The recursive perturbation method and its application to the study of nuclear shell models*

F. T. Hioe

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 24 April 1974)

The recursive perturbation method given previously by the author is generalized and applied to the calculation of the energy levels of nuclear shell models. Specifically, we consider a numerically exactly soluble model of interacting fermions with $SU(3)$ symmetry studied by Li, Klein, and Dreizler. The method can be obviously extended to the study of more general models with $SU(n)$ symmetry.

1. INTRODUCTION

A new perturbation method for quantum mechanical problems was given recently by the author[1] by which the successive terms of a perturbation series for the energies of the system considered are obtained by a set of algebraic recurrence relations rather than by iterations or diagrammatic methods as was done traditionally. The recursive perturbation method, as we shall call it, is not only a powerful method for numerical calculation, but it also provides a new line of analytical approach to the study of quantum systems by means of difference equations. Two types of Hamiltonian were given as examples in Ref. 1: One consists of the boson operators an example of which is the Hamiltonian of an anharmonic oscillator, and the other consists of the spin operators an example of which is the Hamiltonian of a two-level nuclear shell model studied by Lipkin, Meshkov, and Glick.[2] The extension of the application of the recursive perturbation method to the calculation of the energy levels of any quantum system the Hamiltonian of which consists of elements of the $SU(n)$ algebras is straightforward,[3] but it was not explicitly presented. In this paper, we consider a specific three-level nuclear model of N interacting fermions with $SU(3)$ symmetry studied by Li, Klein, and Dreizler,[4] and we give an expression for the energies of the "ground-state band" up to the fourth order terms in the coupling parameters. A recurrence relation is also given by which the higher order terms can be quite readily obtained if needed. The energy series is useful in understanding the variation of the various levels as N, the number of fermions, increases. It is, of course, particularly useful when the coupling parameters are small and when the number of fermions N is very large which makes the direct numerical diagonalization of the Hamiltonian impossible. Previous studies of the $SU(3)$ model were done with the boson expansion[5] and transition-operator boson methods.[6,7] The recursive perturbation method bears some resemblance at first sight to these methods but is in fact quite different in details. The extension of the method to the study of more general models with $SU(n)$ symmetry is straightforward.

2. THE *SU(n)* MODEL

The general relation of the symmetric representations of the $U(n)$ group and some numerically exactly solvable nuclear shell models was discussed by Okubo.[8] Suppose we have n single-particle levels (or shells) with energies $\epsilon_\mu(\mu=1,2,\ldots,n)$ each of which is N-fold degenerate. Let $a_{p\mu}$ and $a_{p\mu}^\dagger$ $(p=1,2,\ldots,N,\ \mu=1,2,\ldots,n)$ be the annihilation and creation operators of a nucleon in a state given by quantum numbers p and μ. The Hamiltonian of the nucleus will be written as

$$H=\sum_{\mu=1}^{n}\sum_{p=1}^{N}\epsilon_\mu a_{p\mu}^\dagger a_{p\mu}+\sum\lambda a_{r\alpha}^\dagger a_{s\beta}a_{p\mu}^\dagger a_{q\nu}, \tag{2.1}$$

where λ is a coupling parameter which depends on all quantum numbers r,s,p,q,α,β,μ, and ν. The form (2.1) is too complicated to solve and in the usual approximation one selects only the terms with $r=s$ and $p=q$, namely we consider a Hamiltonian of the form

$$H=\sum_{\mu=1}^{n}\sum_{p=1}^{N}\epsilon_\mu a_{p\mu}^\dagger a_{p\mu}+\sum_{\mu,\nu,\alpha,\beta=1}^{n}\sum_{p,s=1}^{N}\lambda_{\nu\beta}^{\mu\alpha}a_{s\beta}^\dagger a_{s\alpha}a_{p\nu}^\dagger a_{p\mu}. \tag{2.2}$$

If we set

$$G_\nu^\mu=\sum_{p=1}^{N}a_{p\nu}^\dagger a_{p\mu}, \tag{2.3}$$

then, using the commutation relations

$$\{a_{p\mu},\ a_{q\nu}^\dagger\}=\delta_{pq}\delta_{\mu\nu},$$
$$\{a_{p\mu},\ a_{q\nu}\}=\{a_{p\mu}^\dagger,\ a_{q\nu}^\dagger\}=0, \tag{2.4}$$

the Hamiltonian (2.2) can be written as

$$H=\sum_{\mu=1}^{n}\epsilon_\mu G_\mu^\mu+\sum_{\mu,\nu,\alpha,\beta=1}^{n}\lambda_{\nu\beta}^{\mu\alpha}G_\beta^\alpha G_\nu^\mu, \tag{2.5}$$

where G_ν^μ satisfy the commutation relation

$$[G_\nu^\mu,\ G_\beta^\alpha]=\delta_\beta^\mu G_\nu^\alpha-\delta_\nu^\alpha G_\beta^\mu \tag{2.6}$$

and are generators of the $U(n)$ algebra which becomes the $SU(n)$ algebra if the particle number N is conserved. The ground-state band corresponds to the most symmetric representation of the $SU(n)$ group with signature $(N,0,0,\ldots,0)$. The case $n=2$ is the model of Lipkin, Meshkov, and Glick[2] while the case $n=3$ is the model studied by Li, Klein, and Dreizler.[4]

3. THE RECURSIVE PERTURBATION METHOD

The recursive perturbation method formulated by the author[1] consists essentially of three steps:

(1) Use the Bargmann analytic function representation.[9]

(2) At each stage (order) of the perturbation calculation, the eigenfunction is taken to be a power series consisting of a finite number of terms, thus only a finite number of unknown coefficients to be determined (the number being dependent on the order of the perturbation term being calculated and on the form of the perturbing Hamiltonian).

 Copyright © 1974 American Institute of Physics

(3) The unknown coefficients are determined recursively in terms of the known coefficients of the previous orders by comparing coefficients of the like powers of the expansion variables.

The Bargmann representation of the boson and spin operators are given by the following:

$$a^\dagger \to z, \tag{3.1a}$$

$$a \to \frac{\partial}{\partial z}, \tag{3.1b}$$

$$S^+ \to z_1 \frac{\partial}{\partial z_2}, \tag{3.1c}$$

$$S^- \to z_2 \frac{\partial}{\partial z_1}, \tag{3.1d}$$

and

$$S^z \to \tfrac{1}{2}\left(z_1 \frac{\partial}{\partial z_1} - z_2 \frac{\partial}{\partial z_2}\right), \tag{3.1e}$$

where the z's (except the superscript appearing in S^z which denotes the z component) are arbitrary complex variables. The representation (3.1a) and (3.1b) was in fact used many years ago by Fock[10] and the representation (3.1c), (3.1d), and (3.1e) was obtained from the Schwinger representation[11] of angular momentum

$$S^+ \to a_1^\dagger a_2,$$
$$S^- \to a_2^\dagger a_1,$$
$$S^z \to \tfrac{1}{2}(a_1^\dagger a_1 - a_2^\dagger a_2) \tag{3.2}$$

by replacing the boson operators in (3.2) by z and $\partial/\partial z$ according to Eq. (3.1a) and (3.1b). One of the advantages of the Bargmann analytic function representation is the simplicity of the form of the eigenfunctions.[3] Moreover, it provides a unified and systematic way of treating the boson operators and generators of $SU(n)$ algebras.

For a Hamiltonian consisting of the boson operators, the energy equation in the Bargmann representation is

$$H\left(z, \frac{\partial}{\partial z}\right) f(z) = E f(z), \tag{3.3}$$

where

$$f(z) = \sum_{p=0}^{\infty} c_p z^p, \tag{3.4}$$

and for a Hamiltonian consisting of the spin operators, the energy equation in the Bargmann representation is

$$H\left(\tfrac{1}{2}\left(z_1\frac{\partial}{\partial z_1} - z_2\frac{\partial}{\partial z_2}\right),\; z_1\frac{\partial}{\partial z_2},\; z_2\frac{\partial}{\partial z_1}\right) f(z_1, z_2) = E f(z_1, z_2), \tag{3.5}$$

where

$$f(z_1, z_2) = \sum_{p=0}^{2S} c_p z_2^{2S-p} z_1^p, \tag{3.6}$$

and $[S(S+1)]^{1/2}$ is the total spin. More generally, for a Hamiltonian consisting of generators of an $SU(n)$ algebra, the generators G_ν^μ, $\mu, \nu = 1, 2, \ldots, n$, may be represented by

$$G_\nu^\mu = z_\nu \frac{\partial}{\partial z_\mu}, \quad \mu, \nu = 1, 2, \ldots, n \tag{3.7}$$

with the eigenfunction of the Hamiltonian in the most symmetric representation being represented by

$$f(z_1, \ldots, z_n) = \sum_{j_1, \ldots, j_n = 0}^{N} c_{j_1, \ldots, j_n} z_1^{j_1} \cdots z_n^{j_n}. \tag{3.8}$$

Suppose now the Hamiltonian H can be written as the sum of H_0, the unperturbed part and $\sum_{p=1}^{\pi_n} \lambda_p H_1^{(p)}$, the perturbing part, where π_n is equal to $2\binom{n}{2}$ in general and is equal to $\binom{n}{2}$ if all the λ's are real, then we write, for the eigenvalue of H,

$$E^{\{K\}}(\lambda_1, \ldots, \lambda_{\pi_n}) = \sum_{p_1, \ldots, p_{\pi_n}=0}^{\infty} A^{\{K\}}_{p_1, \ldots, p_{\pi_n}} \lambda_1^{p_1} \cdots \lambda_{\pi_n}^{p_{\pi_n}}, \tag{3.9}$$

and for the eigenfunction of H,

$$f^{\{K\}}(\lambda_1, \ldots, \lambda_{\pi_n}; z_1, \ldots, z_n) = \sum_{p_1, \ldots, p_{\pi_n}=0}^{\infty} B^{\{K\}}_{p_1, \ldots, p_{\pi_n}} \times (z_1, \ldots, z_n) \lambda_1^{p_1} \cdots \lambda_{\pi_n}^{p_{\pi_n}}, \tag{3.10}$$

where $\{K\}$ denotes a set of quantum numbers designating a particular unperturbed energy level considered. The crucial step of the recursive perturbation method (a step which also makes the recursive perturbation method distinctive from the other perturbation methods) is to let $B^{\{K\}}_{p_1, \ldots, p_n}(z_1, \ldots, z_n)$ be a *finite* linear combination of powers of $z_1, \ldots, z_n$. If the highest powers of z_i and $\partial/\partial z_i$ in H_1 are P_i and Q_i respectively, it follows from Ref. 1 that by letting

$$B^{\{K\}}_{p_1, \ldots, p_{\pi_n}}(z_1, \ldots, z_n) = B^{\{K\}}_{0, \ldots, 0}(z_1, \ldots, z_n) \sum_{j_1=-Q_1 p}^{P_1 p} \cdots \sum_{j_{n-1}=-Q_{n-1} p}^{P_{n-1} p}{}' \; b^{\{K\}}_{p_1, \ldots, p_{\pi_n}; j_1, \ldots, j_n} \times z_1^{j_1} \cdots z_n^{j_n}, \tag{3.11}$$

where the prime in the summation denotes the exclusion of the term $j_1 = \cdots = j_{n-1} = 0$ and where $p = p_1 + p_2 + \cdots + p_{\pi_n}$, $j_n = -(j_1 + \cdots + j_{n-1})$, substitutions of (3.11) and (3.9) into the eigenvalue equation for H and comparisons of coefficients of like powers of λ's and z's will lead us to a set of recurrence relations by which the coefficients A's in (3.9) can be determined recursively in a consistent and systematic manner. One sees from (3.11) that as the order of the perturbation term p increases, the number of unknowns, b's, to be determined also increases. But the b's are going to be given in terms of the b's of the previous orders, i.e., the b's are determined recursively and, moreover, the b's of the same order are usually determinable individually (i.e., without having to solve a set of simultaneous equations, say, involving several or increasing number of the b's of the same order). It is also interesting to note that $b_{p_1, \ldots, p_n; j_1, \ldots, j_n}$ becomes zero if one or more of the j's has absolute value greater than N [the easiest way to see this is to construct a simple example and then deduce the general case by deduction; see Ref. 1 and the recurrence relation for the $SU(3)$ model in the following section], and thus the powers of the expansion parameters z's are automatically restricted to the range $-N \le j_i \le N$, $i = 1, 2, \ldots, n$.

4. THE $SU(3)$ MODEL

In this section, we apply the recursive perturbation method to the study of a specific $SU(3)$ model considered by Li, Klein, and Dreizler.[4] The model assumes three

N-fold degenerate single-particle shells with energies ϵ_1, ϵ_2, and ϵ_3 and the Hamiltonian of the model is assumed to be

$$H=\sum_{\mu=1}^{3}\epsilon_\mu G_{\mu\mu}+\lambda_1(G_{23}^2+G_{32}^2)+\lambda_2(G_{13}^2+G_{31}^2)+\lambda_3(G_{12}^2+G_{21}^2), \tag{4.1}$$

where the two-body interactions λ_1, λ_2, and λ_3 existing between shells are assumed to be real, and the operators G's are given by

$$G_{\nu\mu}=\sum_{p=1}^{N} a_{p\nu}^\dagger a_{p\mu} \tag{4.2}$$

$a^\dagger$, a being the fermion creation and annihilation operators. In the Bargmann analytic function representation, the Hamiltonian is written as

$$H=\left(\epsilon_1 u\frac{\partial}{\partial u}+\epsilon_2 v\frac{\partial}{\partial v}+\epsilon_3 w\frac{\partial}{\partial w}\right)+\lambda_1\left(v^2\frac{\partial^2}{\partial w^2}+w^2\frac{\partial^2}{\partial v^2}\right)$$
$$+\lambda_2\left(u^2\frac{\partial^2}{\partial w^2}+w^2\frac{\partial^2}{\partial u^2}\right)+\lambda_3\left(u^2\frac{\partial^2}{\partial v^2}+v^2\frac{\partial^2}{\partial u^2}\right). \tag{4.3}$$

Consider an energy level in the ground state band characterized by the quantum numbers a, b, and c $(=N-a-b)$, i.e., this level has an energy $a\epsilon_1+b\epsilon_2+c\epsilon_3$ in the absence of any interations between shells. The eigenfunction corresponding to this energy level is $u^a v^b w^c$. When the interactions are present, let the same energy level now become

$$E^{abc}(\lambda_1,\lambda_2,\lambda_3)=\sum_{p_1,p_2,p_3=0}^{\infty} A_{p_1,p_2,p_3}^{abc}\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}, \tag{4.4}$$

where $A_{0,0,0}^{abc}=a\epsilon_1+b\epsilon_2+c\epsilon_3$, and let the corresponding eigenfunction now become

$$f^{abc}(\lambda_1,\lambda_2,\lambda_3;\,u,v,w)=\sum_{p_1,p_2,p_3=0}^{\infty} B_{p_1,p_2,p_3}^{abc}(u,v,w)\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}$$
$$=B_{0,0,0}^{abc}(u,v,w)\sum_{p_1,p_2,p_3=0}^{\infty}\beta_{p_1,p_2,p_3}^{abc}(u,v,w)\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}, \tag{4.5}$$

where $B_{0,0,0}^{abc}(u,v,w)=u^a v^b w^c$, $\beta_{0,0,0}^{abc}(u,v,w)=1$ and, for $p\geq 1$,

$$\beta_{p_1,p_2,p_3}^{abc}(u,v,w)=\sum_{i=-2p}^{2p}\sum_{j=-2p}^{2p}{}' b_{p_1,p_2,p_3;\,i,j,k}^{abc}u^i v^j w^k, \tag{4.6}$$

where $p=p_1+p_2+p_3$, $k=-i-j$ and the prime in the summation denotes the exclusion of the term $i=j=0$. The coefficients A's and b's are the coefficients to be determined and the crucial part of the recursive perturbation method for this problem is expressed by Eq. (4.6). With

$$u^2\frac{\partial^2}{\partial v^2}B_{p_1,p_2,p_3}^{abc}(u,v,w)$$
$$=u^a v^b w^c\left(b(b-1)u^2v^{-2}+2bu^2v^{-1}\frac{\partial}{\partial v}+u^2\frac{\partial^2}{\partial v^2}\right)$$
$$\times\beta_{p_1,p_2,p_3}^{abc}(u,v,w) \tag{4.7}$$

and similar expressions for $v^2(\partial^2/\partial u^2)B_{p_1,p_2,p_3}^{abc}(u,v,w)$, $u^2(\partial^2/\partial w^2)B_{p_1,p_2,p_3}^{abc}(u,v,w)$, etc., we get, by substitutions into the energy equation and comparisons of the coefficients of $\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}$,

$$\sum_{p_1,p_2,p_3}\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}\left[\left(\epsilon_1 u\frac{\partial}{\partial u}+\epsilon_2 v\frac{\partial}{\partial v}+\epsilon_3 w\frac{\partial}{\partial w}\right)\right.$$
$$\times\beta_{p_1,p_2,p_3}^{abc}(u,v,w)+\left(c(c-1)v^2w^{-2}+2cv^2w^{-1}\frac{\partial}{\partial w}\right.$$
$$\left.+v^2\frac{\partial^2}{\partial w^2}+b(b-1)v^{-2}w^2+2bv^{-1}w^2\frac{\partial}{\partial v}+w^2\frac{\partial^2}{\partial v^2}\right)$$
$$\times\beta_{p_1-1,p_2,p_3}^{abc}(u,v,w)+\left(c(c-1)u^2w^{-2}+2cu^2w^{-1}\frac{\partial}{\partial w}\right.$$
$$\left.+u^2\frac{\partial^2}{\partial w^2}+a(a-1)u^{-2}w^2+2au^{-1}w^2\frac{\partial}{\partial u}+w^2\frac{\partial^2}{\partial u^2}\right)$$
$$\times\beta_{p_1,p_2-1,p_3}^{abc}(u,v,w)+\left(b(b-1)u^2v^{-2}+2bu^2v^{-1}\frac{\partial}{\partial v}\right.$$
$$\left.+u^2\frac{\partial^2}{\partial v^2}+a(a-1)u^{-2}v^2+2au^{-1}v^2\frac{\partial}{\partial u}+v^2\frac{\partial^2}{\partial u^2}\right)$$
$$\left.\times\beta_{p_1,p_2,p_3-1}^{abc}(u,v,w)\right]$$
$$=\sum_{p_1,p_2,p_3}\lambda_1^{p_1}\lambda_2^{p_2}\lambda_3^{p_3}\sum_{q_1=0}^{p_1}\sum_{q_2=0}^{p_2}\sum_{q_3=0}^{p_3}{}^{*}A_{p_1-q_1,p_2-q_2,p_3-q_3}^{abc}$$
$$\times\beta_{q_1,q_2,q_3}^{abc}(u,v,w),$$

where the star in the summation denotes the exclusion of the term $q_1=p_1$, $q_2=p_2$, and $q_3=p_3$ simultaneously. Comparing the coefficients of like powers of u, v, and w on both sides, we obtain the following recurrence relation by which the A's can be obtained readily by recursion (omitting the superscripts a, b, and c for convenience):

$$(i\epsilon_1+j\epsilon_2+k\epsilon_3)b_{p_1,p_2,p_3;\,i,j,k}$$
$$+(a+i+1)(a+i+2)$$
$$\times(b_{p_1,p_2-1,p_3;\,i+2,j,k-2}+b_{p_1,p_2,p_3-1;\,i+2,j-2,k})$$
$$+(b+j+1)(b+j+2)$$
$$\times(b_{p_1-1,p_2,p_3;\,i,j+2,k-2}+b_{p_1,p_2,p_3-1;\,i-2,j+2,k})$$
$$+(c+k+1)(c+k+2)$$
$$\times(b_{p_1-1,p_2,p_3;\,i,j-2,k+2}+b_{p_1,p_2-1,p_3;\,i-2,j,k+2})$$
$$=\sum_{q_1=0}^{p_1}\sum_{q_2=0}^{p_2}\sum_{q_3=0}^{p_3}{}^{*}A_{p_1-q_1,p_2-q_2,p_3-q_3}b_{q_1,q_2,q_3;\,i,j,k} \tag{4.8}$$

with $b_{0,0,0;\,i,j,k}=\delta_{0i}\delta_{0j}\delta_{0k}$ and $b_{q_1,q_2,q_3;\,i,j,k}=0$ if $|i|$, $|j|$, or $|k|$ is greater than $2(q_1+q_2+q_3)$ or if one or more of the q's is negative [see Eq. (4.6)]. Using (4.8), we readily find:

$p=1$, $A_{1,0,0}^{abc}=A_{0,1,0}^{abc}=A_{0,0,1}^{abc}=0$.

$p=2$,

$$A_{2,0,0}^{abc}=\frac{1}{2(\epsilon_3-\epsilon_2)}\{(b+1)(b+2)c(c-1)$$
$$-(c+1)(c+2)b(b-1)\},$$

$A_{0,2,0}^{abc}$ is obtained from $A_{2,0,0}^{abc}$ by changing $\epsilon_1\to\epsilon_2$, $\epsilon_2\to\epsilon_3$, $\epsilon_3\to\epsilon_1$ and $a\to b$, $b\to c$, and $c\to a$, and $A_{0,0,2}^{abc}$ is obtained from $A_{0,2,0}^{abc}$ by the same permutation,

$$A_{1,1,0}^{abc}=A_{1,0,1}^{abc}=A_{0,1,1}^{abc}=0.$$

$p=3$,

$$A_{1,1,1}^{abc}=(a+1)(a+2)(b+1)(b+2)(c-1)c/$$

$$2(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)+(a+1)(a+2)(b-1)b(c+1)(c+2)/$$
$$2(\epsilon_2-\epsilon_1)(\epsilon_2-\epsilon_3)+(a-1)a(b+1)(b+2)(c+1)(c+2)/$$
$$2(\epsilon_1-\epsilon_2)(\epsilon_1-\epsilon_3)+(a-1)a(b-1)b(c+1)(c+2)/$$
$$2(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)+(a-1)a(b+1)(b+2)(c-1)c/$$
$$2(\epsilon_2-\epsilon_1)(\epsilon_2-\epsilon_3)+(a+1)(a+2)(b-1)b(c-1)c/$$
$$2(\epsilon_1-\epsilon_2)(\epsilon_1-\epsilon_3),$$
$$A^{abc}_{3,0,0}=A^{abc}_{0,3,0}=\cdots=0,$$
$$A^{abc}_{2,1,0}=A^{abc}_{1,2,0}=\cdots=0.$$

$p=4$,

$$A^{abc}_{3,1,0}=A^{abc}_{1,3,0}=A^{abc}_{1,0,3}=\cdots=0,$$
$$\begin{aligned}A^{abc}_{4,0,0}=&\frac{(b+1)(b+2)c(c-1)}{8(\epsilon_3-\epsilon_2)^3}[\tfrac{1}{2}(b+3)(b+4)(c-2)(c-3)\\&-(b+1)(b+2)(c-1)c+(b-1)b(c+1)(c+2)]\\&-\frac{(c+1)(c+2)b(b-1)}{8(\epsilon_3-\epsilon_2)^3}[\tfrac{1}{2}(c+3)(c+4)(b-2)(b-3)\\&-(c+1)(c+2)(b-1)b+(c-1)c(b+1)(b+2)];\end{aligned}$$

$A^{abc}_{0,4,0}$ and $A^{abc}_{0,0,4}$ are obtained from $A^{abc}_{4,0,0}$ by permuting ϵ_1, ϵ_2, ϵ_3, and a, b, c;

$$\begin{aligned}A^{abc}_{2,2,0}=&\frac{(a+1)(a+2)c(c-1)}{8(\epsilon_3-\epsilon_1)}\biggl(\frac{(b-1)b(c-1)c}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_1)}\\&-\frac{(b-1)b(c+1)(c+2)}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_2)}-\frac{(b+1)(b+2)c(c-1)}{(\epsilon_3-\epsilon_2)(\epsilon_3-\epsilon_1)}\\&+\frac{(c+1)(c+2)b(b-1)}{(\epsilon_3-\epsilon_2)(\epsilon_3-\epsilon_1)}\biggr)+\frac{(b+1)(b+2)c(c-1)}{8(\epsilon_3-\epsilon_2)}\\&\biggl(-\frac{(a-1)a(c-1)c}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_2)}+\frac{(a-1)a(c+1)(c+2)}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_1)}\\&-\frac{(a+1)(a+2)c(c-1)}{(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)}+\frac{(c+1)(c+2)a(a-1)}{(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)}\biggr)\\&-\frac{(c+1)(c+2)(b-1)b}{8(\epsilon_3-\epsilon_2)}\biggl(\frac{(a+1)(a+2)(c-1)c}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_1)}\\&-\frac{(a+1)(a+2)(c+1)(c+2)}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_2)}+\frac{(a+1)(a+2)c(c-1)}{(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)}\\&-\frac{(c+1)(c+2)a(a-1)}{(\epsilon_3-\epsilon_1)(\epsilon_3-\epsilon_2)}\biggr)-\frac{(c+1)(c+2)a(a-1)}{8(\epsilon_3-\epsilon_1)}\\&\biggl(-\frac{(b+1)(b+2)(c-1)c}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_2)}+\frac{(b+1)(b+2)(c+1)(c+2)}{(\epsilon_2-\epsilon_1)(\epsilon_3-\epsilon_1)}\\&+\frac{(b+1)(b+2)c(c-1)}{(\epsilon_3-\epsilon_2)(\epsilon_3-\epsilon_1)}-\frac{(c+1)(c+2)b(b-1)}{(\epsilon_3-\epsilon_2)(\epsilon_3-\epsilon_1)}\biggr);\end{aligned}$$

$A^{abc}_{0,2,2}$ and $A^{abc}_{2,0,2}$ are obtained from $A^{abc}_{2,2,0}$ by permuting ϵ_1, ϵ_2, ϵ_3 and a, b, c. By letting $\epsilon_1=-\epsilon_2=\frac{1}{2}\epsilon$, $\epsilon_3=0$, $\lambda_1=\lambda_2=0$, $\lambda_3=\lambda$ (λ_3 denotes the interaction between levels 1 and 2), $c=0$ and $b=N-a$, we obtain the corresponding result for the $SU(2)$ model studied by Lipkin, Meshkov, and Glick[2] for which the Hamiltonian is assumed to be

$$H=\epsilon S^z+\lambda(S^{+2}+S^{-2}). \tag{4.9}$$

Thus, for the ath energy level, the energy is given by

$$E^a(\lambda)=\sum_{p=0}^{\infty}A_p^a\lambda^p, \tag{4.10}$$

where

$$A_0^a=(-\tfrac{1}{2}N+a)\epsilon,$$
$$A_1^a=A_3^a=A_5^a=\cdots=0,$$
$$\begin{aligned}A_2^a=&(1/2\epsilon)[(N-a+1)(N-a+2)a(a-1)\\&-(a+1)(a+2)(N-a)(N-a-1)],\end{aligned}$$
$$\begin{aligned}A_4^a=&(1/16\epsilon^3)(N-a+1)(N-a+2)a(a-1)\\&\times[(N-a+3)(N-a+4)(a-2)(a-3)\\&-2(N-a+1)(N-a+2)a(a-1)\\&+2(a+1)(a+2)(N-a)(N-a-1)]\\&-(1/16\epsilon^3)(a+1)(a+2)(N-a)(N-a-1)\\&\times[(a+3)(a+4)(N-a-2)(N-a-3)\\&-2(a+1)(a+2)(N-a)(N-a-1)\\&+2(N-a+1)(N-a+2)a(a-1)].\end{aligned}$$

Thus, the energy of the first state above the ground level is given by

$$\begin{aligned}E^1(\lambda)-E^0(\lambda)=&\epsilon-(2/\epsilon)(N-1)(N-3)\lambda^2\\&-(2/\epsilon^3)(N-1)(N-3)(N^2-16N+27)\lambda^4\\&+\cdots,\end{aligned} \tag{4.11}$$

which is the result given by Lipkin, Meshkov, and Glick. More generally, the coefficients A's may be obtained from the following recurrence relation:

$$\begin{aligned}&\epsilon j\,b_{p;j,-j}+(N-a-j+1)(N-a-j+2)b_{p-1;j-2,-j+2}\\&\quad+(a+j+1)(a+j+2)b_{p-1;j+2,-j-2}\\&=\sum_{q=0}^{p-1}A_{p-q}b_{q;j,-j}\end{aligned} \tag{4.12}$$

with $b_{0;j,-j}=\delta_{0,j}$, $b_{p;j,-j}=0$ for $|j|>2p$.

5. SUMMARY

We have generalized the recursive perturbation method and have demonstrated how it can be applied to the perturbation calculation of the energies of nuclear shell models with $SU(n)$ symmetry. Specifically, we have calculated the energies of the ground state band of the $SU(3)$ model studied by Li, Klein, and Dreizler up to the fourth order terms in the coupling parameters. We have also presented a simple recurrence relation (4.8) by which the higher order terms can be quite readily computed if needed. The recursive perturbation method is a powerful numerical method particularly suited for the computer, and we feel that it will certainly find application in many other problems in physics.[12]

ACKNOWLEDGMENTS

The author is most grateful to Professor Susumu Okubo for introducing to him the nuclear shell models. The author is also grateful to Professor Elliott Montroll for his interest and support.

*Research partially supported by the National Science Foundation Grant GU-4040.

[1]F.T. Hioe, J. Math. Phys. 15, 445 (1974).

[2]H.J. Lipkin, N. Meshkov and A.J. Glick, Nucl. Phys. **62**, 188 (1965).
[3]F.T. Hioe, J. Math. Phys. **15**, 1174 (1974).
[4]S.Y. Li, A. Klein, and R.M. Dreizler, J. Math. Phys. **11**, 975 (1970).
[5]S.C. Pang, A. Klein, and R.M. Dreizler, Ann. Phys. (N.Y.) **49**, 477 (1968).
[6]R.M. Dreizler and A. Klein, Phys. Rev. C **7**, 512 (1973).
[7]S. Okubo, Phys. Rev. C **9**, 1188 (1974).
[8]S. Okubo, "Symmetric representations of the $U(n)$ group and exactly solvable nuclear model" (unpublished).
[9]V. Bargmann, Comm. Pure Appl. Math. **20**, 1 (1967); Rev. Mod. Phys. **34**, 829 (1962).
[10]V. Fock, Z. Phys. **49**, 339 (1928).
[11]J. Schwinger, "On angular momentum," U.S. Atomic Energy Commission, NYO-3071 (1952), reprinted in *Quantum Theory of Angular Momentum*, edited by L.C. Biedenharn and H. Van Dam (Academic, New York, 1965).
[12]For an application of the recursive perturbation method to a problem in quantum optics, see F.T. Hioe and J.H. Eberly, "Radiative frequency shifts of a two-level system," preprint.

Local bounded perturbations of KMS states

E. Presutti

Istituto Fisico Università dell'Aquila, 67100 L'Aquila, Italy

E. Scacciatelli

Istituto Matematico Università di Roma, P.le delle Scienze 5, 00185 Roma, Italy
Gruppo Nationale per la Fisica Matematica del Consiglio Nazionale Delle Ricerche, Roma, Italy
(Received 22 February 1973)

Local bounded perturbations of an infinite equilibrium state are studied in the C^*-algebraic framework. It is assumed that for the unperturbed state the pressure exists in the thermodynamic limit and that the Dubin–Sewell hypotheses are fulfilled. The following is then shown: At constant temperature the perturbed state is analytic in the perturbation Q, the infinite volume pressure does not depend on Q, and the new state is KMS with respect to the time evolution corresponding to the adiabatic perturbation, as treated in a previous paper.

1. INTRODUCTION AND MAIN RESULTS

The isothermal response of a macroscopic system S to a local perturbation Q has been extensively studied. Recently some results have been obtained. It has been shown[1] that if the macroscopic system is represented by a density matrix in a Hilbert space and the local perturbation is determined by a self-adjoint bounded operator Q, then the perturbed matrix (at constant temperature) is analytic in Q (i.e., in the perturbation strength) and the pertubative series has been explicated.

This provides the necessary tools for the perturbative study of the corresponding state on the C^*-algebra of the system observables[2]. Finally, the convergence of "real and complex time" correlations functions for KMS states has been proved in Refs. 3 and 4.

In this paper we study the perturbative expansion directly in the thermodynamic limit. For this purpose, we define the infinite equilibrium state as a positive functional on a C^*-algebra obtained as the norm closure of the C^*-algebras of the bounded regions and we study the dependence on the local perturbation Q. We assume for technical reasons that Q is described by a bounded operator. This is not a loss of generality if lattice spaces or hard-core particles are considered, while the case of continuous particles with unbounded perturbation requires an *ad hoc* treatment.

The hypotheses we use in this paper concern the behavior of the unperturbed time correlation functions in the thermodynamic limit. We assume the existence of the infinite volume pressure and that the Dubin and Sewell conditions[5] are fulfilled. This assures the existence of dynamics in the Hilbert space in which the unperturbed equilibrium state is represented by a vector via the GNS construction.

The results we then obtain can be summarized as follows:

The perturbed state exists in the thermodynamic limit and can be implemented by a density matrix in the GNS space of the unperturbed state. The state, furthermore, is analytic in Q. This fact could play an important role in the study of the time relaxation of the adiabatically perturbed infinite state which in the remote past was in equilibrium for the unperturbed evolution. Analyticity was an hypothesis in a previous work[6] in which the above problem appeared as the approach to equilibrium of a spin in a thermostat. There it was proved that the first terms (in Q) of the adiabatic response correctly approached the isothermal ones, under suitable hypotheses on the kernel of the master equation describing the time evolution.

The perturbed state verifies the Dubin–Sewell conditions w.r.t. the perturbed time correlations functions, so that it is KMS.

It is shown that the time evolution corresponding to the perturbed state via the Tomita theorem[7] is the same as the one we introduced in a previous paper[8] when the adiabatic perturbation was considered.

The perturbed pressure exists in the thermodynamic limit and does not depend on Q.

2. MATHEMATICAL FRAMEWORK AND NOTATIONS

Notations

(a) We employ the standard symbols $\mathbb{C}, \mathbb{R}, \mathbb{R}_+, \mathbb{Z}, \mathbb{Z}_+$, to denote the complex plane, the real line, the positive reals, the integers, the positive integers.

(b) Let $\mathfrak{A}$ be a C^*-algebra[9] with identity. We denote by $\mathfrak{A}^*, \mathfrak{A}^*_+$ the set of all continuous and the set of all positive continous functionals on $\mathfrak{A}$. We denote by $S(\mathfrak{A})$, the set of the states on $\mathfrak{A}$, i.e.,

$$S(\mathfrak{A}) \equiv \{\psi \in \mathfrak{A}^*_+, \psi(I) = 1\}.$$

Sometimes we denote $\psi(A)$, $\psi \in \mathfrak{A}^*$, $A \in \mathfrak{A}$, by the symbol $\langle\psi;A\rangle$. We denote by the symbol $\mathfrak{A}_r$, the r-ball of $\mathfrak{A}$, i.e.,

$$\mathfrak{A}_r \equiv \{A \in \mathfrak{A}; \|A\| < r\}.$$

(c) Let $\mathfrak{A}$ be a C^*-algebra with identity, $\psi \in S(\mathfrak{A})$. Let $(\mathfrak{H}_\psi; \pi_\psi; \Omega_\psi)$ be the GNS triple induced by ψ. We denote by $\tilde{\psi}$, the extension of ψ to $\pi_\psi(\mathfrak{A})''$, defined by

$$\tilde{\psi}(\cdot) = (\Omega_\psi, (\cdot)\Omega_\psi). \tag{2.1}$$

We denote by $\mathfrak{A}_*(\psi)$ the set of all ultraweakly continuous functionals w.r.t. the π_ψ representation space:

$$\phi \in \mathfrak{A}_*(\psi) \Rightarrow \phi = \tilde{\phi} \circ \pi_\psi. \tag{2.2}$$

We denote by $\mathcal{S}(\psi, A)$ the island of ψ:

$$\mathcal{S}(\psi, \mathfrak{A}) \equiv \{\phi \in S(\mathfrak{A}) : \phi = \tilde{\phi} \circ \pi_\psi\}. \tag{2.3}$$

Definition 2.1: Let ϕ be a state on $\mathfrak{A}$. Let τ_t be an homorphism of the real line into Aut$\mathfrak{A}$. Let $\beta \in \mathbb{R}^+$; we say that Φ satisfies the KMS conditions[10,11] corresponding to (τ_t, β) if $\forall A, B \in \mathfrak{A}$, $\exists$ functions f_{AB}, g_{AB} on the complex plane $\mathbb{C}$, such that:

(i) $\quad f_{AB}(t) = \langle \phi, (\tau_t A)B \rangle \; g_{AB}(t) = \langle \phi, B(\tau_t A) \rangle,$
$\quad \forall A, B \in \mathfrak{A}, \quad \forall t \in \mathbb{R}.$

 Copyright © 1974 American Institute of Physics

(ii) $f_{AB}[g_{AB}]$ is analytic in the strip $\mathrm{Im}z \in \{-\beta, 0\}$ $[\mathrm{Im}z \in \{0, \beta\}]$ and continuous on its boundaries.

(iii) $f_{AB}(z) = g_{AB}(z + i\beta), \quad \forall z \in C.$

Mathematical framework

In the algebraic formulation of statistical mechanics [10,12] one considers the physical space Γ (i.e., $\mathbb{R}^\nu$ for continuous particles or $\mathbb{Z}^\nu$ for spin systems) of a system $\bar{S}$ and the set $L = \{\Lambda\}$ of the bounded regions in which $\bar{S}$ can be confined. To each $\Lambda \in L$ the algebra of observables is represented by a type 1 factor $\mathfrak{U}_\Lambda$ of operators in a Hilbert space $\mathfrak{H}_F$. This will be referred to as the $\mathfrak{H}_F$ representation of the algebra. One further supposes that the algebras $\mathfrak{U}_\Lambda$ are isotonic, that is if $\Lambda, \Lambda' \in L$ and $\Lambda' \supset \Lambda$ then $\mathfrak{U}_{\Lambda'} \supset \mathfrak{U}_\Lambda$.

Let $\mathfrak{U}_L = \bigcup_{\Lambda \in L} \mathfrak{U}_\Lambda$ and $\mathfrak{A}$ be the norm closure of $\mathfrak{U}_L$

$$\mathfrak{A} = \overline{\bigcup_{\Lambda \in L} \mathfrak{U}_\Lambda} \tag{2.4}$$

Thus $\mathfrak{A}$ is a C^*-algebra (with identity), and it is termed the algebra of quasilocal observables for the system. A state of $\bar{S}$ may be represented by a state on $\mathfrak{A}$. To define a Gibbs state for $\bar{S}$, one considers an increasing sequence $\{\Lambda_n\}, \Lambda_n \in L$, such that $\bigcup_n \Lambda_n = \Gamma; n \in \mathbb{Z}_+$. For each Λ_n, one supposes the existence of two self-adjoint operators in $\mathfrak{H}_{F\Lambda_n}$, $H_0^{(n)}$, and $N^{(n)}$, corresponding to the Hamiltonian and particle number for a system $\bar{S}^{(n)}$ occupying Λ_n and subject to prescribed boundary conditions. If for every $n \in Z_+$ the operator $H^{(n)} = H_0^{(n)} - \mu N^{(n)}$ (μ is the chemical potential), is self-adjoint and lower bounded, and if the operator $\exp(-\beta H^{(n)})$ is of trace-class on $\mathfrak{U}_{\Lambda_n}$ for $\beta \in R_+$ it may be defined a state $\phi^{(n)}_{\beta,\mu}$ on $\mathfrak{U}_{\Lambda_n}$ by

$$\langle \phi^{(n)}_{\beta,\mu}; A\rangle = \{\mathrm{Tr}_n \exp(-\beta H^{(n)})\}^{-1} \times \{\mathrm{Tr}_n [\exp(-\beta H^{(n)})]A\}, \quad \forall A \in \mathfrak{U}_{\Lambda_n}. \tag{2.5}$$

If $\lim_n \phi^{(n)}_{\beta,\mu}(A)$ exists $\forall A \in \mathfrak{U}_L$ then, since $\mathfrak{U}_L$ is norm dense in a $\mathfrak{A}$, the limit defines a state $\phi_{\beta,\mu}$ on $\mathfrak{A}$ that is named a Gibbs state for $\bar{S}$:

$$\langle \phi_{\beta,\mu}; A\rangle = \lim_n \langle \phi^{(n)}_{\beta,\mu}; A\rangle, \quad \forall A \in \mathfrak{U}_L. \tag{2.6}$$

(See Ref. 13.) The thermodynamic potentials are, on the other hand studied by means of the partition function; we also consider the possibility that the thermodynamic limit exists for the sequence $|\Lambda_n|^{-1} \ln[\mathrm{Tr}_n \exp(-\beta H^{(n)})]$. The limit (2.6) defines a state which is locally normal w.r.t. the $\mathfrak{H}_F$ representation, i.e.,

$$\sigma_\Lambda = \phi|_{\mathfrak{U}_\Lambda} \in [\pi_F(\mathfrak{A})'']_*, \tag{2.7}$$

where π_F is the $\mathfrak{H}_F$ representation for $\mathfrak{A}$.

In order to consider time translations, one defines

$$\tau_t^{(n)}A: \mathfrak{U}_{\Lambda_n} \to \mathfrak{U}, \quad n \in \mathbb{Z}_+, \tag{2.8}$$

where $\tau_t^{(n)}$ is an homomorphism of the real line into Aut $\mathfrak{A}$:

$$\tau_t^{(n)}A = U^{(n)}(t)AU^{(n)}(-t) = A^{(n)}(t), \tag{2.9}$$

$$U^{(n)}(t) = \exp(iH^{(n)}t). \tag{2.10}$$

To define time translations in the thermodynamical limit, Dubin and Sewell[5] made the following assumptions:

D.S.I.

$$\lim_n \langle \phi^{(n)}; A_1^{(n)}(t_1)\cdots A_k^{(n)}(t_k)\rangle \text{ exists } \forall A_1\cdots A_k \in \mathfrak{U}_L, \quad k \in \mathbb{Z}_+, \quad t_1\cdots t_k \in \mathbb{R}. \tag{2.11}$$

D.S.II

$$\lim_m \lim_n \langle \phi^{(n)}; A_1^{(n)}(t_1)\cdots A_k^{(n)}(t_k) \times A^{(m)}_{k+1}(t_{k+1})\cdots A^{(m)}_{k+s}(t_{k+s})\rangle = \lim_n \langle \phi^{(n)}; A_1^{(n)}(t_1)\cdots A^{(n)}_{k+s}(t_{k+s})\rangle$$
$$\forall A_1\cdots A_{k+s} \in \mathfrak{U}_L, \quad k, s \in \mathbb{Z}_+, \quad t_1\cdots t_{k+s} \in \mathbb{R}. \tag{2.12}$$

We report below their main results.

(i) There exists a Gibbs state ϕ, locally normal w.r.t. the Fock representation, and an homomorphism τ_t of the real line into Aut. $\pi_\phi(\mathfrak{A})''$, such that

$$\lim_n \langle \phi^{(n)}; A_1^{(n)}(t_1)\cdots A^{(n)}_k(t_k)\rangle = \langle \tilde{\phi}; \tau_{t_1}\pi_\phi(A_1)\cdots \tau_{t_k}\pi_\phi(A_k)\rangle,$$
$$\forall A_1\ldots A_k \in \mathfrak{U}_L, \quad k \in \mathbb{Z}_+, \quad t_1\ldots t_k \in R. \tag{2.13}$$

(ii) $$\tau_t\pi_\phi(A) = U(t)\pi_\phi(A)U(-t), \quad \forall A \in \mathfrak{U}_L. \tag{2.14}$$

$U(t)$ is a unitary strongly continuous operator in $\mathfrak{H}_\phi$ and

$$\langle \tilde{\phi}; \tau_t\pi_\phi(A)\rangle = (\Omega\phi, U(t)\pi_\phi(A)U(-t)\Omega\phi) = (\Omega\phi, \pi_\phi(A)\Omega\phi), \quad \forall A \in \mathfrak{U}_L. \tag{2.15}$$

(iii) $\tilde{\phi}$ is a KMS state w.r.t. (τ_t, β).

Remark 2.1: If $\Psi \in S(\mathfrak{A})$ is KMS w.r.t. $(\beta, \gamma(t))$, it follows that Ω_Ψ is both cyclical and separating w.r.t. $\pi_\Psi(\mathfrak{A})''$ [11] (see Ref. 7) and Ψ is invariant w.r.t. $\gamma(t)$. As a consequence, $\gamma(t)$ is unique, via the Tomita theorem, and τ_t, defined in point (ii), is the Tomita automorphism.

In a previous paper[8] the consequence of an adiabatic perturbation of the state ϕ, by a local bounded interaction Q was studied. The main results for the perturbed state ϕ_t are

(i) $$\phi_t \in \mathcal{S}(\phi, \mathfrak{A}), \quad \forall t \in R. \tag{2.16}$$

(ii) A new homomorphism $\tau_p(t)$ (β, μ depending) of the real line into Aut$\pi_\phi(\mathfrak{A})''$ is defined

$$\tau_p(t)A = U_p(t)AU_p(-t), \quad \forall A \in \pi_\phi(\mathfrak{A})'', \tag{2.17}$$

$$U_p(t) = \exp\{i[H + \pi_\phi(Q)]t\}, \tag{2.18}$$

where

$$H = \text{s-}\lim_{t\to 0} t^{-1}(U(t) - I). \tag{2.19}$$

In Eq. (2.16) ϕ_t can be written as

$$\phi_t(A) = \langle \tilde{\phi}; \tau_p(t)\pi_\phi(A)\rangle, \quad \forall A \in \mathfrak{U}_L. \tag{2.20}$$

3. ISOTHERMAL PERTURBATION OF Φ

In this paper we want to study the isothermal perturbation of ϕ, via a local bounded interaction Q.

Let Q belong to $\mathfrak{U}_{\Lambda_{n_0}}$, $Q = Q^+$. Let us consider the increasing sequence $\{\Lambda_n\}$ above defined, with $n > n_0$. Let us consider again the sequence of systems $\mathcal{S}^{(n)}$, with Hamiltonians $H_{0p}^{(n)} = H_0^{(n)} + Q$ and particle numbers $N^{(n)}$, so that the operators

$$H_p^{(n)} = H_{0p}^{(n)} - \mu N^{(n)} = H^{(n)} + Q \tag{3.1}$$

are self-adjoint and lower bounded. The operators

$$\exp(-\beta H_p^{(n)}) \tag{3.2}$$

are of trace class in $\mathfrak{H}_{\Lambda_n}$, $\forall n > n_0$, $\beta \in \mathbb{R}_+$ (see Ref. 1). We can then define the sequence of states

$$\langle \phi_p^{(n)}; A\rangle = \{\mathrm{Tr}_n \exp(-\beta H_p^{(n)})\}^{-1}\{\mathrm{Tr}_n[\exp(-\beta H_p^{(n)})]A\},$$
$$\forall A \in \mathfrak{U}_{\Lambda_n}, \quad n > n_0. \tag{3.3}$$

Again in order to consider time translations, we define

$$\tau_p^{(n)}(t)\colon \mathfrak{U}_{\Lambda_n} \to \mathfrak{U}_{\Lambda_n}, \quad \forall n > n_0, \tag{3.4a}$$

$$\tau_p^{(n)}(t)A = U_p^{(n)}(t)AU_p^{(n)}(-t), \quad \forall A \in \mathfrak{U}_{\Lambda_n}, \quad n > n_0, \tag{3.4b}$$

$$U_p^{(n)}(t) = \exp(iH_p^{(n)}t). \tag{3.5}$$

We also consider the sequence of perturbed functions

$$Z_p^{(n)}(\beta, \mu) = \mathrm{Tr}_n \exp(-\beta H_p^{(n)}). \tag{3.6}$$

Theorem 3.1: Under D.S.I, D.S.II hypotheses [Eqs. (2.11), (2.12)] and with the above definitions and assumptions, the following holds:

(i) $\lim_n \phi_p^{(n)}(A) = \phi_p(A), \quad \forall A \in \mathfrak{U}_L,$ (3.7)

where $\phi_p \in \mathcal{S}(\phi, \mathfrak{A})$.

(ii) for fixed β and μ there exists an unique homomorphism of the real line into $\pi_{\phi_p}(\mathfrak{A})''$ such that

$$\lim \langle \phi_p^{(n)}; \hat{A}_1^{(n)}(t_1)\cdots A_k^{(n)}(t_k)\rangle = \langle \tilde{\phi}_p; (\tau_p(t_1)\pi_{\phi_p}(A)) \cdots (\tau_p(t_k)\pi_{\phi_r}(A_k))\rangle, \quad \forall A_1\cdots A_k \in \mathfrak{U}_L, \quad k \in \mathbb{Z}_+,$$
$$t_1\cdots t_k \in R, \tag{3.8}$$

where

$$\tau_p(t)\pi_{\phi_p}(A) = U_p(t)\pi_{\phi_p}(A)U_p(-t), \quad \forall A \in \mathfrak{U}_L, \tag{3.9}$$

$$U_p(t) = \exp\{i[H + \pi_p(Q)]t\}, \tag{3.10}$$

$$H = \operatorname*{s-lim}_{t\to 0}(U(t) - I)t^{-1}. \tag{3.11}$$

(iii) $\tilde{\phi}_p$ is KMS w.r.t. $(\beta, \tau_p(t))$.

In order to prove Theorem (3.1) we have to enunciate some lemmas.

Remark 3.1: Let $\mathfrak{H}$ be an Hilbert space. Let $\mathfrak{B}(\mathfrak{H})$, $\mathfrak{B}_1(\mathfrak{H})$ be the sets of all bounded operators and of all trace-class operators on $\mathfrak{H}$, respectively. $\mathfrak{B}_1(\mathfrak{H})$ is a closed space w.r.t. the norm $\|\cdot\|_1$ defined by

$$\|\rho\|_1 = \mathrm{Tr}_{\mathfrak{H}}(\rho^+\rho)^{\frac{1}{2}}, \quad \forall \rho \in \mathfrak{B}_1(\mathfrak{H}). \tag{3.12}$$

$\mathfrak{B}_1(\mathfrak{H})$ is isomorphic to the norm closure of the strongly continuous functionals on $\mathfrak{B}(\mathfrak{H})$ [w.r.t. the norm of $\mathfrak{B}(\mathfrak{H})^*$].

We shall sometimes denote with the same symbol the image of ρ in $\mathfrak{B}(\mathfrak{H})^*$.

Lemma 3.1: Let Q, H, be self-adjoint operators on a Hilbert space $\mathfrak{H}$ and let $Q \in \mathfrak{B}(\mathfrak{H})$. Further, let $S(\alpha, H)$, $S(\alpha, H+Q) \in \mathfrak{B}_1(\mathfrak{H})$, $\forall \alpha \in \mathbb{R}_+$, where

$$S(\alpha,H) = \exp(-\alpha H); \quad S(\alpha, H+Q) = \exp[-\alpha(H+Q)]. \tag{3.13}$$

Then it follows that:

(i) $S(\alpha; H+Q) = \sum_{n\geq 0}^{\infty} S_n^Q(\alpha, H)$ (3.14)

with

$$S_0^Q(\alpha, H) = S(\alpha, H), \tag{3.15}$$

$$S_n^Q(\alpha,H) = \int_0^\alpha d\tau S_0^Q(\alpha-\tau, H)QS_{n-1}^Q(\tau, H). \tag{3.16}$$

The convergence in Eq. (3.14) is w.r.t. the norm $\|\cdot\|_1$ defined in Eq. (3.12), $\forall \alpha \in \mathbb{R}_+$. The n integrals in Eq. (3.15) are defined as strongly Bôchner.[14]

(ii) $\forall A \in \mathfrak{B}(\mathfrak{H})$:

$$\mathrm{Tr}_{\mathfrak{H}}[AS(\alpha, H+Q)] = \sum_{n\geq 0}^{\infty} \int_0^\alpha d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-1}} d\tau_n \times \mathrm{Tr}_{\mathfrak{H}}[AS(\alpha-\tau_1, H)QS(\tau_1-\tau_2, H)\cdots QS(\tau_n, H)]. \tag{3.17}$$

The proof of the lemma can be found in Ref. 1.

Lemma 3.2: Let us consider the operators $\exp(-\beta H^{(n)}) = S(\beta, H^{(n)})$ defined in Sec. 2.

For each $k \in \mathbb{Z}_+$ we define the domains $\mathfrak{D}_k$ in C^k

$$\mathfrak{D}_K \equiv \{(z_1, \ldots, z_k) \in C^k\colon -\beta < \mathrm{Im}z_1 < \cdots < \mathrm{Im}z_k < 0\}. \tag{3.18}$$

Then $\forall A_1\cdots A_k \in \mathfrak{U}_{\Lambda_n}$, $(z_1, \ldots, z_k) \in \overline{\mathfrak{D}}_k$ the operators

$$S^{(n)}_{A_1\ldots A_k z_1\ldots z_k} = S(\beta - \tau_1, H^{(n)})A_1^{(n)}(t_1)\cdots A_k^{(n)}(t_k) \times S(\tau_k, H^{(n)}) \in \mathfrak{B}_1(\mathfrak{H}_{\Lambda_n}) \tag{3.19}$$

with

$$\tau_j = \mathrm{Im}z_j, \quad t_j = \mathrm{Re}z_j, \quad A_j^{(n)}(t_j) = \tau_{t_j}^{(n)}A_j. \tag{3.20}$$

Proof: It follows from the following equation[3]

$$\|S_{A_1\ldots A_k z_1\ldots z_k}\|_1 = \sup_{\|A\|<1} |\mathrm{Tr}_n S^{(n)}_{A_1\ldots A_k z_1\ldots z_k} A| \leq \|A_1\|\ldots\|A_k\|\,\|S(\beta, H^{(n)})\|_1. \tag{3.21}$$

Definition 3.1: For every $A_1\ldots A_k \in \mathfrak{U}_{\Lambda_n}$, $\{z_1\ldots z_k\} \in \mathfrak{D}_k$ let us define the functionals

$$\psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}(B) = \{\mathrm{Tr}_n S(\beta; H^{(n)})\}^{-1}\cdot\{\mathrm{Tr}_n S^{(n)}_{A_1\ldots A_k z_1\ldots z_k} B\},$$
$$\forall B \in \mathfrak{U}_{\Lambda_n}. \tag{3.22}$$

By Lemma 3.2 $\psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k} \in (\mathfrak{U}_{\Lambda_n})_*$ and

$$\|\psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}\|_{(\mathfrak{U}_{\Lambda_n})_*} \leq \|A_1\|\cdots\|A_k\|. \tag{3.23}$$

We point out that

$$\psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}(I) = \psi^{(n)}_{A_1\ldots A_k z_1+z, z_2+z\ldots z_k+z}(I),$$
$$\forall z \in C. \tag{3.24}$$

So that $\psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}(I)$ defines the complex functions

$$F^{(n)}_{A_1\ldots A_k}(z_1\cdots z_k), \quad \forall A_1\ldots A_k \in \mathfrak{U}_{\Lambda_n},$$
$$\{z_1\ldots z_k\} \in \overline{\mathfrak{D}}_k \supset \overline{\mathfrak{D}}_k, \text{ where}$$

$$\mathfrak{D}'_k \equiv \{(z_1\ldots z_k) \in C^k\colon \mathrm{Im}z_1 < \mathrm{Im}z_2 < \cdots < \mathrm{Im}z_k < \mathrm{Im}z_1 + \beta. \tag{3.25}$$

Remark 3.2: The unperturbed states $\phi^{(n)}$ are KMS w.r.t. $(\beta, \tau_t^{(n)})$. Then the complex functions defined by Eq. (3.24)

$$F^{(n)}_{A_1\ldots A_k}(z_1\ldots z_k) = \psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}(I),$$
$$\forall A_1\ldots A_k \in \mathfrak{U}_{\Lambda_n}, z_1\ldots z_k \in \bar{\mathfrak{D}}'_k, k \in \mathbb{Z}_+ \tag{3.26}$$

have the following properties:

(i) $$F^{(n)}_{A_1\ldots A_k}(z_1\ldots z_k) = F^{(n)}_{A_s\ldots A_k A_1\ldots A_{s-1}} \times (z_s\ldots z_k, z_{1+i}\beta\ldots z_{s-1} + i\beta). \tag{3.27}$$

(ii) $F^{(n)}_{A_1\ldots A_k}(z_1\ldots z_k)$ are analytic functions in the domain $\mathfrak{D}'_k$ and are continuous on its boundary. For $\mathrm{Im}z_n = \ldots = \mathrm{Im}z_s, \mathrm{Im}z_{j+1} = \ldots = \mathrm{Im}z_k = \mathrm{Im}z + \beta$ they have the values $\phi^{(n)}(A^{(n)}_{j+1}(t_{j+1})\cdots A_k^{(n)}(t_k)A_1^{(n)}(t_1)\cdots A_j^{(n)}(t_j))$.

(iii) $F^{(n)}_{A_1\ldots A_k}(z_1\ldots z_k)$ are analytic functions of $z_1\ldots z_s$, $s \leqslant k$ for $\{z_1\ldots z_k\} \in \bar{\mathfrak{D}}'_k$ such that

$$z_1\cdots z_s \in \mathfrak{D}'_s \quad \text{and} \quad \mathrm{Im}z_{s+1} = \cdots = \mathrm{Im}z_k < \mathrm{Im}z_1 + \beta \quad \text{or}$$
$$\mathrm{Im}z_{s+1} = \cdots = \mathrm{Im}z_k = \mathrm{Im}z_1 + \beta. \tag{3.28}$$

For instance, by Eq. (3.24),

$$F^{(n)}_{A_1\ldots A_k}(z_1\cdots z_s t_{s+1}\cdots t_k) = F^{(n)}_{A_1\ldots A_k}(z_1 + z\ldots z_s + z\ t_{s+1} + z\ldots t_k + z),$$
$$\forall z \in C, \quad t_{s+1}\cdots t_k \in \mathbb{R} \tag{3.29}$$

and the functions $F^{(n)}_{A_1\ldots A_k}(z_1\ldots z_s t_{s+1}\ldots t_k)$ are analytic in $z_1\cdots z_s$ if

$$\mathrm{Im}(z_1 + z) < \cdots < \mathrm{Im}(z_s + z) < \mathrm{Im}z < \mathrm{Im}(z_1 + z) + \beta; \tag{3.30}$$

that is,

$$-\beta < \mathrm{Im}z_1 < \cdots < \mathrm{Im}z_s < 0. \tag{3.31}$$

Remark 3.3: With the above notations we can write for every $A \in \mathfrak{U}_{\Lambda_n}$, $n > n_0$

$$\begin{aligned}\langle\phi_p^{(n)};A\rangle &= \{\mathrm{Tr}_n S(\beta, H^{(n)})\}\{\mathrm{Tr}_n S(\beta, H_p^{(n)})\}^{-1} \cdot\{\mathrm{Tr}_n S(\beta, H^{(n)})\}^{-1}\cdot\{\mathrm{Tr}_n S(\beta, H_p^{(n)})A\}\\ &= \left\{\sum_{h\geqslant 0}^{\infty}\int_0^\beta d\tau_1\cdots\int_0^{\tau_{n-1}}\psi^{(n)}_{hQ,(-i\tau_1),\ldots,(-i\tau_h)}(I)\right\}^{-1} \cdot\left\{\sum_{h\geqslant 0}^{\infty}\int_0^\beta d\tau_1\ldots\int_0^{\tau_{h-1}}d\tau_h\,\psi^{(n)}_{hQ,(-i\tau_1),\ldots,(-i\tau_h)}(A)\right\},\end{aligned} \tag{3.32}$$

where hQ means $Q\ldots Q$ h-times.

Lemma 3.3: Under D.S.I, D.S.II hypotheses, it follows that:

(i) $$\lim \psi^{(n)}_{A_1\ldots A_k z_1\ldots z_k}(B) = \psi_{A_1\ldots A_k z_1\ldots z_k}(B),$$
$$\forall A_1\ldots A_k \in \mathfrak{U}_L, \quad z_1\ldots z_k \in \bar{\mathfrak{D}}_k, \quad k \in \mathbb{Z}_+ \tag{3.33}$$

with $z_1\ldots z_s \in \mathfrak{D}_s$, $0 \leqslant s \leqslant k, s \in \mathbb{Z}_+$

The limit uniform in $z_1\ldots z_k$ on the compact sets in the domain $\mathfrak{D}_k$ and the limiting functions

$$\psi_{A_1\ldots A_k z_1\ldots z_k}(B) = F_{A_1\ldots A_k B}(z_1\cdots z_k 0) \tag{3.34}$$

are analytic in $\mathfrak{D}_s$, continuous on its boundaries, and satisfy the KMS conditions:

$$F_{A_1\ldots A_k B}(z_1\ldots z_k 0) = F_{A_h\ldots A_k B A_1\ldots A_{h-1}} (z_h\ldots z_k 0 z_{1+i}\beta\ldots z_{h-1} + i\beta). \tag{3.35}$$

(ii) The boundary value of $F_{A_1\ldots A_k B}(z_1\ldots z_k 0)$ for $\mathrm{Im}z_1 = \ldots = \mathrm{Im}z_j = -\beta, \mathrm{Im}z_{j+1} = \ldots = \mathrm{Im}z_k = 0$ is the function

$$(\Omega_\phi, (\tau_{j+1}(t_{j+1})\pi_\phi(A_{j+1}))\cdots\pi_\phi(E)\cdots(\tau_{t_j}\pi_\phi(A_j))\Omega_\phi), \tag{3.36}$$

where $t_j = \mathrm{Re}z_j$ and $\Omega_\phi, \tau_t, \pi_\phi$ are defined in Sec. 2.

(iii) $$\psi_{A_1\ldots A_k z_1\ldots z_k} \in (\pi_\phi(\mathfrak{A})'')_*. \tag{3.37}$$

The proof can be found in Ref. 3. With regards to points (ii) and (iii) we report the integral representation for

$$\begin{aligned}&F^{(n)}_{A_1A_2A_3A_4}(z_1z_2t_3t_4)F_{A_1A_2A_3A_4}(z_1z_2t_3t_4)f(z_2 - z_1) \times f(-z_1) = -\frac{1}{4\pi}\int_{-\infty}^{+\infty}dt\frac{f(t)}{t - z_2 + z_1}\\ &\times\left(\int_{-\infty}^{+\infty}dt'\frac{f(t)'}{t' + z_2}\langle\phi^{(n)}iA_1A_2^{(n)}(t)A_3^{(n)}(t_3 + t') \times A_4^{(n)}(t_4 + t')\rangle - \int_{-\infty}^{+\infty}dt'\frac{f(t' + i\beta)}{t' + i\beta + z_2}\langle\phi^{(n)}; A_3^{(n)}(t_3 + t')A_4^{(n)}(t_4 + t')A_1A_2(t)\rangle\right) + \frac{1}{4\pi^2}\\ &\times\int_{-\infty}^{+\infty}\frac{dt}{t + z_1 - 2}\left(\int_{-\infty}^{+\infty}dt'\frac{f(t - t' + i\beta)f(i\beta - t')}{t' - i\beta - z_2} \times\langle\phi^{(n)};A_2^{(n)}(t)A_3^{(n)}(t_3)A_4^{(n)}(t_4)A_1^{(n)}(t')\rangle\right.\\ &\left.-\int_{-\infty}^{+\infty}dt'\frac{f(t' - t)f(-t')}{t' - z_2}\langle\phi^{(n)}; A_1^{(n)}(t')A_2^{(n)}(t) \times A_3^{(n)}(t_3)A_4^{(n)}(t_4)\rangle\right).\end{aligned} \tag{3.38}$$

In Eq. (3.38) $f(\xi)$ is an analytic nonzero function decreasing sufficiently fast as $|\mathrm{Re}\xi| \to \infty$ for $0 < \mathrm{Im}\xi < \beta$.

By the Lebesgue theorem the lim can be performed before the integrals in Eq. 3.38, which exists for D.S.I For point (iii) it is sufficient to consider[9]

$$\mathrm{A}_m \in \pi_\phi(\mathfrak{A})'', \quad \|\mathrm{A}_m\| \leqslant |\forall n, \quad s: \lim_m \mathrm{A}_m = A.$$

The integral representation (3.38) is still valid in the thermodynamic limit and again Lebesgue theorem applies. In fact $\lim_m \tilde{\psi}_{A_1\ldots A_k z_1\ldots z_k}(\mathrm{A}_m)$ can be performed inside the integrals, and

$$\lim_m\langle\tilde{\phi}; (\tau_{t_1}\pi_\phi(A_1))\cdots(\tau_{t_k}\pi_\phi(A_k))\,A_m\rangle = \langle\tilde{\phi}; (\tau_{t_1}\pi_\phi(A_1))\cdots(\tau_{t_k}\pi_\phi(A_k))B\rangle.$$

Remark 3.4: By Lemma 3.3 the following limits exist:

$$\lim_n \psi^{(n)}_{A_1\ldots A_k z_1\ldots z_n}(B_1^{(n)}(t_1)\cdots B_s^{(n)}(t_s)),$$
$$\forall A_1\ldots A_n B_1\ldots B_s \in \mathfrak{U}_L, \quad z_1\ldots z_k \in \mathfrak{D}_k, \quad t_1\ldots t_s \in \mathbb{R}, \quad k, s \in \mathbb{Z}_+. \tag{3.39}$$

Lemma 3.4: Let us consider the operator $\hat{A}^{(n)}(t) \in \mathfrak{U}_{\Lambda_n}$:

$$\hat{A}^{(n)}(t) = \tau_p^{(n)}(t)A = U_p^{(n)}(t)A\,U_p^{(n)}(-t), \quad \forall A \in \mathfrak{U}_{\Lambda_n}, \quad t \in R.$$

It follows that

$$\lim_n \psi^{(n)}_{A_1\dots A_k z_1\dots z_k}(A^{(n)}(t)) = \lim_N \sum_{h\geq 0}^{N} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n$$

$$\lim_n \psi^{(n)}_{A_1\dots A_k z_1\dots z_k}(\eta_Q^{(n)}(t_h)\cdots\eta_Q^{(n)}(t_1)A^{(n)}(t)) \tag{3.40}$$

In Eq. (3.40) $\eta_Q^{(n)}(t): R \to \mathfrak{B}(\mathfrak{U}_{\Lambda_n})$ as

$$\eta_Q^{(n)}(t)A = i[Q^{(n)}(t)_1 A], \quad \forall A \in \mathfrak{U}_{\Lambda_n}. \tag{3.41}$$

Proof: $\hat{A}^{(n)}(t) = \tau_p^{(n)}(t)A$, may be expressed in terms of $\tau_t^{(n)}A$ by the interaction representation formula[16]

$$\tau_p^{(n)}(t)A \lim_n \sum_{h>0}^{N} \int_0^t dt_1 \dots \int_0^{t_{n-1}} dt_h\, \eta_Q^{(n)}(t_h)\dots\eta_Q^{(n)}(t_1)\tau_t^{(n)}A. \tag{3.42}$$

The integrals in Eq. (3.42) are strong limits of Riemann sums. The norm of the summand is majorized by $\|A\|(2\|Q\|\,|t|^h/h!)$. Then $\hat{A}^{(n)}(t)$ may be expressed as strong limit of a sequence of elements of $\mathfrak{U}_{\Lambda_n}$; $\mathfrak{U}_{\Lambda_n}$ is a Von Neumann Algebra, therefore strongly closed. $\psi^{(n)}_{A_1\dots A_k z_1\dots z_k}$ is strongly continuous in any bounded region of $\mathfrak{U}_{\Lambda_n}$. At last, the sequence

$$\lim_N \sum_{h\geq 0}^{N} \int_0^t dt_1 \dots \int_0^{t_{n-1}} dt_h\, \psi^{(n)}_{A_1\dots A_k z_1\dots z_k} \times (\eta_Q^{(n)}(t_h)\cdots\eta_Q^{(n)}(t_1)A(t)) \tag{3.43}$$

is uniformly convergent w.r.t. n. [the summand is majorized in modulo by $\|A\|\cdot\|\cdots\|A_k\|(2\|Q\|\,|t|^h)/h!)$].

Remark 3.5: We point out that the terms in the sequence (3.42) are in the ball of radius $r < \|A\| \exp(2\cdot\|Q\|\cdot|t|)$.

Furthermore, the product of operators belonging to the unit ball is continuous in the weak topology.[9]

Now we can express the product $\hat{B}_1(t_1)\cdots\hat{B}_k(t_k)$, as a product of k sequences of Eq. (3.42) type. Every element of this product is in the ball of radius $r < Ce^{2\cdot\|Q\|\cdot|T|}$, with $C \equiv \|B_1\|\cdots\|B_K\|$, $T = |t_1| + \cdots + |t_K|$. It follows that

$$\lim_n \psi^{(n)}_{A_1\dots A_k z_1\dots z_k}(\hat{B}_1^{(n)}(t_1)\cdots\hat{B}_s^{(n)}(t_s))$$

$$= \lim_{n_1}\dots\lim_{n_s}\sum_{h_1\geq 0}^{N_1}\cdots\sum_{h_s\geq 0}^{N_s}\int_0^{t_1} d\hat{t}_1\dots\int_0^{\hat{t}'_{h_1-1}} dt_{h_1}\dots\int_0^{t_s} d\hat{t}^s\dots$$

$$\int_0^{\hat{t}^s_{h_s-1}} d\hat{t}^s_{h_s} \lim_n \psi_{A_1\dots A_k z_1\dots z_k}(\eta_Q^{(n)}(\hat{t}'_{h_1})$$

$$\dots\eta_Q^{(n)}(\hat{t}'_1)B_1^{(n)}(t_1)\cdots\eta_Q^{(n)}(t^s_{h_s})\cdots\eta_Q^{(n)}(\hat{t}^s_1)B_s^{(n)}(t_s)),$$

$$\forall A_1\dots A_k B_1\dots B_s \in \mathfrak{U}_{\Lambda n}\, z_1\dots z_k \in \mathfrak{D}_k, \quad t_1\dots t_s \in R. \tag{3.44}$$

Remark 3.6: Let us introduce the notations

$$\sum_{h\geq 0}^{L}(\beta,\tau_1\dots\tau_n) \equiv \sum_{h\geq 0}^{L}\int_0^\beta d\tau_1\int_0^{\tau_1} d\tau_2\dots\int_0^{\tau_{h-1}} d\tau_h, \tag{3.45}$$

$$\sum_{h\geq 0}^{L}(t,t_1\dots t_h) \equiv \sum_{h\geq 0}^{L}\int_0^t dt_1\int_0^{t_1} dt_2\dots\int_0^{t_{h-1}} dt_K. \tag{3.46}$$

Then we can write

$$\lim_m \lim_n \langle\phi_p^{(n)}; \hat{B}_1^{(n)}(t_1)\dots\hat{B}_k^{(n)}(t_k)\hat{B}_{k+1}^{(m)}(t_{k+1})\cdots\hat{B}_{k+s}^{(m)}(t_{k+s})\rangle$$

$$= \lim_m \lim_n \left(\lim_L \sum_{h\geq 0}^{L}(\beta,\tau_1\dots\tau_h)\psi^{(n)}_{h\cdot Q\cdot i\tau_1\dots - i\tau_n}(I)\right)^{-1}$$

$$\times\left(\lim_L \sum_{h\geq 0}^{L}(\beta,\tau_1\dots\tau_n)\lim N_1\dots N_{K+s}\right.$$

$$\times\sum_{h_1\geq 0}^{L}(t_1,\hat{t}_1'\dots\hat{t}'_{h_1})\dots\sum_{h_{K+s}\geq 0}^{N_{K+s}}(t_{k+s},t_1^{k+s}\dots\hat{t}^{k+s}_{hs}).$$

$$\times\psi^{(n)}_{hQ\dots i\tau_h\dots i\tau_h}(\eta_Q^{(n)}(\hat{t}'_{h_1})\dots\eta_Q^{(n)}(\hat{t}'_1)B^{(n)}(t_1)$$

$$\times\dots\eta_Q^{(n)}(\hat{t}^k_{h_k})\dots\eta_Q^{(n)}(\hat{t}^k_1)B_k^{(n)}(t_k)\eta_Q^{(m)}(\hat{t}^{k+1}_{h_{k+1}})\cdots\eta_Q^{(m)}(\hat{t}^{k+1}_1)$$

$$\left.\times B^{(m)}_{k+1}(t_{k+1})\cdots\eta_Q^{(m)}(\hat{t}^{k+s}_{h_{k+1}})\cdots\eta_Q^{(m)}(\hat{t}_1^{k+s})B^{(m)}_{k+s}(t_{k+s})\right). \tag{3.47}$$

We note that

$$\lim_L\left|\sum_{h\geq 0}^{L}(\beta\tau_1\dots\tau_h)\lim_{N_1\dots N_{R+s}}\sum_{h_1\geq 0}^{N_1}(t_1\dots)\dots \times\sum_{h_{k+s}\geq 0}^{N_{k+s}}(t_{k+s}\dots)\psi^{(n)}_{hQ\dots}\text{ [as in Eq. (3.47)]}\right|$$

$$\leq \exp(\beta\|Q\|).\ \exp[2\|Q\|(t_1| + \dots + |t_{K+s}|)]. \tag{3.48}$$

That is the limits (3.48) are uniform w.r.t. n, m. It follows

$$\lim_m \lim_n\{\text{as in Eq. 3.47}\} = [(\lim_L \sum_{h>0}^{L}(\beta\tau_1\cdots\tau_h)$$

$$\times\lim_n \psi^{(n)}_{\dots}(I))^{-1}\lim_L\sum_{h\geq 0}^{L}(\beta\tau_1\dots\tau_h)$$

$$\times\lim_{N_1\dots N_{k+s}}\sum_{h_1\geq 0}^{N_1}(t_1\dots)\sum_{h_{k+s}\geq 0}^{N_{k+s}}(t_{k+s}\dots)\lim_m\lim_n\psi^{(n)}_{\dots}$$

$$\text{(as in Eq. (3.47))}]. \tag{3.49}$$

Proof of the theorem 3.1: Under D.S.I, D.S.II hypotheses, Lemma (3.3) shows that the limits in Eq. (3.47) can be performed. It follows that:

(i) $$\lim_n\langle\phi_p^{(n)};\hat{A}_1(t_1)\cdots\hat{A}_k(t_k)\rangle \quad \text{exists } \forall A_1\dots A_k \in \mathfrak{U}_L,$$
$$t_1\dots t_k \in \mathbb{R}, \quad k \in \mathbb{Z}_+. \tag{3.50}$$

(ii) $$\lim_m\lim_n\langle\phi_p^{(n)};\hat{A}_1^{(n)}(t_1)\dots\hat{A}_k^{(n)}(t_k)\hat{A}^{(m)}_{k+1}(t_{k+1})\times\dots\hat{A}^{(m)}_{k+s}(t_{k+s})\rangle$$
$$= \lim_n\langle\phi_p^{(n)};\hat{A}_1^{(n)}(t_1)\cdots\hat{A}^{(n)}_{k+s}(t_{k+s})\rangle,$$
$$\forall A_1\cdots A_{k+s} \in \mathfrak{U}_L$$
$$t_1\cdots t_{k+s} \in r, k, s \in \mathbb{Z}_+. \tag{3.51}$$

Then the D.S. results are valid for the perturbed state ϕ_p. Now we have only to prove that ϕ_p (the limit of $\phi_p^{(n)}$), belongs to $\mathcal{S}(\phi,\mathfrak{A})$, and that $U_p(t) = \exp\{i[H + \pi_\phi(Q)]t\}$; $H = \lim_{t\to 0} t^{-1}[U(t) - I]$. By Eq. (3.49), we see that

$$\phi_p(A) = C\sum_{h\geq 0}^{\infty}\int_0^\beta d\tau_1\cdots\int_0^{\tau_{h-1}} d\tau_h\,\psi_{h\cdot Q-i\tau_1\dots -i\tau_h}(A), \quad \forall A \in \mathfrak{U}_L \tag{3.52}$$

$$C = \sum_{h\geq 0}^{\infty}\int_0^\beta d\tau_1\cdots\int_0^{\tau_{h-1}} d\tau_h\,\psi_{h\cdot Q-i\tau_1\dots -i\tau_h}(I). \tag{3.53}$$

We know that $\tilde{\psi}_{hQ-i\tau_1\dots -\tau_h}$, i.e., the extension of $\psi_{hQ-i\tau_1\dots -\tau_h}$ belongs to $(\pi_\phi(\mathfrak{A})'')_*$.

But

$$|\psi_{hQ-i\tau_1\cdots-i\tau_h}(A)| \leqslant \lim_n |\psi_{hQ-i\tau_1\cdots-i\tau_h}(A)| \leqslant \|Q\|^{h}\cdot\|A\| \quad \forall A \in \mathfrak{U}_L$$

by Eq. (3.31). It follows that ϕ_p is a norm limit of elements of $(\pi_\phi(\mathfrak{A})'')_*$ w.r.t. the norm of $\mathfrak{A}^*$.

At the end we see that,

$$\lim_n \langle \phi_p^{(n)}; A\hat{B}^{(n)}(t)\rangle = \langle \tilde{\phi}_p; \pi_\phi(A)\tau_p(t)\pi_\phi(B)\rangle$$

$$\lim_n \sum_{h\geqslant 0}^{N} \int_0^t dt_1 \cdots \int_0^{t_{h-1}} dt_h \langle \tilde{\phi}_p; \pi_\phi(A)\eta_{\pi_\phi(Q)}(t_h) \cdots \eta_{\pi_\phi(Q)}(t_1)\tau_t\pi_\phi(B)\rangle, \quad (3.54)$$

so that $\tau_p(t)$ can be written as strong limit of elements of $\pi_\phi(\mathfrak{A})''$, in the form,

$$\tau_p(t) = \text{s-}\lim_n \sum_{h\geqslant 0}^{N} \int_0^t dt_1 \cdots \int_0^{t_{h-1}} dt_h \, \eta_{\pi_\phi(Q)}(t_h)\cdots\eta_{\pi_\phi(Q)}(t_1)\tau_t. \quad (3.55)$$

Theorem 3.2: We assume that the hypotheses of Theorem 3.1 hold and that the infinite volume unperturbed pressure exists.

Then the following limit

$$\lim_n |\Lambda_n|^{-1}.\ \ln \operatorname{Tr}_n S(\beta, H_p^{(n)})$$

exists and it is independent of Q.

Proof: We have

$$\lim_n |\Lambda_n|^{-1} \ln \operatorname{Tr}_n S(\beta, H_p^{(n)}) = \lim_n |\Lambda_n|^{-1} \times \ln\left[\frac{\operatorname{Tr}_n S(\beta, H_p^{(n)})}{\operatorname{Tr}_n S(\beta, H^{(n)})}\right] + \lim_n |\Lambda_n|^{-1} \ln \operatorname{Tr}_n S(\beta, H^{(n)}). \quad (3.56)$$

By Eq. (3.21) one obtains

$$\exp - \beta\|Q\| \leqslant \frac{\operatorname{Tr}_n S(\beta, H_p^{(n)})}{\operatorname{Tr}_n S(\beta, H^{(n)})} \leqslant \exp - \beta\|Q\|. \quad (3.57)$$

Therefore, the first term in the l.h.s. of Eq. (3.56) is zero for Theorem 3.1 and Eq. (3.57). The existence of the second term is in the hypotheses of the theorem.

Remark 3.7: Theorem 3.2 shows that the thermodynamic potentials like the pressure, which are additive in the volume, are uneffected by local perturbations. In this sense, therefore, the effect of the perturbations remains localized during an isothermal transformation.

ACKNOWLEDGMENTS

We are indebted to G. L. Sewell for useful conversations.

[1]D. A. Uhlenbrock, J. Math. Phys. **12**, 2505 (1972).
[2]D. A. Uhlenbrock, J. Math. Phys. **13**, 993 (1972).
[3]J. Ginibre and C. Gruber, Commun. Math. Phys. **11**, 198 (1969).
[4]H. Araki, Publ. RIMS Kyoto Univ. Ser. A. **4**, 361 (1968).
[5]D. Dubin and G. L. Sewell J. Math. Phys. **11**, 2990 (1970).
[6]C. Marchioro, E. Presutti, and E. Scacciatelli, Physica (Utr.) **57**, 79 (1972).
[7]M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras and Its Applications* (Springer, New York, 1970).
[8]E. Presutti, E. Scacciatelli, G. L. Sewell, and F. Wanderling, J. Math. Phys. **13**, 1085 (1972).
[9]J. Dixmier, *Les C*-algebras et leurs répresentations* (Gauthier-Villars, Paris, 1964); J. Dixmier, *Les algebras d'operateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1960).
[10]R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **6**, 189 (1967).
[11]M. Winnink, *Cargese Lectures 1969*, edited by D. Kastler.
[12]D. Ruelle, Lecture Notes, Summer School of Theoretical Physics, Cargese, Corsica (1965)
[13]From now on we drop β, μ from the symbols representing Gibbs states.
[14]E. Hille and R. S. Phyllips, *Functional Analysis and Semigroups* (Am. Math. Soc., Providence, R. I., 1957).
[15]M. Guenin, Commun. Math. Phys. **3**, 120 (1966).

Mathematical theory of the R matrix. I. The eigenvalue problem*

F. J. Narcowich

Texas A&M University, Department of Mathematics, College Station, Texas 77840
(Received 17 May 1973)

This is the first paper in a two part series aimed at placing the theory of Wigner's R matrix on a mathematically rigorous footing. In Paper I of the series, we will show that the eigenvalue problem associated with the R matrix can be solved for a large class of potentials, including Coulomb-like potentials. We will do this for the case in which the boundary of the internal region is a smooth surface—although the results remain true for a much larger class of surfaces. In Paper II of the series, we will show that the R matrix exists for the class of potentials mentioned, is a compact operator, and can be approximated uniformly (i.e., normwise) by the usual expansions associated with the R matrix.

I. INTRODUCTION

A. Origin of the problem

In a series of papers,[1-3] in 1946 and 1947, E.P. Wigner and L. Eisenbud introduced the concept of the "reaction matrix" or "R matrix" to calculate cross sections of nuclear reactions near resonance. The main object of these calculations was to justify the Breit-Wigner "one-level formula"[4] for the decay of a compound nucleus using as few assumptions about the nuclear potential as possible.

The basic idea behind these calculations is really very simple, although the calculations themselves are rather complicated. Consider a system of N spinless particles (spin merely complicates the argument and adds nothing essentially new) interacting via some potential V. The configuration space of this system (E^{3N}) is divided into two regions: a bounded region I (the *internal* region) and its complement, $E^{3N} - I$ (the *external* region). The region I is chosen to enclose the center of mass of the system and physically represents a region in configuration space where all of the particles interact via nuclear forces. It we restrict the collision process to low enough energies, the reaction products will be a pair of nuclear fragments which are essentially free when they are far enough apart to be outside of the internal region. By consistently matching the normal derivative and the value of the wavefunction on the surface of the internal region with the same quantities from the external region, one can obtain the solution to Schrödinger's equation in the external region and, for large separation distance, the asymptotic form of the wavefunction.

The value of the stationary wavefunction of energy E and its normal derivative on S, the surface of the internal region, are not, however, independent quantities. Wigner and Eisenbud heuristically constructed an operator $R(E)$ which takes the normal derivative of the wavefunction on S into the value of the wavefunction on S. By specifying different values of the normal derivative, we get different asymptotic states. Thus, in effect, the operator $R(E)$, (the R matrix) is supplying the same information as the collision matrix. In fact, Wigner and Eisenbud calculated the collision matrix from the R matrix and, from the collision matrix, the cross section for the reaction.[3]

Unfortunately, the derivation and expansions used for the R matrix were completely formal and, except in the trivial case of one dimension, it was never proved that the various expansions converged. Moreover, no conditions were placed on the potential V or the surface S, thus leaving the eigenvalue problem itself open to difficulties.

B. Statement of the problem

The eigenvalue problem introduced by Wigner and Eisenbud in conjunction with the R matrix is a variation of the Neumann problem. We are required to find a complete set of orthonormal functions spanning $L^2(I)$ and satisfying the following in I:

$$-\sum_{i,j} \frac{\partial}{\partial x_i} A_{ij} \frac{\partial}{\partial x_j} u_l + V u_l = E_l u_l \tag{1}$$

and

$$\partial_A u_l = \sum_{i,j} n_i A_{ij} \frac{\partial u_l}{\partial x_j} = 0 \tag{2}$$

on S. The A_{ij} are components of an $n \times n$ Hermitian, positive definite matrix and, in the usual scattering problem, are constants. If the center of mass has not been separated out, the matrix is diagonal. On the other hand, when the center of mass has been separated out, off-diagonal terms (the so-called Hughes-Eckart or specific mass corrections[5]) arise. The n_i are components of the unit outward normal to S and, finally, V is some potential.

Roughly speaking, if the surface S, the potential V, and the matrix A_{ij} (whose elements we allow to depend on x) are smooth and, in addition, $A_{ij}(x)$ is uniformly positive definite, then it is known that the solution to the eigenvalue problem exists and that the eigenfunctions are smooth and satisfy the boundary conditions pointwise.[6,7,8] Our purpose in writing this paper is twofold: First, we wish to show that the regularity assumptions on V can be weakened considerably. Second, we wish to provide a framework for the construction of the R matrix and the discussion of its properties which we will give in a followup paper.

This paper is organized as follows: In Sec. II, we present a short review of the Sobolev theory. This section is entirely expository and is included for the convenience of the reader. In Sec. III, using some results of Schechter,[6,9] we first give a rigorous discussion of the eigenvalue problem with $V = 0$. By using a theorem of Kato,[10] we then show that the eigenvalue problem can be solved for a class of potentials which are "Kato small"[5] in comparison to the operator associated with $V = 0$. We then specialize these results to a class of potentials which we call "R-admissible." This is the natural class of potentials for which the R matrix is defined. We then give results involving eigenfunction expansions coming from R-admissible poten-

Copyright © 1974 American Institute of Physics

tials. In an appendix, we show that potentials with Coulomb-like singularities are R-admissible.

II. SOBOLEV SPACES

For the convenience of the reader, this section will include a summary of certain results from the theory of Sobolev spaces.

A. Notation

In all that follows, the symbol I denotes a bounded, open region of Euclidean n-space, E^n. The boundary (surface) of I, S, is an infinitely differentiable, $(n-1)$-dimensional manifold.[11,12] S is orientable and I lies entirely on one side of S. The inner products in $L^2(I)$ and $L^2(S)$ are denoted by $(\ ,\ )_I$ and $(\ ,\ )_S$, respectively; the norms are denoted by $\|\ \|_I$, and $\|\ \|_S$, again, respectively. Inner products are conjugate linear in the left and linear in the right. As is customary, $C_0^\infty(I)$ denotes the set of all functions infinitely differentiable in I and vanishing outside of some closed set contained in I. $C^\infty(\bar{I})$ denotes the set of all functions infinitely differentiable in $\bar{I}$, the closure of I.

For derivatives, we will use the standard multi-index notation: Namely, let $D_i = \partial/\partial x_i$ and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be nonnegative integers. We denote

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n},$$

where $\alpha = \{\alpha_1, \ldots, \alpha_n\}$, $|\alpha| = \sum \alpha_i$.

B. Generalized derivatives and the spaces $W_2^l(I)$

We will more or less follow Sobolev[13] or Agmon[14] in our discussion. We begin with the concept of generalized derivative.

Definition (Generalized derivative): Let $f \in L^2(I)$. If there exists $h \in L^2(I)$ such that

$$(f, (-1)^{|\alpha|} D^\alpha g)_I = (h, g)_I$$

for all $g \in C_0^\infty(I)$, then h is called a generalized derivative of f and we write

$$h = D^\alpha f.$$

Moreover, we denote the set of all $f \in L^2(I)$ having all generalized derivatives of order $j \le l$ by $W_2^l(I)$.

A generalized derivative, if it exists, is unique (modulo sets of measure zero in I). This follows directly from its definition coupled with the fact that $C_0^\infty(I)$ is dense in $L^2(I)$. In addition, if a function f is l times continuously differentiable in $\bar{I}$, then f has all generalized derivatives of order $j \le l$ and these agree with the ordinary derivatives (modulo sets of measure zero). Consequently, $C^\infty(\bar{I})$ is a subset of $W_2^l(I)$ for all l.

Obviously, generalized derivatives are densely defined linear operators on $L^2(I)$. In particular, $W_2^l(I)$ is a linear manifold in $L^2(I)$. We can define an inner product and norm for functions in $W_2^l(I)$ as follows: Let $f, g \in W_2^l(I)$; then set

$$[f,g]_{l,I} = \sum_{|\alpha| \le l} (D^\alpha f, D^\alpha g)_I + (f,g)_I,$$

$$|f|_{l,I} = ([f,f]_{l,I})^{1/2}.$$

With this norm and inner product, $W_2^l(I)$ becomes a Hilbert space in its own right (see Agmon,[14] p. 4).

All that we have said so far is independent of the fact that I is a bounded region. In fact, precisely the same statements hold for an arbitrary open region in E^n, including E^n itself. The following lemma, which we will use later, is an important "density" result for the space $W_2^l(E^n)$.

Lemma II.1: The set of all functions infinitely differentiable and vanishing outside some closed bounded region, $C_0^\infty(E^n)$, is dense in $W_2^l(E^n)$.

Proof: See Lions and Magenes,[15] p. 37.

The following result, due to Calderón, provides an intimate connection between $W_2^l(I)$ and $W_2^l(E^n)$:

Lemma II.2 (Calderón extension theorem): There exists a bounded linear transformation T of $W_2^l(I)$ into $W_2^l(E^n)$ such that if $u \in W_2^l(I)$, and $\hat{u} = Tu$, the restriction of $\hat{u}$ to I, $\hat{u}|_I$, coincides with u. That is, $\hat{u}|_I = u$.

Proof: See Agmon,[14] p. 171 or, for a more general statement and proof, see Calderón,[16] Theorem 12.

We will use both of these results in proving that Coulomb-like potentials belong to the class of R-admissible potentials (see Sec. III and Appendix B). Also, the combination of these two lemmas gives a very important density theorem for $W_2^l(I)$.

Theorem II.1. $C^\infty(\bar{I})$ is dense in $W_2^l(I)$.

Proof: See Lions and Magenes,[15] p. 44.

Much of the great utility that the Sobolev spaces enjoy comes from the fact that sets which are bounded in the $|\ |_{l,I}$ norm are relatively compact in the space $W_2^j(I)$ for all $j < l$. That is, if B is a bounded subset of $W_2^l(I)$, then every sequence which may be extracted from B has a subsequence which is actually convergent in $W_2^j(I)$ for all $j < l$. We state this theorem, which is due to Rellich, and refer the reader to Agmon,[14] p. 30, for the proof.

Theorem II.2 (Rellich): Every bounded set in $W_2^l(I)$ is relatively compact in $W_2^j(I)$ if $j < l$.

We remark that this theorem is false if I is replaced by E^n. It is true, however, that the theorem holds for a much larger class of regions than the one I belongs to. For a discussion and more references, see Lions and Magenes,[15] p. 111.

We close this rather terse section with an inequality which will be of some value to us.

Theorem II.3 (Interpolation theorem): Let ϵ be a positive real number such that $0 < \epsilon \le 1$. If $u \in W_2^l(I)$ for some $l \ge 2$, and if $1 \le j \le l-1$, then

$$|u|_{j,I}^2 \le \gamma(\epsilon^{l-j} |u|_{l,I}^2 + \epsilon^{-j} \|u\|_I^2),$$

where $\gamma = \gamma(I, l)$ depends only on I and l.

Proof: See Agmon,[14] p. 24.

C. The trace of a function in $W_2^l(I)$. Intregration by parts

In boundary value problems, we must be able to define various functions on the surface S. For an arbitrary function in $L^2(I)$, this is an impossibility because the boundary is a set of measure zero in $\bar{I}$. It is, however, possible to make sense out of such a definition if the functions belong to $W_2^1(I)$. This is done as follows: First, we define an operator τ, the trace operator (see Ref. 15), by setting

$$\tau f = f|_S \ (= \text{restriction of } f \text{ to } S)$$

for all $f \in C^\infty(\bar{I})$. By Theorem 1, we can approximate any $u \in W_2^1(I)$ in the norm $|\ |_{1,I}$ by a sequence of functions in $C^\infty(\bar{I})$. The trace is then extended to all functions in $W_2^1(I)$ by taking limits. The following theorem, which is a version of a theorem of Sobolev,[13] guarantees that the trace is well-defined for all $f \in W_2^1(I)$.

Theorem II. 4: For all $f \in C^\infty(\bar{I})$ and every $\epsilon > 0$, τf satisfies

$$\|\tau f\|_S \leq \epsilon |f|_1 + C(\epsilon)\|f\|_I,$$

where $C(\epsilon)$ depends on I and ϵ but not on f. Hence, τ may be extended to all functions in $W_2^1(I)$ and, when so extended, is a compact map from $W_2^1(\bar{I})$ into a dense subset of $L^2(S)$.

Proof: The inequality may be found in Ladyzhenskaya and Utral'tseva,[7] p. 49. Other proofs and versions of this theorem are given in Sobolev,[13] p. 85; Lions and Magenes,[15] p. 44; Agmon,[14] p. 38. In Sobolev's work, the compactness of the operator is proved via estimates on integral kernels. The compactness also follows directly from the inequality given below. The only portion of the theorem which requires comment is the claim the $W_2^1(I)$ is mapped by τ into a dense subset of $L^2(S)$. This follows from the version of the theorem in Lions and Magenes.[15] They prove that τ fills the space [$H^{1/2}(S)$ in their notation, $W_2^{1/2}(S)$ in ours] which amounts to a "half-order" Sobolev space. On p. 40, they show that this space includes the set of all functions infinitely differentiable on S and that this latter space is dense in $L^2(S)$. It immediately follows that the range of τ is dense in $L^2(S)$. QED

Using Theorem II. 4, various formulas involving integration by parts may be justified. For example, if $u, v \in W_2^1(I)$,

$$\int_I \frac{\partial \bar{u}}{\partial x_i} v d^n x = -\int_I \bar{u} \frac{\partial v}{\partial x_i} d^n x + \int_S dS n_i \overline{\tau u} \tau v.$$

This is established by first using the corresponding formula for functions in $C^\infty(I)$ and then taking limits (see Smirnov,[17] p. 337).

So far we have avoided the question of any intrinsic meaning for the trace. Such a meaning does exist: Given any function f in $W_2^1(I)$, there exists a function $\tilde{f}$, equal almost everywhere to f, such that $\tilde{f}$ is defined on S and on surfaces "parallel to S". For example, in the case of S being the unit sphere, the parallel surfaces are concentric spheres with radii less than 1. Moreover, if the parallel surface S_p is close to S, then the difference between $\tilde{f}|_{S_p}$ and $\tilde{f}|_S$ will be small. Again using S as the unit sphere in E^n, we have

$$\| \tilde{f}_{r=\lambda}| - \tilde{f}_{r=1}| \|_S \to 0$$

as $\lambda \to 1$ from below. The trace of f is simply the restriction of $\tilde{f}$ to S. For a more complete discussion of this topic, see Sobolev,[13] p. 85, Lions and Magenes,[15] p. 205, and, Narcowich,[18] p. 28.

As a comprehensive reference for the spaces we have been discussing, Lions and Magenes[15] contains the most material. The work of Smirnov[17] is less comprehensive, but also tends to be less abstract. Of course, the original work on the subject is done in Sobolev's book,[13] Agmon[14] and Ladyszhenskaya-Ural'tseva[7] provide a quick overview. The latter work discusses more general surfaces than those with which we are working.

III. THE EIGENVALUE PROBLEM

In the first section of this chapter, we will discuss various aspects of the eigenvalue problem associated with $V = 0$. This is done primarily to collect and make firm the information that is found in various references. We will heavily rely on some results of Schechter,[6,9] which we collect in Appendix A. In the second section, we will apply the results we obtain in the first section to the case in which V is nonzero.

A. The case $V = 0$

We place three conditions on the partial differential operator Q,

$$Q = -\sum_{i,j} D_i A_{ij}(x) D_j,$$

which is associated with the eigenvalue problem given in Sec. I. These three conditions are

(C1) $A_{ij}(x) \in C^\infty(\bar{I})$,

(C2) $\overline{A_{ij}(x)} = A_{ji}(x)$,

(C3) $\mu_0 |\xi|^2 \leq \sum_{i,j} \overline{\xi_i} A_{ij}(x) \xi_j \leq \mu_1 |\xi|^2$,

where μ_0, μ_1 are positive constants and ξ is an arbitrary n-component complex valued vector with norm $|\xi|$. (C1), (C2), and (C3) hold for all $x \in \bar{I}$. In the rest of the paper, we consider the derivatives in Q to be generalized and allow Q to operate on any function in $W_2^2(I)$. This makes sense because only second order derivatives appear in Q. We now define an operator H_0 as follows:

(1) $H_0 = Q|_{D(H_0)}$,

where $D(H_0)$ is the set of all $f \in W_2^2(I)$ such that

(2) $\partial_A f \equiv \sum_{i,j} n_i A_{ij}(x) D_j f = 0$ on S,

where (2) holds in the sense of the trace (see Sec. II).

The main theorem of this section, which is more or less a collection of results which are known, is

Theorem III. 1: The operator H_0 is positive and self-adjoint. Given any E not in the spectrum of H_0, the operator $T(E) = (H_0 - E)^{-1}$ maps $W_2^l(I)$ into $W_2^{l+2}(I)$ for all $l \geq 0$. In particular, $T(E)$ maps $C^\infty(\bar{I})$ into $C^\infty(\bar{I})$. Also, $T(E)$ maps $L^2(I)$ compactly into $W_2^1(I)$ and, hence, compactly into $L^2(I)$. The spectrum of H_0 consists of countably many nonnegative eigenvalues with $+\infty$ being the only limit point. The eigenfunctions of H_0 are in $C^\infty(\bar{I})$ and pointwise satisfy the boundary conditions.

We will postpone the proof of Theorem III. 1. The content of Theorem III. 1 is very simple: The eigenvalue problem associated with the R matrix for $V = 0$ is classically solvable and the operator $(H_0 - E)^{-1}$ has a smoothing effect on functions.

In the proof of Theorem III. 1, we will need the following lemmas.

Lemma III.1: Q is properly elliptic and the boundary operator $\partial_A = \sum_{i,j} n_i A_{ij}(x) D_j$ covers Q.

Proof: The result follows directly from the definitions given in Appendix A coupled with properties (C_1), (C_2), and (C_3). QED

Lemma III.2: Let $f \in C^\infty(\bar{I})$ and suppose $\partial_A f = 0$ on S. Then the following inequalities hold:

(3a) $(f, Qf)_I \geq 0$,

(3b) $\| (Q+1)f \|_I \geq \| f \|_I$,

(3c) $\| (Q+1)f \|_I \geq \| Qf \|_I$,

(3d) $|f|_{2,I} \leq K(\| Qf \|_I + \| f \|_I)$,

where K is a constant which depends only on I, ∂_A, and Q.

Proof: By an easy integration by parts, we have

(4) $$(u, Qv)_I = \sum_{i,j} (D_i u, A_{ij} D_j v)_I - (u, \partial_A v)_S.$$

Setting $u = v = f$, $\partial_A f = 0$ in (4) and using (C3), we obtain (3a). To obtain (3b) and (3c), we need only use

$$\| (Q+1)f \|_I^2 = \| Qf \|_I^2 + 2(Qf, f)_I + \| f \|^2$$

plus (3a). Finally, (3d) follows from Lemmas III. 1 and A.1.

The next two lemmas concern the Hermitian form defined by

(5) $$\langle u, v \rangle \equiv \sum_{i,j} (D_i u, A_{ij} D_j v)_I + (u, v)_I,$$

which is obviously defined for all u, v in $W_2^1(I)$. Corresponding to (5), we define

(6) $$\langle u \rangle = (\langle u, u \rangle)^{1/2}.$$

Lemma III.3: With $\langle\ ,\ \rangle$ as the inner product and $\langle\ \rangle$ as the norm, $W_2^1(I)$ is again a Hilbert space. Moreover, there exists a constant $C > 0$ such that

(7) $$C^{-1} |f|_{1,I} \leq \langle f \rangle \leq C |f|_{1,I}$$

That is, the norm $\langle\ \rangle$ is equivalent to the norm $|\ |_{1,I}$.

Proof: The form $\langle\ ,\ \rangle$ satisfies all the algebraic axioms of an inner product for $W_2^1(I)$. The only properties that require any proof are the completeness of $W_2^1(I)$ in the norm $\langle\ \rangle$ and the equivalence of $\langle\ \rangle$ and $|\ |_{1,I}$—i.e., the inequality (7). If (7) holds however, a Cauchy sequence v_k in $\langle\ \rangle$ norm is also a Cauchy sequence in $|\ |_{1,I}$ norm. Since it is known that $W_2^1(I)$ is complete in the $|\ |_{1,I}$ norm, v_k must converge in $|\ |_{1,I}$ norm to a function $v \in W_2^1(I)$. But then, applying (7) to $v_k - v$, we have

$$\langle v_k - v \rangle \leq C |v - v_k|_{1,I}.$$

It is obvious that v_k converges to v in the $\langle\ \rangle$ norm. Hence, if (7) holds, completeness is assured.

To obtain (7), we begin by using property (C3) with $\xi_i = D_i f$, $f \in W_2^1(I)$. This gives

$$\mu_0 \sum_i |D_i f|^2 \leq \sum_{i,j} D_i \bar{f} A_{ij} D_j f \leq \mu_1 \sum_j |D_i f|^2.$$

Setting $C^2 = \max(\mu_1, \mu_0^{-1})$ and observing that $C^2 > 1$, we obtain

$$C^{-2}\Big(\sum_i |D_i f|^2 + |f|^2\Big) \leq \sum_{i,j} D_i \bar{f} A_{ij} D_i f + |f|^2 \leq C^2\Big(\sum_i |D_i f|^2 + |f|^2\Big).$$

Integrating this last inequality over all x in I gives

$$C^{-2}|f|_{1,I}^2 \leq \langle f, f \rangle \leq C^2 |f|_{1,I}^2,$$

from which (7) follows trivially. QED

Lemma III.4: Given any $f \in L^2(I)$, there exists a unique $u \in W_2^1(I)$ such that

(8) $$\langle u, v \rangle = (f, v)_I$$

for all $v \in W_2^1(I)$. If $f \in W_2^l(I)$, then $u \in W_2^{2+l}(I)$ for all $l \geq 0$. If $f \in C^\infty(\bar{I})$, then $u \in C^\infty(\bar{I})$. Moreover, $u \in D(H_0)$ and $(H_0 + 1)u = f$ if and only if (8) holds.

Proof: For fixed f, we have

$$|(f, v)_I| \leq \| f \|_I \| v \|_I \leq C \| f \|_I \langle v \rangle.$$

Hence, the inner product $(f, v)_I$ is a bounded linear functional on $W_2^1(I)$. By Lemma III. 3 and the Riesz-representation theorem (see Riesz-Nagy,[19] p. 61), there exists a vector $u \in W_2^1(I)$ such that

$$\langle u, v \rangle = (f, v)_I$$

for all $v \in W_2^1(I)$. To see that u is unique, suppose that $\tilde{u}$ is any vector satisfying

$$\langle \tilde{u}, v \rangle = (f, v)_I$$

for all $v \in W_2^1(I)$. Then, by subtracting this equation from the last,

$$\langle \tilde{u} - u, v \rangle = 0.$$

Hence, $\tilde{u} - u$ is orthogonal to all of $W_2^1(I)$ and $\tilde{u} - u = 0$, whence u is unique.

If $f \in W_2^l(I)$, Lemma A.3 immediately implies that $u \in W_2^{2+l}(I)$, provided $l \geq 1$. Also, if $f \in C^\infty(\bar{I})$, then Lemma A. 3 implies $u \in C^\infty(\bar{I})$. When f is in $L^2(I)$, we must resort to another tactic because $\langle u, v \rangle$ is not defined for $v \in L^2(I)$ and Lemma A. 3 does not apply.

First of all, if $f \in C^\infty(\bar{I})$, we have already seen $u \in C^\infty(\bar{I})$. We may integrate the inner product $\langle u, v \rangle$ by parts (see Sec. IIC) to obtain

(*) $$((Q+1)u, v)_I + (\partial_A u, v)_S = (f, v)_I$$

for all $v \in W_2^1(I)$. By picking $v \in C_0^\infty(I)$, the surface term vanishes and we must have that $(Q+1)u = f$, since the last equation holds with zero surface term for the dense set [in $L^2(I)$] $C_0^\infty(I)$. But then, using $(Q+1)u = f$, we have

$$((Q+1)u, v)_I = (f, v)_I$$

for all $v \in W_2^1(I)$ and the surface term in (*) vanishes:

$$(\partial_A u, v)_S = 0.$$

By Theorem II. 4, the trace is dense in $L^2(S)$. Hence, $\partial_A u$ is orthogonal to a dense set and we must have that $\partial_A u = 0$. Thus, if $f \in C^\infty(\bar{I})$, $u \in D(H_0)$.

In general, if $f \in L^2(I)$, there exists a sequence $f_k \in C^\infty(\bar{I})$ which tends to f in $L^2(I)$. For each f_k, we have a unique $u_k \in C^\infty(\bar{I})$ such that $\partial_A u_k = 0$ on S. By inequalities (3b), (3c), and (3d),

$$|u_k - u_l|_{2,I} \le 2K(\|(Q+1)(u_k - u_l)\|_I)$$

since $f_k = (Q+1)u_k$, this implies

$$|u_k - u_l|_{2,I} \le 2K(\|f_k - f_l\|_I).$$

But f_k is a convergent sequence in $L^2(I)$ and hence Cauchy. The last inequality then tells us that u_k is a Cauchy sequence in $W_2^2(I)$. Since this space is a Hilbert space, the sequence u_k is convergent in $W_2^2(I)$ to some vector $u \in W_2^2(I)$. Taking limits in the equation

$$\langle u_k, v\rangle = (f_k, v)_I$$

gives

$$\langle u, v\rangle = (f, v)_I$$

for all $v \in W_2^1(I)$. Hence, $u \in W_2^2(I)$ even when $f \in L^2(I)$.

Since $u \in W_2^2(I)$, an integration by parts coupled with a repetition of the argument used in the $C^\infty(\bar{I})$ case gives $\partial_A u = 0$ on S (in the sense of trace) and $(Q+1)u = f$. Hence, by definition of $D(H_0)$, $u \in D(H_0)$ and

$$(H_0 + 1)u = f.$$

Conversely, if $(H_0+1)u = f$, an integration by parts shows (8) must hold for all $v \in W_2^1(I)$. QED

We are now ready to prove Theorem III. 1.

Proof of Theorem III. 1: An integration by parts coupled with property (C3) of Q shows that H_0 is both symmetric and nonnegative. By Lemma III. 4, the range of $H_0 + 1$ consists of all $L^2(I)$. Since any symmetric operator whose range coincides with the whole space is self-adjoint (see Naimark,[20] p. 103), $H_0 + 1$ and, hence, H_0 are self-adjoint.

Given any E not in the spectrum of H_0, $T(E)$ is a bounded map from $L^2(I)$ to $L^2(I)$. We wish to show that if $f \in W_2^l(I)$, then $T(E)f \in W_2^{l+2}(I)$.

Set $u = T(E)f$. Then $(H_0 - E)u = f$ and Lemma III. 4 implies

$$((H_0+1)u, v)_I = \langle u, v\rangle = ((E+1)u + f, v)_I$$

for all $v \in W_2^1(I)$. We will use induction on l to show $f \in W_2^l(I)$ implies $u \in W_2^{l+2}(I)$.

If $l = 0$, Lemma III. 4 insures that $u \in W_2^2(I)$. Suppose that, for $l = k$, $f \in W_2^k(I)$ implies $u \in W_2^{k+2}(I)$. To complete the induction proof, we must show that $f \in W_2^{k+1}(I)$ implies that $u \in W_2^{k+3}(I)$. Clearly, $f \in W_2^{k+1}(I)$ implies $f \in W_2^k(I)$ and, by hypothesis, $u \in W_2^{k+2}(I)$. But then both u and f belong to $W_2^{k+1}(I)$ and, hence, so does $(E+1)u + f$. By Lemma III. 4, we immediately have that $u \in W_2^{k+3}(I)$, which completes the induction proof.

Next, we wish to show that $T(E)$ is a compact map from $L^2(I)$ to $W_2^1(I)$. Again let $u = T(E)f$, $(H_0 - E)u = f$. By taking limits in (3d), the inequality given there holds for u. That is,

$$|u|_{2,I} \le K(\|H_0 u\|_I + \|u\|_I). \tag{9}$$

Using the fact that $T(E)$ is bounded and $(H_0 - E)u = f$, this last inequality gives

$$|u|_{2,I} \le K'\|f\|_I,$$

where K' depends on E, but not f. Hence $T(E)$ maps bounded sets in $L^2(I)$ into bounded sets in $W_2^2(I)$. By Theorem II. 2, a bounded set in $W_2^2(I)$ is relatively compact in $W_2^1(I)$ and $L^2(I)$. Thus $T(E)$ maps $L^2(I)$ compactly into $W_2^1(I)$ and $L^2(I)$.

Since $T(E) = (H_0 - E)^{-1}$ maps $L^2(I)$ into $L^2(I)$ compactly, its spectrum consists of countably many eigenvalues with 0 as the only limit point (see Ringrose,[21] p. 51). Hence, $T(E)^{-1} + E = H_0$ has a spectrum consisting of countably many nonnegative eigenvalues with $+\infty$ as the only limit point.

We conclude by proving that any eigenfunction u_j of H_0 is actually in $C^\infty(\bar{I})$ and classically solves the eigenvalue problem. u_j being an eigenfunction of H_0 implies that

$$T(E)u_j = (E_j - E)^{-1}u_j,$$

where E_j is the eigenvalue corresponding to u_j. From the start, we know $u_j \in W_2^2(I)$. But then $T(E)u_j$ and hence u_j itself belong to $W_2^4(I)$. Applying the result again tells us that $u_j \in W_2^6(I)$ etc. Continuing in this way, we see that $u_j \in W_2^l(I)$ for all $l \ge 0$. By Lemma A. 2 (Sobolev's lemma), $u_j \in C^\infty(\bar{I})$. Finally, the trace of a function in $C^\infty(\bar{I})$ is simply the restriction of the function to S. Hence $\partial_A u_j = 0$ pointwise on S. Thus u_j classically solves the eigenvalue problem. QED

B. The case $V \ne 0$

The approach we take in this section is to add a suitably restricted potential V to H_0 using a theorem of Kato and Rellich. By some simple arguments, we then obtain that not only is $H_0 + V$ self-adjoint on $D(H_0)$, but that its spectrum is discrete. We then restrict our attention to a class of potentials, which we call R-admissible, which will turn out to be the natural class of potentials for which the R matrix is defined. For such potentials, $H_0 + V$ is bounded below and the completion of the Hermitian form $((H_0 + V + \lambda + 1)u, v)_I$, where λ is a certain positive constant, induces an inner product on $W_2^1(I)$ whose associated norm is equivalent to the usual norm in $W_2^1(I)$. We conclude with a result involving the expansion of functions in terms of the eigenfunctions of $H_0 + V$, where V is R-admissible.

Lemma III. 5 (Kato-Rellich theorem): Let A be self-adjoint and let B be an Hermitian operation obeying:

(a) $D(B) \supset D(A)$,

(b) There is an $a < 1$ and $b > 0$ such that

$$\|B\psi\| \le a\|A\psi\| + b\|\psi\|$$

for all $\psi \in D(A)$. Then $A + B$ defined on $D(A)$ is self-adjoint.

Proof: See Kato,[10] pp. 287-89.

An Hermitian operator B satisfying (a) and (b) with respect to a self-adjoint operator A is said to be *Kato-small* with respect to A (see Simon,[5] p. 206).

Theorem III. 2: Let V be an Hermitian operator which is Kato-small with respect to H_0. Then the operator

$$H = H_0 + V, \quad D(H) = D(H_0)$$

is self-adjoint and the spectrum of H consists of countably many eigenvalues with $\pm\infty$ being the only limit points. Moreover, the operator $(H-E)^{-1}$ maps $L^2(I)$ compactly into $W_2^1(I)$ and, hence, $L^2(I)$ for any complex number E not in the spectrum of H.

Proof: The self-adjointness follows directly from Lemma III. 5. We need only address our attention to the compactness of $(H-E)^{-1}$ and the discreteness of the spectrum of H.

Suppose that E is not in the spectrum of H, then $(H-E)^{-1}$ exists and is a bounded operator on $L^2(I)$. Given any $f \in L^2(I)$, let $v = (H-E)^{-1}f$. Applying $H_0 + V - E = H - E$ to v and rearranging terms, we have

$$H_0 v = f - Vv + Ev.$$

Because V is Kato-small with respect to H_0,

$$\| Vv \|_I \le a \| H_0 v \| + b \| v \|,$$

where $0 < a < 1$ and $b > 0$. Applying the triangle inequality to the expression for $H_0 v$ and using the last inequality, we have that

$$\| H_0 v \|_I \le [1/(1-a)][\| f \|_I + (b + |E|)\| v \|_I].$$

Using this last inequality coupled with the boundedness of $(H-E)^{-1}$, we have

$$\| H_0 v \|_I + \| v \|_I \le C \| f \|_I,$$

where C is a constant depending on a, b, and E, but not v or f. By inequality (9), we have that

$$| v |_{2,I} \le K(\| H_0 v \|_I + \| v \|_I)$$

and, hence,

$$(10) \quad | v |_{2,I} \le KC \| f \|_I.$$

Hence, by (10), $(H-E)^{-1}$ maps a set bounded in $L^2(I)$ into a set bounded in $W_2^2(I)$. By Theorem II. 2, any bounded set in $W_2^2(I)$ is relatively compact in $W_2^1(I)$ and, hence, in $L^2(I)$. Be definition, $(H-E)^{-1}$ maps $L^2(I)$ compactly into $W_2^1(I)$ and $L^2(I)$.

The discreteness of the spectrum of H is simply a repetition of the argument used to prove the discreteness of the spectrum of H_0. QED

For the purpose of the R matrix, the class of potentials which are Kato-small with respect to H_0 is too broad. We now define a restricted, but physically interesting class of potentials.

Definition (R-admissible operators): Let V be an Hermitian operator with $D(V) \subset L^2(I)$. V is said to be *R-admissible* if $D(V) \supset W_2^1(I)$ and if, for all $f \in W_2^1(I)$,

$$(11) \quad \| Vf \|_I \le M | f |_{1,I},$$

where M is a constant which is independent of f.

Clearly, any bounded Hermitian operator on $L^2(I)$ is R-admissible. In Appendix B, we will show that the many-particle coulomb potential is also R-admissible—along with Coulomb-like potentials (the Yukawa potential, for example).

We now show that every R-admissible potential is Kato-small with respect to H_0 and, hence, that Theorem III. 2 holds for such potentials.

Corollary III. 1: Let V be any R-admissible potential. Then V is Kato-small with respect to H_0 and, hence, Theorem III. 2 holds for such potentials.

Proof: Let $u \in D(H_0)$. Then, by Theorem III. 1, $u \in W_2^2(I)$ and, hence, $u \in W_2^1(I)$. By the interpolation theorem (Theorem II. 3),

$$| u |_{1,I}^2 \le \gamma(\epsilon | u |_{2,I}^2 + \epsilon^{-1} \| u \|_I^2)$$

for all $0 < \epsilon \le 1$. Combining this with (9) and (11), we have that

$$\| Vu \|_I^2 \le M^2 \gamma [\epsilon K^2 (\| H_0 u \|_I + \| u \|_I)^2 + \epsilon^{-1} \| u \|_I^2].$$

By choosing ϵ so that

$$\epsilon < \tfrac{1}{4}(M^2 \gamma K^2)^{-1},$$

we have

$$\| Vu \|_I \le \tfrac{1}{2} \| H_0 u \| + b \| u \|_I,$$

where b depends on M, γ, and K. Hence, V is Kato-small with respect to H_0 and Theorem III. 2 applies. QED

In the next theorem, we will introduce an inner product on $W_2^1(I)$ which will play a vital role in the construction of the R matrix. For the most part, the next theorem is the reason for introducing R-admissible potentials.

Theorem III. 3: Let V be an R-admissible operator. Then there exist positive constants λ, ρ_1, and ρ_2 such that

$$(12) \quad \rho_1 | u |_{1,I}^2 \le \langle u, u \rangle + (Vu, u)_I + \lambda \| u \|_I^2 \le \rho_2 | u |_{1,I}^2$$

for all $u \in W_2^1(I)$. Here $\langle u, u \rangle$ is defined by (5). Hence, the Hermitian form

$$(13) \quad \langle u, v \rangle_{v,\lambda} = \langle u, v \rangle + (Vu, v)_I + \lambda (u, v)_I$$

is another inner product on $W_2^1(I)$ and the norm

$$(14) \quad \langle u \rangle_{V,\lambda} = (\langle u, u \rangle_{V,\lambda})^{1/2}$$

is equivalent to the usual norm on $W_2^1(I)$. Moreover, given any $u \in D(H_0)$,

$$(15) \quad \langle u, v \rangle_{V,\lambda} = ((H_0 + V + \lambda + 1)u, v)_I$$

for all $v \in W_2^1(I)$. Hence, the operator $H = H_0 + V$ is bounded below.

Proof: We will work with the lower half of (12) first. From Schwartz's inequality and Lemma III. 3, we have that

$$\langle u, u \rangle + (Vu, u)_I \ge C^{-1} | u |_{1,I}^2 - \| Vu \| \, \| u \|.$$

Coupling this with (11), which holds because V is R-admissible, we obtain

$$(*) \quad \langle u, u \rangle + (Vu, u)_I \ge C^{-1} | u |_{1,I}^2 - M \| u \| \, | u |_{1,I}.$$

For any three positive numbers a, b, ϵ, it is obvious that

$$ab \le \tfrac{1}{2}(\epsilon a^2 + \epsilon^{-1} b^2).$$

Applying this to (*) gives

$$\langle u,u\rangle + (Vu,u)_I \geq (C^{-1} - \tfrac{1}{2}M\epsilon)\,|u|^2_{1,I} - \tfrac{1}{2}M\epsilon^{-1}\|u\|^2_I.$$

Choosing $\epsilon = C^{-1}M^{-1}$ and $\lambda = \frac{1}{2}M^2C$, we have $\frac{1}{2}C|f|^2_{1,I} \leq \langle u,u\rangle + (Vu,u)_I + \lambda(u,u)_I \leq (C+M+\lambda)|u|^2_{1,I}$, where the right half of the inequality follows from Lemma III. 3, inequality (11), and the fact that $\|u\|^2_I \leq |u|^2_{1,I}$.

Finally, (15) follows directly from Lemma III. 4 and the definition of $\langle u,v\rangle_{V,\lambda}$. QED

We remark that (15) implies that the inner product $\langle u,v\rangle_{V,\lambda}$ could have been obtained by completing $D(H_0)$ in the inner product $((H_0+V+\lambda+1)u,v)_I$.

We conclude with a theorem concerning the expansions in the eigenfunctions of $H = H_0 + V$ for V R-admissible.

Corollary III.2: Let V be R-admissible and let u_j be the orthonormal eigenfunctions of $H = H_0 + V$ belonging to the eigenvalues E_j. For any $v \in W_2^1(I)$, let

$$a_j = (u_j, v)_I;$$

then the expansion

$$\sum_{j=0}^{\infty} a_j u_j$$

converges to v in both $L^2(I)$ and $W_2^1(I)$. Moreover, we have

$$\langle v\rangle^2_{V,\lambda} = \sum_{j=0}^{\infty} (E_j+\lambda+1)\,|a_j|^2.$$

Proof: By Theorem III. 3, the set of functions

$$\tilde{u}_j = (E_j+\lambda+1)^{-1/2}u_j$$

is obviously orthonormal in the inner product $\langle\ ,\ \rangle_{V,\lambda}$. Moreover, any function $v \in W_2^1(I)$ for which

$$\langle \tilde{u}_j, v\rangle_{V,\lambda} = 0$$

for all j must vanish. This follows because

$$\langle \tilde{u}_j v\rangle_{V,\lambda} = ((H+\lambda+1)\tilde{u}_j, v)_I$$

and

$$((H+\lambda+1)\tilde{u}_j, v)_I = (E_j+\lambda+1)^{1/2}a_j.$$

Hence, all the Fourier coefficients of v vanish and $v = 0$. But then, the orthogonal complement of the span of the $\tilde{u}_j$ in $W_2^1(I)$ is the space consisting of the 0 vector. Hence, the $\tilde{u}_j$ span $W_2^1(I)$ (see Riesz-Nagy,[19] p. 72) and the $\tilde{u}_j$ form a complete orthonormal set in $W_2^1(I)$. The rest of the theorem follows from the properties of such a set plus some minor computations. QED

In closing, we remark that if V is a smooth function, then much of the regularity theory presented in the last section carries directly over to the case of nonzero V.

IV. CONCLUDING REMARKS

Although we have avoided the question of spin dependent systems, such systems present no real difficulty as long as the spin dependent interactions are confined to the potential term V. For the case in which the spin dependence is carried in the kinetic energy terms (e.g., the Dirac equation), the eigenvalue problem is different and our results do not apply.

Finally, if the surface S of the internal region has finitely many "corners" and "edges," our results obviously hold.

ACKNOWLEDGMENTS

Most of the results presented here stem from the author's thesis.[18] The author gratefully acknowledges the help and suggestions given by his advisor, Eugene P. Wigner. In addition, the author wishes to thank V. Bargmann, J. Kohn, E. Nelson, and D. Spencer for helpful suggestions.

APPENDIX A

In this appendix, we collect some definitions and lemmas given by Schechter.[6,9] We will use these in Sec. III.

1. Ellipticity, proper ellipticity, covering set

A $(2k)$th-order partial differential operator

$$L = -\sum_{|\alpha|\leq 2k} \Lambda_\alpha(x)D^\alpha,$$

$\Lambda_\alpha(x) \in C^\infty(\bar{I})$, is said to be *elliptic* in $\bar{I}$ if the characteristic polynomial

$$P_L(x,\xi) \equiv \sum_{|\alpha|=2k} \Lambda_\alpha(x)\xi^\alpha,$$

$$\xi^\alpha = \xi_1^{\alpha_1}\xi_2^{\alpha_2}\cdots\xi_n^{\alpha_n},$$

vanishes for no real n-component vector ξ. The elliptic operator L is said to be *properly elliptic* in $\bar{I}$ if for every $x_0 \in S$, every real nonzero vector $\mathbf{T}$ tangent to S at x_0, and every real nonzero vector $\mathbf{N}$ normal to S at x_0, the polynomial

$$P(z) \equiv P_L(x_0, \mathbf{N} + z\mathbf{T})$$

has exactly k roots,

$$\lambda_1(\mathbf{T},\mathbf{N}), \ldots, \lambda_k(\mathbf{T},\mathbf{N}),$$

with positive imaginary parts.

By a boundary operator, we mean an operator of the form

$$B = \sum_{|\alpha|\leq m} b_\alpha(x_0)D^\alpha,$$

where the coefficients $b_\alpha(x_0)$ need only be defined on the boundary S, but are assumed infinitely differentiable there.

A set of k boundary operators $\{B_j\}_{j=1}^k$,

$$B_j = \sum_{|\alpha|\leq m_j} b_{j\alpha}D^\alpha,$$

where $m_j < 2k$, is said to *cover* the properly elliptic operator L if at every point $x_0 \in S$, and for every real nonzero vector $\mathbf{T}$ tangent to S at x_0 and every real nonzero vector $\mathbf{N}$ normal to S at x_0, the polynomials

$$Q_j(z) \equiv \sum_{|\alpha|=m_j} b_{j\alpha}(x_0)(\mathbf{N} + z\mathbf{T})^\alpha$$

are linearly independent modulo the polynomial

$$S(z) = \prod_{j=1}^{k} [z - \lambda_j(\mathbf{T},\mathbf{N})],$$

where $\lambda_j(\mathbf{T},\mathbf{N})$ are the roots of $P(z)$ with positive imaginary parts. Said another way, the equation

$$\sum_{j=1}^{k} C_j(x_0)Q_j(z) = M(z)S(z),$$

where $M(z)$ is some polynomial in z, holds only if $C_j = M(z) = 0$ for $j = 1, \ldots, k$.

For more details, we refer the reader to Schechter's papers.[6,9]

2. Three important lemmas

We now state three lemmas: The first involves an inequality for properly elliptic operators, the second and third are regularity theorems.

Lemma A.1 (Schechter): There exists a constant K depending only on I, L, and the B_j such that

$$|v|^2_{2k,I} \le K(\|Lv\|^2_I + \|v\|^2_I)$$

for all $v \in C^\infty(\bar{I})$ satisfying

$$B_j v = 0 \quad \text{on } S, \quad j = 1, \ldots, k,$$

if and only if L is *properly elliptic* and the set $\{B_j\}_{j=1}^k$ covers L.

Proof: See Schechter[9] for the statement and references.

Lemma A.2 (Sobolev): If $u \in W_2^l(I)$ for all $l \ge 0$, then $u \in C^\infty(I)$ after correction on a set of measure zero.

Proof: See Sobolev,[13] p. 69. The statement we use may be found in Schechter,[6] Lemma 6.1.

Lemma A.3 (Schechter): Let E_j, E_j' be partial differential operators of order $\le m_j \le m$. For all $f, g \in W_2^m(I)$, define

$$\langle f,g\rangle = \sum_j [E_j f, E_j' g]_{m-m_j,I}.$$

Further suppose that there exists a constant c such that

$$c^{-1}|f|^2_{m,I} \le \langle f,f\rangle \le c|f|^2_{m,I}$$

for all $f \in W_2^m(I)$. If $u \in W_2^m(I)$ and $g \in W_2^l(I)$ satisfy

$$\langle u,v\rangle = (g,v)_I$$

for all $v \in W_2^l(I)$, then $u \in W_2^{2m+l}(I)$. Moreover, if $g \in C^\infty(\bar{I})$, then so is u.

Proof: See Schechter,[6] Theorem 6.1, for a more general statement of the theorem and for the proof.

APPENDIX B

Consider a system of N particles with positions $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and let x stand for the $3N$-dimensional position in configuration space. We wish to show that the many particle Coulomb-like potential

$$(1) \quad V(x) = \sum_{i\ne j} [C_{ij}(x)/|\mathbf{x}_i - \mathbf{x}_j|],$$

where $C_{ij}(x) = C_{ji}(x)$ is a real, bounded measurable function of x in $\bar{I}$, is R-admissible. Supposing that f is in the domain of the operators

$$|\mathbf{x}_i - \mathbf{x}_j|^{-1},$$

we have

$$(2) \quad \|Vf\|_I \le \sum_{i\ne j} M_{ij}\|\,|\mathbf{x}_i - \mathbf{x}_j|^{-1}f\|_I$$

where M_{ij} bounds $C_{ij}(x)$. To show that $V(x)$ is R-admissible then reduces to showing that for all $f \in W_2^1(I)$, there exists a constant K, independent of f, such that

$$(3) \quad \|\,|\mathbf{x}_i - \mathbf{x}_j|^{-1}f\|_I \le K|f|_{1,I}.$$

Theorem B.1: $V(x)$ is R-admissible

Proof: Let $f \in C_0^\infty(E^{3N})$ and let $\mathbf{X} = \mathbf{x}_i - \mathbf{x}_j$ and $r = |\mathbf{X}|$. Holding all the coordinates except X_1, X_2, X_3 constant, we have the following inequality due to Courant,[22] p. 446:

$$\int \frac{|f|^2}{r^2}\, d^3X \le 4\int\left(\left|\frac{\partial f}{\partial X_1}\right|^2 + \left|\frac{\partial f}{\partial X_2}\right|^2 + \left|\frac{\partial f}{\partial X_3}\right|^2\right)d^3X,$$

where the integrals are over all values of $\mathbf{X}$. Integrating over the remaining coordinates of f and using the obvious inequality

$$\left|\frac{\partial f}{\partial X_1}\right|^2 + \left|\frac{\partial f}{\partial X_2}\right|^2 + \left|\frac{\partial f}{\partial X_3}\right|^2 \le C^2|\nabla f|^2,$$

where C comes from changing coordinates, we have

$$\|f/r\|_{E^{3N}} \le 2C\|\nabla f\|_{E^{3N}}$$

Also,

$$\|\nabla f\|_{E^{3N}} \le |f|_{1,E^{3N}}.$$

Hence

$$(4) \quad \|f/r\|_{E^{3N}} \le 2C|f|_{1,E^{3N}}.$$

By Lemma II.1, $C_0^\infty(E^{3N})$ is dense in $W_2^1(I)$. Hence, by taking limits, (4) holds for all $f \in W_2^1(E^{3N})$. By the Calderón extension theorem (Lemma II.2), given any $f \in W_2^1(I)$, there exists a bounded linear transformation T from $W_2^1(I)$ to $W_2^1(E^{3N})$ such that

$$Tf|_I = f.$$

Using Tf in (4) along with the obvious inequality,

$$\|Tf/r\|_I = \|f/r\|_I \le \|Tf/r\|_{E^{3N}},$$

we have

$$(5) \quad \|f/r\|_I \le 2C|Tf|_{k,E^{3N}} \le 2CC'|f|_{1,I},$$

where the upper inequality follows from the boundedness of T. Hence, (3) holds for all $f \in W_2^1(I)$ with $K = 2CC'$. By our earlier discussion, it immediately follows that $V(x)$ is R-admissible. QED

Two remarks are now in order: First of all, $V(x)$ not only includes the case of the Coulomb potential, but also the Yukawa potential; secondly nothing precludes separating out the center of mass and the result holds even in that case.

*Supported by the U. S. Air Force Office of Scientific Research under Grant No. 73-2484.

[1]E. P. Wigner, Phys. Rev. **70**, 15 (1946).
[2]E. P. Wigner, Phys. Rev. **70**, 606 (1946).
[3]E. P. Wigner, Phys. Rev. **72**, 29 (1947).
[4]G. Breit and E. P. Wigner, Phys. Rev. **49**, 515 (1936).
[5]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, N. J., 1971).
[6]M. Schechter, Commun. Pure Appl. Math. **12**, 457 (1959).
[7]O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and Quasilinear Elliptic Equations* (Academic, New York and London, 1968).
[8]J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations* (Springer, New York, 1971).
[9]M. Schechter, Commun. Pure Appl. Math. **12**, 561 (1959).
[10]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[11]N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N. J., 1965).
[12]M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965).
[13]S. L. Sobolev, *Applications of Functional Analysis in Mathematical Physics* (American Mathematical Society, Providence, R. I., 1963).
[14]S. Agmon, *Lectures on Elliptic Boundary Value Problems* (Van Nostrand, New York, 1965).
[15]J. L. Lions and E. Magenes, *Problèms aux limites non homogènes et applications* (Dunod, Paris, 1968), Vol. I.
[16]A. P. Calderón, "Lebesque Spaces on Differentiable Functions," in *Proceedings of the Fourth Symposia in Pure and Applied Mathematics* (American Mathematical Society; Providence, R. I., 1961).
[17]V. I. Smirnov, *A Course of Higher Mathematics* (Addison-Wesley, Reading, Mass., 1964), Vol. V.
[18]F. J. Narcowich, Ph.D. thesis (Princeton University, 1972) (unpublished).
[19]F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955).
[20]M. A. Naimark, *Normed Rings* (Noordhoof, Groningen, The Netherlands, 1964).
[21]J. R. Ringrose, *Compact Non-Self-Adjoint Operators* (Van Nostrand-Reinhold, London, 1971).
[22]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1966), Vol. I.

The mathematical theory of the R matrix. II. The R matrix and its properties*

Francis J. Narcowich

Department of Mathematics and Department of Physics, Texas A&M University, College Station, Texas 77843
(Received 12 June 1973)

In this paper, it is shown that Wigner's R matrix, for a certain class of unbounded potentials which may be nonlocal or have Coulomb-type singularities, exists, is a compact operator, and that the expansions associated with the R-matrix converge. For the same class of potentials, a perturbation theory is constructed and conditions are given for the convergence of the resulting Born-type expansions.

1. INTRODUCTION

A. Background

The existence and properties of the R matrix and the convergence of the expansions associated with it have been rigorously examined only in one-dimensional cases or cases in which separation of variables is possible,[1] even though the R matrix has been used extensively since its inception.[2] In this paper, we will show that, in a general setting, the R matrix exists, is a compact operator, and that the usual expansions associated with it converge. In addition, we will construct a perturbation theory for the R matrix and give conditions for the convergence of the Born-type expansions that arise.

In Paper I,[3] we discussed a slightly generalized version of the eigenvalue problem associated with the R matrix. Similarly, we shall also discuss a generalized version of the R matrix. In the usual R matrix theory, the configuration space of a system of particles with n spinless degrees of freedom is divided into two regions: the internal region I, which is bounded and has a smooth surface S; and, the external region, which is the complement of I. Given, in I, a solution ψ to the time-independent Schrödinger equation,

$$(-\Delta + V)\psi = E\psi, \tag{1}$$

the R matrix takes $\partial\psi/\partial n$ (= normal derivative of ψ on S) into $\psi|_S$ (= restriction of ψ to S or "value of ψ on S"). Instead of the time-independent Schrödinger equation, we shall consider the partial differential equation

$$(Q+V)\psi \equiv -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} A_{ij}(x) \frac{\partial\psi}{\partial x_j} + V\psi = E\psi. \tag{2}$$

Here, the $A_{ij}(x)$ are smooth and form the components of a uniformly positive definite Hermitian matrix. In this more general case, we define the R matrix as the operator which takes the derivative

$$\partial_A\psi \equiv \sum_{i,j} n_i A_{ij}(x) \frac{\partial\psi}{\partial x_j}, \tag{3}$$

where the n_i are components of the outward-drawn normal to S, into $\psi|_S$. In case $A_{ij}(x)$ is the unit matrix, we are back to (1) and the usual R matrix.

A formal expansion for the "generalized" R matrix can be obtained in the same way as the formal expansion for the usual R matrix. First, the eigenvalue problem

$$(Q+V)U_k = E_k U_k, \quad \partial_A U_K = 0 \quad \text{on } S \tag{4}$$

is solved; then, ψ is expanded in the U_k,

$$\psi = \sum_{k=0}^{\infty} A_k U_k. \tag{5}$$

Using Green's theorem, it is easily seen that

$$A_k = (E_k - E)^{-1} (U_k|_S, \partial_A\psi)_S,$$

where $(\ ,\ )_S$ is the inner product in $L^2(S)$ with the usual surface measure and $U_k|_S$ is the restriction of U_k to S. Substituting the expression for A_k into the expansion for ψ and evaluating on the boundary S, we obtain

$$\psi|_S = \sum_{k=0}^{\infty} \frac{1}{E_k - E} U_k|_S (U_k|_S, \partial_A\psi)_S.$$

Thus the R matrix, $R(E)$, has the formal operator expansion

$$R(E) = \sum_{k=0}^{\infty} \frac{1}{E_k - E} P_k, \tag{6}$$

where, for any $\sigma \in L^2(S)$,

$$P_k\sigma \equiv U_k|_S (U_k|_S, \sigma)_S.$$

There are two major difficulties with the approach outlined above: we have assumed (a) the existence of a ψ satisfying (2) and having roughly arbitrary surface derivative $\partial_A\psi$, and (b) the convergence of the expansion (5) to ψ on S. (a) is crucial, for without it, $R(E)$ may not be densely defined and, hence, may not be an operator! In Sec. 3, we will show that it is possible to resolve these difficulties for a large class of Hermitian operators V, which we call R admissible and which includes potentials with Coulomb-type singularities (see Sec. 2 C).

B. Organization and summary

The remainder of the paper is divided into two sections. In Sec. 2, we establish notation and summarize Paper I. In Sec. 3, we construct the R matrix for R-admissible operators (see Sec. 2 C), show that it is actually a compact operator, and prove that the expansion (6) converges to $R(E)$ in the uniform topology of $L^2(S)$. In the last part of Sec. 3, we construct a perturbation theory for the R matrix and give conditions for the convergence of the resulting Born-type expansions.

2. SUMMARY OF PAPER I

A. Notation

In what follows, the symbol I denotes a bounded, open region of Euclidean n-dimensional space, E^n. The boundary (surface) of I, S, is an infinitely differentiable, $(n-1)$-dimensional manifold. S is orientable and I lies entirely on one side of S.

 Copyright © 1974 American Institute of Physics

We will use a number of spaces in the course of the paper. $L^2(I)$, $L^2(S)$, and $W_2^l(I)$ are, respectively, the spaces of complex-valued square integrable functions on I, S, and the space of complex-valued functions on I which have all $0 \leq j \leq l$ square integrable generalized derivatives.[3,4] The inner products and norms are denoted by $(\ ,\ )_I$, $\|\ \|_I$; $(\ ,\ )_S$, $\|\ \|_S$; and $[\ ,\ ]_{l,I}$, $|\ |_{l,I}$, respectively. All inner products are linear on the right and conjugate linear on the left.

In addition to these Hilbert spaces, we will also use the spaces $C^\infty(\bar{I})$, $C^\infty(S)$, and $C_0^\infty(I)$. These are, respectively, the set of all complex-valued, infinitely differentiable functions on $\bar{I}$ (= closure of I), S and the subset of $C^\infty(\bar{I})$ whose elements vanish outside of some compact subset of I.

By Q, we denote the partial differential operator,

$$Q = -\sum_{i,j=1} \frac{\partial}{\partial x_i} A_{ij}(x) \frac{\partial}{\partial x_j}, \tag{7}$$

where $A_{ij}(x)$ satisfies

$(C_1)\, A_{ij}(x) \in C^\infty(\bar{I})$,

$(C_2)\, \overline{A_{ij}(x)} = A_{ji}(x)$,

$(C_3)\, \mu_0 |\xi|^2 \leq \sum_{i,j} \bar{\xi}_i A_{ij}(x) \xi_j \leq \mu_1 |\xi|^2$,

and where μ_0, μ_1 are positive constants and ξ is an arbitrary n-component complex-valued vector with norm $|\xi|$. (C_2) and (C_3) hold for all $x \in \bar{I}$. By ∂_A we denote the boundary operator,

$$\partial_A = \sum_{i,j} n_i A_{ij}(x) \frac{\partial}{\partial x_j}, \tag{8}$$

where n_i are the components of the outward unit normal to S.

B. The trace of a function in $W_2{}^l(I)$

Given an arbitrary function f in $L^2(I)$, it is impossible to assign any meaning to the restriction of f to S. However, for functions in $W_2{}^l(I)$, $l \geq 1$, this is possible, as the following "trace theorem" shows.

Theorem 2.1: For all $f \in C^\infty(\bar{I})$ and every $\epsilon > 0$, $f|_S$ satisfies

$$\|f|_S\|_S \leq \epsilon |f|_{1,I} + C(\epsilon) \|f\|_I,$$

where $C(\epsilon)$ depends on I and ϵ, but not f. Hence, the linear map $\tau f \equiv f|_S$ can be extended to all functions in $W_2{}^1(I)$ and, when so extended, is a compact map from $W_2{}^1(I)$ into a dense subset of $L^2(S)$.

Proof: The proof and a discussion may be found in Paper I (Theorem 1.4). Further references are given there.

As we pointed out in Paper I, the trace can be used to extend the formulae for integration by parts to all functions in $W_2{}^l(I)$. This is because $f \in W_2{}^l(I)$ implies all derivatives of f of order l-1 or less are in $W_2{}^1(I)$.

C. The eigenvalue problem

Given an Hermitian operator V on $L^2(I)$, the eigenvalue problem associated with the R matrix is to find a complete set of eigenfunctions U_k such that

$$(Q+V)U_k = E_k U_k \tag{9}$$
$$\partial_A U_k = 0 \text{ on } S,$$

where Q is defined by (1) and ∂_A by (2). The last equation is taken to hold in the sense of trace. Precisely, let

$$H_0 = Q|_{D(H_0)}, \tag{10}$$

where $D(H_0)$ consists of all f in $W_2{}^2(I)$ such that $\partial_A f = 0$ on S. We then have the following theorems from Paper I.

Theorem 2.2: The operator H_0 is positive and self-adjoint. Given any complex number E not in the spectrum of H_0, the operator $(H_0 - E)^{-1}$ maps $W_2{}^l(I)$ into $W_2{}^{l+2}(I)$ for all $l \geq 0$. In particular, $(H_0 - E)^{-1}$ maps $C^\infty(\bar{I})$ into $C^\infty(\bar{I})$. Also, $(H_0 - E)^{-1}$ maps $L^2(I)$ compactly into $W_2{}^1(I)$. The spectrum of H_0 consists of countably many nonnegative eigenvalues with $+\infty$ being the only limit point. The eigenfunctions of H_0 are in $C^\infty(\bar{I})$ and pointwise satisfy the boundary conditions.

Proof: See Paper I, Theorem 1.2.

Theorem 2.3: Let V be an Hermitian operator on $L^2(I)$ satisfying

(a) $D(V) \supset D(H_0)$,

(b) there is an $a < 1$ and $b > 0$,

such that

$$\|Vf\|_I \leq a \|H_0 f\|_I + b\|f\|_I,$$

for all f in $D(H_0)$. (i.e., V is Kato-small with respect to H_0). Then, the operator

$$H = H_0 + V, \quad D(H) = D(H_0)$$

is self-adjoint and the spectrum of H consists of countably many eigenvalues with $\pm\infty$ being the only limit points. Moreover, the operator $(H-E)^{-1}$ maps $L^2(I)$ compactly into $W_2{}^1(I)$ and, hence, $L^2(I)$, for any complex E not in the spectrum of H.

Proof: See Theorem 2.2, Paper I.

An Hermitian operator V with $D(V) \subset L^2(I)$ is said to be *R-admissible* if $D(V) \supset W_2{}^1(I)$ and if for all $f \in W_2{}^1(I)$,

$$\|Vf\|_I \leq M |f|_{1,I}, \tag{11}$$

where M is independent of f.

We remark that, besides including all bounded Hermitian operators, the class of R-admissible operators includes the physically interesting Coulomb-like potential

$$V(x) = \sum_{i \neq j} [C_{ij}(x)/|\mathbf{x}_i - \mathbf{x}_j|],$$

where $C_{ij}(x)$ is a real-valued bounded function of x and $\mathbf{x}_i$ is the position of the ith particle in an N particle system.

R-admissible operators will play a central role in the construction of the R matrix.

Theorem 2.4: Let V be R-admissible. Then V is *Kato-small* with respect to H_0 and, hence, Theorem 3 holds for such V. Moreover, $H = H_0 + V$ is bounded below and there exists a constant λ such that the Hermitian form

$$\langle f, g\rangle_V \equiv \int_I \sum_{i,j} \frac{\overline{\partial f}}{\partial x_i} A_{ij}(x) \frac{\partial g}{\partial x_j}\, d^n x + (Vf, g)_I + \lambda(f, g)_I, \qquad (12)$$

which is defined for all $f, g \in W_2^1(I)$, forms a new inner product on $W_2^1(I)$ whose norm,

$$\langle f\rangle_V = (\langle f, f\rangle_V)^{1/2} \qquad (13)$$

is equivalent to the usual norm on $W_2^1(I)$. Finally, if $f \in D(H_0)$, $g \in W_2^1(I)$

$$\langle f, g\rangle_V = ((H_0 + V + \lambda)f, g)_I, \qquad (14)$$

and, conversely, if there exists $h \in L^2(I)$ such that

$$\langle f, g\rangle_V = (h, g)_I,$$

then $f \in D(H_0)$ and $h = (H_0 + V + \lambda)f$.

Proof: See Theorems 2.1, 2.3, Corollary 2.1 and Lemma 2.4 of Paper I.

Concerning the inner product $\langle\ ,\ \rangle_V$, we have the following important corollary:

Corollary 2.1: Let V be R-admissible and let U_k be the orthonormal eigenfunctions of $H = H_0 + V$ belonging to the eigenvalues E_k. For any $v \in W_2^1(I)$, let

$$A_k \equiv (U_k, v)_I.$$

Then the expansion

$$\sum_{k=0}^{\infty} A_k U_k$$

converges to v in both $L^2(I)$ and $W_2^1(I)$. Moreover, we have

$$\langle v\rangle_V^2 = \sum_{k=0}^{\infty} (E_k + \lambda)|A_k|^2.$$

Proof: See Corollary 2.2 of Paper I.

We close by remarking that if V preserves $W_2^l(I)$ for all $l \geq 0$, then the part of Theorem 2 concerning the regularity of $(H_0 - E)^{-1}f$ holds for $(H - E)^{-1}f$ as well.

3. CONSTRUCTION AND PROPERTIES OF THE R-MATRIX

To construct the R matrix, we first solve the boundary value problem

$$(Q + V)\psi = E\psi, \qquad (15)$$

$$\partial_A \psi = \sigma,$$

where $\sigma \in C^\infty(S)$, $\psi \in W_2^2(I)$, E is not an eigenvalue of $H = H_0 + V$, and $D(V) \supset C^\infty(\bar{I})$. Next, we define a linear operator $U(E)$ which maps σ into ψ. If V is R-admissible $U(E)$ can then be extended to a compact map from $L^2(S)$ to $W_2^1(I)$. Finally, the composition $\tau U(E)$ is the R matrix, $R(E)$. $\tau U(E)$ maps $\sigma = \partial_A \psi$ into the value of ψ on the surface S.

The procedure outlined above is carried out in Secs. 3 A and B. In Sec. 3 C, we discuss some of the properties of the R matrix and show that the Mittag—Leffler expansion given in Sec. I A converges in the uniform topology of $L^2(S)$. Finally, in the last section, we discuss the perturbation of R matrix and give conditions for the convergence of Born-type expansions.

A. Solution of the boundary value problem

To solve the boundary value problem (15), we will borrow a trick from the theory of partial differential equations: Pick any $v \in W_2^2(I)$ and suppose $\partial_A v = \sigma$. If ψ solves (15) and $\psi \in W_2^2(I)$, then $\psi - v \in D(H)$ because $\partial_A(\psi - v) = \sigma - \sigma = 0$. Applying $H - E$ to $\psi - v$ and using the definition of H, we have

$$(H - E)(\psi - v) = (Q + V - E)\psi - (Q + V - E)v.$$

Since ψ solves (15), the first term on the right vanishes. After multiplying by $(H - E)^{-1}$ and rearranging terms, we obtain

$$\psi = v - (H - E)^{-1}(Q + V - E)v. \qquad (16)$$

Conversely, given ψ satisfying (2) with v having the properties stated earlier, ψ obviously satisfies the boundary value problem (15). With this in mind, we now prove the following theorem:

Theorem 3.1: Given any $\sigma \in C^\infty(S)$ and any Hermitian operator V such that V is Kato-small compared to H_0 (see Theorem 2.3), and such that $D(V) \supset C^\infty(\bar{I})$, the boundary value problem

$$(Q + V)\psi = E\psi, \quad \partial_A \psi = \sigma,$$

where E is not an eigenvalue of $H = H_0 + V$, has a unique solution in $W_2^2(I)$. Moreover, if V maps $C^\infty(\bar{I})$ into $C^\infty(\bar{I})$, then $\psi \in C^\infty(\bar{I})$ and the boundary conditions are satisfied pointwise.

Proof: By a theorem of Schechter (Ref. 5, Corollary 4.1), it is possible to construct a function $v \in C^\infty(\bar{I})$ such that $\partial_A v = \sigma$. If we define ψ by (2), we see that $\psi \in W_2^2(I)$ (see Theorems 2.2 and 2.3). If V preserves $C^\infty(\bar{I})$, then, by the remark at the end of Section 2.3, $(H - E)^{-1}$ preserves $C^\infty(\bar{I})$. Since $v \in C^\infty(\bar{I})$, $(Q + V - E)\, v \in C^\infty(\bar{I})$ and hence, $\psi \in C^\infty(\bar{I})$. Successively applying $(Q + V - E)$ and ∂_A to ψ, we see that ψ, as defined by (16), solves (15). Moreover, the continuity of the derivatives of ψ in $\bar{I}$ guarantee that the boundary conditions on taken on pointwise, otherwise they are taken on in the sense of trace.

Finally, we will show that ψ is unique and, hence, independent of our choice of v. Suppose $\psi' \in W_2^2(I)$ also solves (15). Then $\psi - \psi' \in D(H) = D(H_0)$, for $\partial_A(\psi - \psi') = \sigma - \sigma = 0$. Applying $(H - E)$ to $\psi - \psi'$ gives

$$(H - E)(\psi - \psi') = (Q + V - E)\psi - (Q + V - E)\psi' = 0.$$

Since E is not in the spectrum of H, we must have $\psi = \psi'$. Hence, ψ is unique and depends only on σ. QED

Two remarks are now in order. First, Theorem 3.1 holds for any R-admissible V because $D(V) \supset C^\infty(\bar{I})$. Second, $\psi = 0$ if and only if $\sigma = 0$: Obviously, if $\sigma = 0$, $v \in D(H)$ and (2) vanishes identically. Conversely, if $\psi = 0$, (2) implies that $v \in D(H)$ and hence, $\sigma = 0$.

Since ψ is uniquely determined by σ, we may define the following map $U(E)$:

$$U(E): \sigma \to \psi. \qquad (17)$$

$U(E)$ is obviously linear and is defined for all $\sigma \in C^\infty(S)$. The next section will be devoted to studying $U(E)$.

B. The operator $U(E)$

The main result of this section is the following theorem:

Theorem 3.2: Let V be R-admissible. Then the operator $U(E)$ defined by (17) can be extended to all of $L^2(S)$

and when so extended is a compact map from $L^2(S)$ to $W_2^1(I)$. Moreover, setting $\psi = U(E)\sigma$ and letting $f \in W_2^1(I)$

$$\langle f, \psi\rangle_V = (\tau f, \sigma)_S + (E+\lambda)(f,\psi)_I, \tag{18}$$

where $\langle\ ,\ \rangle_V$ is defined by formula (12) and τf is the trace of f. Finally, $U(E)\sigma = 0$ if and only if $\sigma = 0$.

Proof: We begin by justifying (18) for $\sigma \in C^\infty(S)$. To do this, we need only note that $\psi = U(E)\sigma \in W_2^2(I)$ by Theorem 1 and integration by parts is justified. Hence,

$$\int_I \sum_{i,j} \frac{\partial \bar f}{\partial x_i} A_{ij}(x) \frac{\partial \psi}{\partial x_j}\, d^n x$$

$$= \int_I - \sum_{i,j} \bar f \frac{\partial}{\partial x_i} A_{ij}(x) \frac{\partial \psi}{\partial x_j} + \int_S dS\, \overline{\tau f}\, \partial_A \psi.$$

Since $\partial_A \psi = \sigma$,

$$-\sum_{i,j} \frac{\partial}{\partial x_i} A_{ij}(x) \frac{\partial \psi}{\partial x_j} = Q\psi, \tag{19}$$

formula (18) follows immediately.

Setting $E = -\lambda$ (which obviously cannot be in the spectrum of H) in (18), we have

$$\langle f, U(-\lambda)\sigma\rangle_V = (\tau f, \sigma)_S, \tag{20}$$

where $f \in W_2^1(I)$ and $\sigma \in C^\infty(S)$. However, (20) implies that $U(-\lambda)\sigma$ is nothing more than the adjoint of the compact operator τ, relative to the inner product $\langle\ ,\ \rangle_V$ on $W_2^1(I)$, when restricted to $\sigma \in C^\infty(S)$. Hence, we may extend $U(-\lambda)$ to all of $L^2(S)$ by this correspondence. Since τ is a compact map from $W_2^1(I)$ to $L^2(S)$ (see Theorem 2.1), the adjoint of τ relative to $\langle\ ,\ \rangle_V$ is a compact map from $L^2(S)$ to $W_2^1(I)$ (see Riesz–Nagy, Ref. 6, p. 217). Hence, $U(-\lambda)$ can be extended to a compact map from $L^2(S)$ to $W_2^1(I)$.

Again restricting σ to $C^\infty(S)$, we obtain from (16) the following relation between $U(-\lambda)\sigma$ and $U(E)\sigma$ by substituting $U(-\lambda)\sigma$ for v:

$$U(E)\sigma = U(-\lambda)\sigma - (H-E)^{-1}(Q+V-E)U(-\lambda)\sigma.$$

Since $(Q+V)U(-\lambda)\sigma = -\lambda U(-\lambda)\sigma$, this last formula implies

$$U(E)\sigma = U(-\lambda)\sigma + (\lambda+E)(H-E)^{-1}U(-\lambda)\sigma$$

or, for $\sigma \in C^\infty(S)$,

$$U(E) = U(-\lambda) + (\lambda+E)(H-E)^{-1}U(-\lambda). \tag{21}$$

By means of (21), we can extend $U(E)$ to all of $L^2(S)$. Moreover, $U(E)$ is compact as a map from $L^2(S)$ to $W_2^1(I)$ because it is the sum of the compact operator $U(-\lambda)$ plus the product of the compact operator $(H-E)^{-1}$ with $U(-\lambda)$ [$(H-E)^{-1}$ maps $L^2(I)$ compactly into $W_2^1(I)$ by Theorem 2.3, it therefore maps $W_2^1(I)\ C\ L^2(I)$ compactly into $W_2^1(I)$].

Formula (18) may be established for all $\sigma \in L^2(S)$ by first noting that this has already been accomplished for $C^\infty(S)$, which is dense in $L^2(S)$ (see Ref. 4, p. 40) and then by taking limits.

Finally, $U(E)\sigma = 0$ implies $(\tau f, \sigma)_S = 0$ for all $f \in W_2^1(I)$. Since the range of τ is dense in $L^2(S)$ (Theorem 1.1), $\sigma = 0$. Conversely, we have already seen that $\sigma = 0$ implies $\psi = 0$. QED

From now on, we shall mean $U(E)$ in the extended sense given by Theorem 2.

We remark that even for arbitrary σ, $U(E)\sigma$ still satisfies the boundary value problem (15) in a generalized sense. It is relatively easy to show that if $Q_0 \equiv Q|_{D(Q_0)}$, where $D(Q_0)$ is the set of all $f \in C_0^\infty(I)$, then

$$(Q_0^* + V)(U(E)\sigma) = EU(E)\sigma.$$

The boundary conditions are then satisfied in an inner product sense.

The following corollary to Theorem 2 will be crucial in establishing the convergence of the R matrix expansions:

Corollary 3.1: Let V be R-admissible and let U_k, and E_k be as in Corollary 2.1. Then, if $E \neq E_k$,

$$A_k = (U_k, U(E)\sigma)$$

is given by

$$A_k = (E_k - E)^{-1}(\tau U_k, \sigma)_S, \tag{22}$$

and the conclusions of Corollary 2.1 hold.

Proof: Let $\psi = U(E)\sigma$, by Theorem 2.4,

$$\langle f, \psi\rangle_V = ((H+\lambda)f, \psi)_I$$

for any $f \in D(H)$. Setting $f = U_k$, we have

$$\langle U_k, \psi\rangle_V = (E_k + \lambda)\,(U_k, \psi)_I.$$

By (19), however, we also have

$$\langle U_k, \psi\rangle_V = (\tau U_k, \sigma) + (E+\lambda)(U_k, \psi)_I.$$

Solving these two equations for $(U_k, \psi)_I = A_k$ gives (22).

QED

C. The R matrix

In what follows, we assume that V is R-admissible and that E is any complex number not in the spectrum of H.

As we pointed out in the introductory paragraphs to this section, the R matrix is defined by

$$R(E) = \tau U(E). \tag{23}$$

For $\sigma \in C^\infty(S)$, $R(E)$ maps the derivative $\partial_A(U(E)\sigma)$ into the value of $U(E)\sigma$ on S.

The next theorem gives several important properties of $R(E)$.

Theorem 3.3: $R(E)$ is a compact map from $L^2(S)$ to $L^2(S)$. The spectrum of $R(E)$ consists of countably many eigenvalues with 0 being the only limit point. In addition, the Hermitian from $(\alpha, R(E)\sigma)_S$, $\alpha, \sigma \in L^2(S)$, has the absolutely convergent Mittag–Leffler expansion

$$(\alpha, R(E)\sigma) = \sum_{k=0}^{\infty} (E_k - E)^{-1}(\alpha, \tau U_k)_S\,(\tau U_k, \sigma)_S. \tag{24}$$

Hence, $R(E)^* = R(\bar E)$ and, for real E, $R(E)$ is self-adjoint.

Proof: Since $R(E)$ is the composition of the compact maps $U(E)$ (Theorem 3.2) and τ (Theorem 2.1), $R(E)$ is itself compact. The characterization of the spectrum of $R(E)$ is simply the characterization of the spectrum of an arbitrary compact operator (see Widom, Ref. 7, p. 23).

To establish (24), consider formula (18) with $f = U(E)\sigma$, and $\psi = U(E)\alpha$,

$$\langle U(E)\sigma, U(E)\alpha\rangle_V = (\tau U(E)\sigma, \alpha)_S + (E+\lambda)(U(E)\sigma, U(E)\alpha)_I.$$

Hence,

$$(\tau U(E)\sigma, \alpha)_S = \langle U(E)\sigma, U(E)\alpha\rangle_V - (E+\lambda)(U(E)\sigma, U(E)\alpha)_I.$$

Using Corollary 3.1 and Corollary 2.1, this last formula becomes

$$(R(E)\sigma, \alpha) = \sum_{k=0}^{\infty} \frac{E_k+\lambda}{|E_k - E|^2}(\sigma, \tau U_k)_S (\tau U_k, \alpha)_S - \sum_{k=0}^{\infty} \frac{E+\lambda}{|E_k - E|^2}(\sigma, \tau U_k)_S (\tau U_k, \alpha)_S.$$

After simplification, we obtain

$$(R(E)\sigma, \alpha)_S = \sum_{k=0}^{\infty} \frac{1}{E_k - \bar{E}}(\sigma, \tau U_k)_S (\tau U_k, \alpha)_S.$$

Upon conjugation, we obtain (14). The expansion is absolutely convergent because it is the difference of two absolutely convergent expansions. Finally, to see that $R(E)^* = R(\bar{E})$, replace E by $\bar{E}$ in (9), interchange the roles of σ and α, and conjugate. This gives

$$(R(\bar{E})\alpha, \sigma)_S = \sum_{k=0}^{\infty} \frac{1}{E_k - E}(\alpha, \tau U_k)_S (\tau U_k, \sigma)_S.$$

Hence, for all α, $\sigma \in L^2(S)$,

$$(R(\bar{E})\alpha, \sigma)_S = (\alpha, R(E)\sigma)_S.$$

This is only possible if $R(\bar{E}) = R(E)^*$. QED

Theorem 3.3 already establishes the convergence of the matrix expansions for $R(E)$. In the next theorem, which is the main result of this paper, we will prove that the operator expansion (6) converges in the uniform topology of $L^2(S)$.

Theorem 3.4: Let P_k be the projection

$$P_k\sigma \equiv \tau U_k, \sigma)_S.$$

Then, $R(E)$ has the operator expansion,

$$R(E) = \sum_{k=0}^{\infty} \frac{1}{E_k - E} P_k, \tag{25}$$

where the expansion holds in the uniform topology of $L^2(S)$ [i.e. (25) converges norm-wise to $R(E)$. See Ref. 6, p. 150].

Proof: By Corollaries 1 and 2.1,

$$U(E)\sigma = \sum_{k=0}^{\infty} \frac{1}{E_k - E} \tau U_k (\tau U_k, \sigma)_S,$$

where the expansion given converges to $U(E)\sigma$ in the norm of $W_2^{\,1}(I)$. Since τ, the trace, is a compact map from $W_2^{\,1}(I)$ to $L^2(S)$ (see Theorem 2.1), it is also continuous. Hence,

$$\tau U(E)\sigma = \sum_{k=0}^{\infty} \frac{1}{E_k - E} \tau U_k(\tau U_k, \sigma)_S.$$

or, for any fixed $\sigma \in L^2(S)$,

$$R(E)\sigma = \sum_{k=0}^{\infty} \frac{1}{E_k - E} P_k\sigma.$$

Thus, the expansion (25) converges to $R(E)$ in the strong sense. To prove uniform convergence, we must show that (25) converges independently of σ.

First of all, we note that for real E, the operator sequence $R_N(E)$, defined by

$$R_N(E) \equiv \sum_{k=0}^{N} \frac{1}{E_k - E} P_k,$$

is an increasing sequence of operators which is bounded above:

$$(\sigma, R_N(E)\sigma) = \sum_{k=0}^{N} \frac{1}{E_k - E} |(\tau U_k, \sigma)_S|^2 \le \sum_{k=0}^{\infty} \frac{1}{E_k - E} |(\tau U_k, \sigma)_S|^2.$$

By (25) we have,

$$(\sigma, R_N(E)\sigma)_S \le (\sigma, R(E)\sigma)_S.$$

Hence, by a theorem of Vigier (see Ref. 6, p. 263), $R_N(E)$ has a uniform limit. Since the uniform limit and the strong limit are the same, provided the former exists,

$$R(E) = \underset{N\to\infty}{\text{uniform-limit}}\, R_N(E)$$

and (25) holds uniformly for real E. For complex E, a similar argument holds after breaking $R(E)$ into real and imaginary parts. QED

D. Perturbation of the R matrix

Let V_1 and V be R-admissible operators. Obviously, the sum $V_1 + V$ is also R-admissible and, assuming E is not in the spectrum of either $H_1 = H_0 + V$, $H = H_0 + V_1 + V$, we can form the operators $U_1(E)$, $R_1(E)$ and $U(E)$, $R(E)$ associated with H_1 and H. Two questions naturally arise: (1) How are $U_1(E)$, $R_1(E)$ and $U(E)$, $R(E)$ related? (2) Can we obtain Born-type expansions for $U(E)$ and $R(E)$ in terms of $U_1(E)$, $R_1(E)$ and "powers" of V?

To answer the first question, let $\sigma \in C^\infty(S)$ and let $\psi_1 = U_1(E)\sigma$ and $\psi = U(E)\sigma$. By Theorem 1,

$$(Q + V_1)\psi_1 = E\psi_1$$

and

$$(Q + V_1 + V)\psi = E\psi.$$

Subtracting the first equation from the second and noting that $\partial_A\psi_1 = \sigma = \partial_A\psi$ implies that $\psi_1 - \psi \in D(H_0)$, we have

$$(H_1 - E)(\psi - \psi_1) = -V\psi,$$

or, returning to $\psi = U(E)\sigma$, $\psi_1 = U_1(E)\sigma$,

$$U_1(E)\sigma = [1 + (H_1 - E)^{-1}V]\,U(E)\sigma, \tag{26}$$

for all $\sigma \in C^\infty(S)$. Conversely, a similar argument shows that

$$U(E)\sigma = [1-(H-E)^{-1}V]\,U_1(E)\sigma, \tag{27}$$

which holds for all $\sigma \in C^\infty(S)$. This leads us to our next theorem,

Theorem 3.5: If E is not in the spectrum of H or H_1, then both $(H_1-E)^{-1}V$ and $(H-E)^{-1}V$ map $W_2{}^1(I)$ compactly into itself; in addition, the following hold:

(a) $1-(H-E)^{-1}V = [1+(H_1-E)^{-1}V]^{-1}$,

(b) $U(E) = [1+(H_1-E)^{-1}V]^{-1}\,U_1(E)$,

(c) $R(E) = \tau[1+(H_1-E)^{-1}V]^{-1}\,U_1(E)$.

Proof: The R-admissible of V guarantees that V maps $W_2{}^1(I)$ continuously into $L^2(I)$. By Theorems 2.3 and 2.4, $(H_1-E)^{-1}$ and $(H-E)^{-1}$ map $L^2(I)$ compactly into $W_2{}^1(I)$. Hence, the composition maps $(H_1-E)^{-1}V$, $(H-E)^{-1}V$ are compact.

To prove (a), we first note that both $1-(H-E)^{-1}V$ and $1+(H_1-E)^{-1}V$ are bounded maps on $W_2{}^1(I)$. Hence, no question of domains arises.

Multiplying the two together, we have,

$$\begin{aligned}
&[1-(H-E)^{-1}V][1+(H_1-E)^{-1}V] \\
&\quad = 1-\{(H-E)^{-1}-(H_1-E)^{-1}+(H-E)^{-1}V(H_1-E)^{-1}\}V \\
&\quad = 1-(H-E)^{-1}\{H_1-E)-(H-E)+V\}(H_1-E)^{-1}V \\
&\quad = 1-(H-E)^{-1}\{H_1+V-E)-(H-E)\}(H_1-E)^{-1}V.
\end{aligned}$$

Since the term in the braces vanishes identically, the product of the two operators is 1. An identical argument shows that

$$[1+(H_1-E)^{-1}V][1-(H-E)^{-1}V]=1.$$

Hence, $1+(H_1-E)^{-1}$ and $1-(H-E)^{-1}V$ are mutually inverse.

(b) follows from the boundedness of $[1-(H-E)^{-1}V]$ $U_1(E)$ coupled with the fact that (27) holds on $C^\infty(S)$, which is dense in $L^2(S)$. (c) follows from (b) and the definition of $R(E)$. QED

For convenience, define

$$T(E) \equiv (H_1-E)^{-1}V. \tag{28}$$

Under the assumption that the operator norm of $T(E)$ in $W_2{}^1(I)$ is less than unity, it is clear that

$$[1+T(E)]^{-1} = \sum_{l=0}^{\infty}(-1)^l T(E)^l, \tag{29}$$

where the expansion converges in the uniform topology of bounded operators on $W_2{}^1(I)$. By substituting this expansion into (b) and (c) of Theorem 5, we obtain the following Born-type expansions for $U(E)$ and $R(E)$:

$$U(E) = U_1(E) + \sum_{l=1}^{\infty}(-1)^l\,T(E)^l\,U_1(E) \tag{30}$$

and

$$R(E) = R_1(E) + \sum_{l=1}^{\infty}(-1)^l\,\tau[T(E)]^l\,U_1(E). \tag{31}$$

The condition on the $W_2{}^1(I)$ operator norm of $T(E)$ is inconvenient. First of all, there are many norms on $W_2{}^1(I)$; secondly, they are hard to compute with. As it turns out, there is a weaker condition which implies the convergence of (29), (30), and (31): $T(E)$ is actually a bounded operator on $L^2(I)$; if the $L^2(I)$ operator-norm of $T(E)$ is less than unity, then the expansions (29), (30), and (31) all converge in the stated topologies.

Before we state and prove the next theorem, let us recall a few facts about the inner product associated with V_1, $\langle\ ,\ \rangle_{V_1}$. By Theorem 2.4, there exists a constant λ_1 such that the Hermitian form

$$\langle f,g\rangle_{V_1} = \int_I d^n x \sum_{i,j}\frac{\partial \bar f}{\partial x_i}A_{ij}(x)\frac{\partial g}{\partial x_j} + (V_1 f, g)_I + \lambda_1(f,g)_I \tag{32}$$

is an inner product on $W_2{}^1(I)$ whose associated norm, $\langle\ \rangle_{V_1}$, is equivalent to the usual norm on $W_2{}^1(I)$, $|\ |_{1,I}$. This means that there exist constants ρ_1 and ρ_2 such that

$$\rho_1\langle f\rangle_{V_1} \le |f|_{1,I} \le \rho_2\langle f\rangle_{V_1}.$$

Thus, for the R-admissible operator V (also for V_1), there exists a constant $M' = \rho_2 M$ such that

$$\|Vf\|_I \le M'\langle f\rangle_{V_1}, \tag{33}$$

for all $f \in W_2{}^1(I)$. Finally, if $f \in D(H_0) = D(H_1)$ and $g \in W_2{}^1(I)$,

$$\langle g,f\rangle_{V_1} = (g,(H_0+V_1+\lambda_1)f) = (g,(H_1+\lambda_1)f). \tag{34}$$

We are now ready to prove a theorem and a corollary concerning the convergence of the expansions (29), (30), and (31).

Theorem 3.6: Let E not be an eigenvalue of H_1 or H and let $T(E)$ be defined by (28). Then, $T(E)$ can be extended to a compact, and, hence, bounded, operator on $L^2(I)$. Moreover, letting $N(E)$ be the $L^2(I)$ operator norm of $T(E)$, the condition $N(E)<1$ implies the convergence of the expansions (29), (30), and (31) in the stated topologies.

Proof: The show that $T(E)$ can be extended to all of $L^2(I)$ as a compact operator, consider the formal adjoint of $T(E)$, $V(H_1-\bar E)^{-1}$. By Theorems 1.3 and 1.4, $(H_1-\bar E)^{-1}$ maps $L^2(I)$ compactly into $W_2{}^1(I)$. Moreover, the R-admissibility of V implies that V maps $W_2{}^1(I)$ continuously into $L^2(I)$. Hence, $V(H_1-\bar E)^{-1}$ is the composition of a continuous operator and a compact operator and is, therefore, compact. We may then extend $T(E)$ by setting $T(E) = [V(H_1-\bar E)^{-1}]^*$. For functions in $W_2{}^1(I)$, this coincides with (28). $T(E)$ is then compact because it is the adjoint of a compact operator (see Ref. 6, p. 217).

In order to show that (30) and (31) converge for $N(E)<1$, we need only show that (29) converges in the uniform operator topology of $W_2{}^1(I)$. To do this, consider the operator identity

$$(1+T(E))^{-1} - \sum_{l=0}(-1)^l\,T(E)^l = (-1)^{L+1}\,T(E)^{L+1}(1+T(E))^{-1}. \tag{35}$$

Let $\phi = T(E)f$, where $f = T(E)^L(1+T(E))^{-1}g$, $g \in W_2{}^1(I)$. Using the norm $\langle\ \rangle_{V_1}$, we have by (34) and (28)

$$\langle\phi\rangle_{V_1}^2 = (\phi,(H_1+\lambda_1)(H_1-E)^{-1}Vf)_I.$$

Making an algebraic manipulation and using the hermiticity of V_1 we obtain

$$\langle\phi\rangle_{V_1}^2=(E+\lambda)\|\phi\|_I^2+(V\phi,f)_I.$$

Next, using Schwartz's inequality, (33), and the inequality

$$ab\le\tfrac12(\epsilon^2a^2+\epsilon^{-2}b^2),$$

which holds for all a, b, $\epsilon>0$, we see that

$$\langle\phi\rangle_{V_1}^2\le|E+\lambda|\,\|\phi\|_I^2+\tfrac12M'(\epsilon^2\langle\phi\rangle_{V_1}^2+\epsilon^{-2}\|f\|_I^2).$$

Setting $\epsilon^{-2}=M'$, using $\|\phi\|_I\le N(E)\,\|f\|_I$ and rearranging, we get

$$\langle\phi\rangle_{V_1}^2\le[2|E+\lambda|N(E)^2+M'^2]\|f\|_I^2. \tag{36}$$

If we now substitute $f=T(E)^L(1+T(E))^{-1}g$ into (36) and use the facts that

$$\|(1+T(E))^{-1}g\|_I\le(1-N(E))^{-1}\|g\|_I,$$

which holds for $N(E)<1$, and

$$\|g\|_I\le|g|_{1,I}\le\rho_2\langle g\rangle_{V_1}$$

we have

$$\langle\phi\rangle_{V_1}\le C(E)[N(E)]^L\langle g\rangle_{V_1}. \tag{37}$$

Here, $C(E)$ is the accumulation of the various constants and depends on E, but not L. As $L\to\infty$, (37) implies that $\langle\phi\rangle_{V_1}\to0$ uniformly in g. Hence, the right side of (35) tends to zero in the uniform topology of $W_2{}^1(I)$. This establishes the convergence of (29) and, hence, (30) and (31).

Corollary 3.2: Let E_k be an eigenvalue of H_1, $N(E)$ be as in Theorem 6, M' as in (33), and $F=E+\lambda_1=|F|e^{i\alpha}$. Then,

$$N(E)\le M'\sup_{k\ge0}\left(\frac{E_k+\lambda_1}{|E_k-E|^2}\right)^{1/2}\le\frac{M'}{2}|F|^{-1/2}\csc\left(\frac{\alpha}{2}\right). \tag{38}$$

Hence, (29), (30), and (31) converge if one of the above is less than unity.

Proof: Using the definition of the norm of a bounded operator coupled with the fact that this coincides with the norm of its adjoint (see Ref. 6, p. 201), we have

$$N(E)=\sup_f\|T(E)^*f\|_I\quad(f\in L^2(I),\ \|f\|_I=1).$$

Since $T(E)^*=V(H_1-E)^{-1}$, we may use (33) to obtain a bound on $\|T(E)^*f\|_I$:

$$\|T(E)^*f\|_I\le M'\langle(H_1-\bar E)^{-1}f\rangle_{V_1}.$$

Letting the U_k's be the orthonormal eigenvectors of H_1 corresponding to E_k, we can expand f in the series

$$f=\sum_{k=0}^\infty A_kU_k,$$

where $A_k=(f,U_k)_I$ and $\Sigma|A_k|^2=1$. Applying $(H_1-\bar E)^{-1}$ to f and using Corollary 2.1,

$$\langle(H_1-\bar E)^{-1}f\rangle_{V_1}^2=\sum_{k=0}^\infty\frac{E_k+\lambda_1}{|E_k-\bar E|^2}|A_k|^2.$$

Hence,

$$\|T(E)^*f\|_I\le M'\sup_k\left(\frac{E_k+\lambda_1}{|E_k-E|^2}\right)^{1/2}.$$

Since the right side of the last inequality is independent of f, and since $|E_k-\bar E|=|E_k-E|$,

$$N(E)\le M'\sup_{k\ge0}\left(\frac{E_k+\lambda_1}{|E_k-E|^2}\right)^{1/2},$$

which is the lower half of (38). To obtain a bound on the right side of this inequality, we may use ordinary calculus to maximize

$$q(t)=(t/|t-F|^2)^{1/2},\quad t\ge0.$$

This gives the far right term in (38). Finally, by Theorem 3.6, if any one of the terms in (38) is less than unity, (29), (30), and (31) must converge. QED

We remark that Corollary 2 implies that by choosing $|F|$ large and fixing α, $0<\alpha<2\pi$, the expansions (28), (29), and (30) can always be made to converge.

If we replace V by γV, where γ is a real constant, Theorem 3.6 implies that the expansion

$$R_\gamma(E)=R_1(E)+\sum_{l=1}^\infty(-1)^l\gamma^l\,\tau T(E)^lU_1(E)$$

will converge for all γ such that

$$|\gamma|N(E)<1.$$

We note that if E is real, it is relatively easy to show that to each order of the coupling constant γ, the approximation to $R_\gamma(E)$ is self-adjoint. As Duke and Wigner[8] point out, a self-adjoint approximation to the R matrix always yields a unitary approximation to the collision matrix. Moreover, the approximation to the collision matrix will be in terms of rational functions because the R-matrix approximation is in terms of polynomials. (The usual Born approximation to the collision matrix is a polynomial approximation and is only approximately unitary. This would seem to be a disadvantage of the method.) Finally, since the approximation is rational, there is some hope that it remains valid even when the expansion for the R matrix is not.

4. CONCLUDING REMARKS

First of all, we wish to remark that the theory we have constructed can easily be extended to spin dependent systems, as long as the spin dependence in the Hamiltonian is confined to the operator V. Secondly, the R matrix potentially contains all the information required to solve eigenvalue problems of the form

$$(Q+V)U_k=E_kU_k$$
$$\partial_AU_k=bU_k|_S$$

or even,

$$U_k|_S=0.$$

Finally, it should be possible to extend the ideas behind the R matrix to higher-order partial differential equations, although this will surely require a modification of the approach we have used.

ACKNOWLEDGMENTS

The work presented here and in Paper I stems mainly from the author's thesis at Princeton University.[9] The author wishes to thank his advisor, Eugene P. Wigner, for much help and encouragement.

*Supported in part by the Air Force Office of Scientific Research under Grant No. 73-2484.

[1]A.M. Lane and R.G. Thomas, Rev. Mod. Phys. 30, 257 (1958).
[2]E.P. Wigner, Phys. Rev. 70, 15 (1946); 70, 606 (1946); with L. Eisenbud Phys. Rev. 72, 29 (1947).
[3]F.J. Narcowich, J. Math. Phys. 15, 1626 (1974).
[4]J.L. Lions and E. Magenes, *Problems and Limites non Homogènes et Applications* (Dunod, Paris, 1968), Vol I.
[5]M. Schechter, Comm. Pure Appl. Math 12, 457 (1959).
[6]F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955).
[7]H. Widom, *Lectures on Integral Equations* (Van Nostrand—Rheinhold, New York, 1969).
[8]C.B. Duke and E.P. Wigner, Rev. Mod. Phys. 36, 584 (1964).
[9]F.J. Narcowich, Ph.D dissertation, Princeton University, 1972 (unpublished).

The Clebsch–Gordan problem and coefficients for the three-dimensional Lorentz group in a continuous basis. III

N. Mukunda* and B. Radhakrishnan

Tata Institute of Fundamental Research, Bombay, 5, India
(Received 22 February 1973)

Along the lines of two previous papers, the Clebsch–Gordan problem for products of representations of $SU(1,1)$ of the form $D^+ \otimes C$ is related to the properties of the Lorentz group $O(3,1)$. The structure of the Clebsch–Gordan series for this case is understood in a new way as being due to the properties of $O(3,1)$ spherical harmonics on the timelike and spacelike hyperboloids in Minkowski space. The Clebsch–Gordan coefficients in a continuous basis are then evaluated.

INTRODUCTION

In the two previous papers[1] of this series, we have described a new approach to the Clebsch–Gordan problem for the three-dimensional Lorentz group $O(2,1)$. We are concerned here with direct products of unitary irreducible representations (UIR's) of this group, and with decomposing such direct products into direct sums of UIR's. Using a particular construction of the unitary representations of $O(2,1)$, we showed that the Clebsch–Gordan problem for this group can be related to properties of special representations of four-dimensional (pseudo) orthogonal groups. Thus for example, the reduction of products of the form $D^+ \otimes D^+$ (or $D^- \otimes D^-$) is achieved by analyzing the representation of the orthogonal group $O(4)$ carried by functions on the unit sphere in four (real) dimensions; and in the case of products of the form $D^+ \otimes D^-$ we were led to the group $O(2,2)$. In these two cases, the representations of the appropriate four-dimensional group were needed in an $O(2) \otimes O(2)$ basis, and our analysis led also to explicit expressions for the Clebsch–Gordan coefficients of $O(2,1)$ in an $O(1,1)$ basis.

The present paper is devoted to the analysis of direct product representations of the type $D^+ \otimes C$ and the related type $D^- \otimes C$, and to the computation of the related Clebsch–Gordan coefficients in the $O(1,1)$ basis. The "symmetry group" we shall be led to in the present case is the homogeneous Lorentz group $O(3,1)$; the solution of our problem entails the construction of a complete set of "spherical harmonics" for this group. There is a particular property of the group $O(3,1)$ that makes the analysis of products of the form $D^+ \otimes C$ specially interesting, in comparison to those of forms $D^\pm \otimes D^\pm$ and $C \otimes C$. In the latter cases, the symmetry groups one is led to are $O(4)$ and $O(2,2)$, with a suitable subgroup being singled out. (The case $C \otimes C$ will be analyzed in the next and concluding, paper of this sequence.) Now both the groups $O(4)$ and $O(2,2)$ can be expressed, locally, as direct products of "smaller" groups, namely one has $O(4) \approx O(3) \otimes O(3)$ and $O(2,2) \approx O(2,1) \otimes O(2,1)$. Making use of this fact, the problem of constructing spherical harmonics in these two cases simplifies a great deal: In fact, this construction is provided by the regular representations of $O(3)$ and $O(2,1)$, respectively. [In dealing with $D^+ \otimes D^-$ and $C \otimes C$, we need the $O(2,2)$ spherical harmonics in two different descriptions, and these are provided by the regular representation of $O(2,1)$ in two different descriptions.] In contrast, the group $O(3,1)$ does not break up in this way, so the construction of its spherical harmonics is decidedly nontrivial. With respect to $O(3,1)$, four-dimensional real space (Minkowski space) splits up into two distinct types of regions, the timelike and the spacelike regions, with very different properties. There is one set of spherical harmonics associated with each region. For the timelike region, they are relatively easy to construct,[2] since one can fall back upon the theory of the regular representation of $O(3,1)$. For the spacelike region, this is not the case, and the construction of the spherical harmonics is somewhat harder. It involves analyzing the representation of $O(3,1)$ associated with functions on the spacelike hyperboloid in Minkowski space, and we have carried out this analysis elsewhere.[3] The results of this analysis will be used here.

In Sec. 1 we set up the unitary representation $\mathcal{D}^+ \otimes \mathcal{C}$ of $O(2,1)$, the components $\mathcal{D}^+$ and $\mathcal{C}$ in the product being the generating representations for the UIR's D^+ and C_q^ϵ of $O(2,1)$. We show how the group $O(3,1)$ describes the symmetry properties of the representation $\mathcal{D}^+ \otimes \mathcal{C}$, set up the relations among the various Casimir operators, and specify the natures of the uncoupled and coupled basis vectors for the total Hilbert space, with whose help the C–G series and coefficients are to be determined. Section 2 explains the construction of a complete set of $O(3,1)$ spherical harmonics for the timelike regions in the space with metric $(+++-)$, while Sec. 3 contains the analogous steps for the spacelike region. With the help of these results, the two types of basis vectors are constructed in Sec. 4, and from their "quantum numbers" one immediately reads off the structure of the C–G series for a product of the form $D^+ \otimes C$. Section 5 calculates the C–G coefficients in the $O(1,1)$ basis for this kind of product and in Sec. 6 the related ones for $D^- \otimes C$ are given.

Appendix A describes the calculations concerning the normalization of wavefunctions belonging to the UIR $\{j_0, 0\}$ of $O(3,1)$. Appendix B contains the details regarding the determination of a phase associated with the occurrence of the UIR's (s,ϵ) in the product $D_k^+ \otimes C_q^\epsilon$.

1. THE REPRESENTATION $\mathcal{D}^+ \otimes \mathcal{C}$ OF SU(1,1)

Let us combine the two unitary representations $\mathcal{D}^+$ and $\mathcal{C}$ of $SU(1,1)$, acting in Hilbert spaces $\mathcal{H}(+,12)$ and $\mathcal{H}(\mathcal{C},34)$, respectively, into their direct product $\mathcal{D}^+ \otimes \mathcal{C}$. Here, 1 and 2 label the variables used in constructing the generators of $\mathcal{D}^+$, 3 and 4 those of $\mathcal{C}$, in the manner of Sec. II of Paper I. The space $\mathcal{H}$ for the product representation $\mathcal{D}^+ \otimes \mathcal{C}$ is the product $\mathcal{H}(+,12) \otimes \mathcal{H}(\mathcal{C},34)$ and so consists of functions $f(x_1, x_2, x_3, x_4)$ subject to

 Copyright © 1974 American Institute of Physics

$$\|f\|^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dx_1\,dx_2\,dx_3\,dx_4\,|f(x_1,x_2,x_3,x_4)|^2<\infty. \quad (1.1)$$

We then have the four sets of oscillator operators a_j, $a_j^\dagger$ obeying

$$[a_j,a_k^\dagger]=\delta_{jk},[a_j,a_k]=[a_j^\dagger,a_k^\dagger]=0,$$
$$a_j=\frac{-i}{\sqrt{2}}\left(x_j+\frac{\partial}{\partial x_j}\right),\quad a_j^\dagger=\frac{i}{\sqrt{2}}\left(x_j-\frac{\partial}{\partial x_j}\right),$$
$$j,k=1,2,3,4. \quad (1.2)$$

Using them, the three generators J_α of $\mathcal{D}^+\otimes\mathcal{C}$, which are sums of the individual generators $J_\alpha(+,12)$ and $J_\alpha(\mathcal{C},34)$, are

$$J_0=\tfrac{1}{2}(a_1^\dagger a_1+a_2^\dagger a_2+1+a_3^\dagger a_3-a_4^\dagger a_4),$$
$$J_1=\tfrac{1}{4}(a_1^\dagger a_1^\dagger+a_2^\dagger a_2^\dagger+a_1a_1+a_2a_2+a_3^\dagger a_3^\dagger-a_4^\dagger a_4^\dagger+a_3a_3-a_4a_4),$$
$$J_2=(-i/4)(a_1^\dagger a_1^\dagger+a_2^\dagger a_2^\dagger-a_1a_1-a_2a_2+a_3^\dagger a_3^\dagger+a_4^\dagger a_4^\dagger$$
$$-a_3a_3-a_4a_4). \quad (1.3)$$

The invariance properties of these generators become evident if in place of the a's and their adjoints, we work with operators b_μ, $b_\mu^\dagger$ and a metric tensor $g_{\mu\nu}$ as follows:

$$b_1=a_1,\quad b_2=a_2,\quad b_3=a_3,\quad b_4=-a_4^\dagger,$$
$$g_{11}=g_{22}=g_{33}=-g_{44}=1,\quad g_{\mu\nu}=0\quad\text{if }\mu\neq\nu. \quad (1.4)$$

The tensors $g_{\mu\nu}$, $g^{\mu\nu}$ will be used henceforth for lowering and raising indices. Equation (1.2) can then be transcribed into this form

$$[b_\mu,b_\nu^\dagger]=g_{\mu\nu},\quad [b_\mu,b_\nu]=[b_\mu^\dagger,b_\nu^\dagger]=0,$$
$$b_\mu=(-i/\sqrt{2})(x_\mu+\partial_\mu),\quad b_\mu^\dagger=(i/\sqrt{2})(x_\mu-\partial_\mu),$$
$$\partial_\mu\equiv\frac{\partial}{\partial x^\mu}. \quad (1.5)$$

At the same time, the J_α take on a simple appearance:

$$J_0=\tfrac{1}{2}(g^{\mu\nu}b_\mu^\dagger b_\nu+2),$$
$$J_1=\tfrac{1}{4}g^{\mu\nu}(b_\mu^\dagger b_\nu^\dagger+b_\mu b_\nu),$$
$$J_2=(-i/4)g^{\mu\nu}(b_\mu^\dagger b_\nu^\dagger-b_\mu b_\nu). \quad (1.6)$$

It is immediately evident from the two sets of equations above that both the basic commutation rules among the primitive variables b_μ, $b_\mu^\dagger$ and the forms of the generators J_α are preserved when we perform a real linear transformation

$$x_\mu\to O_\mu{}^\nu x_\nu,\quad b_\mu\to O_\mu{}^\nu b_\nu,\quad b_\mu^\dagger\to O_\mu{}^\nu b_\nu^\dagger \quad (1.7)$$

that leaves the indefinite quadratic form $x^2\equiv x^\mu x_\mu$ invariant. The space $\mathcal{H}$ thus carries a unitary representation of the group of the matrices $\|O_\mu{}^\nu\|$ and this representation commutes with the representation $\mathcal{D}^+\otimes\mathcal{C}$ of $SU(1,1)$. The group of matrices $\|O_\mu{}^\nu\|$ contains the identity component which we shall refer to as $O(3,1)$, and three other components containing improper transformations. The identity component is generated by six operators $M_{\mu\nu}$ which are

$$M_{\mu\nu}=-M_{\nu\mu}=i(b_\mu^\dagger b_\nu-b_\nu^\dagger b_\mu)=i(x_\mu\partial_\nu-x_\nu\partial_\mu). \quad (1.8)$$

Among the improper transformations it will suffice to consider these two:

$$\mathbb{R}:f(x_\mu)\to f(-x_\mu),$$
$$P_{34}:f(x_1,x_2,x_3,x_4)\to f(x_1,x_2,-x_3,-x_4). \quad (1.9)$$

These are not independent of one another in the sense that

$$\mathbb{R}=P_{34}\exp(i\pi M_{12}). \quad (1.10)$$

In other words, $\mathbb{R}$ and P_{34} belong to the same connected component of the full group of transformations (1.7). The operator $\mathbb{R}$ has the virtue of commuting with the transformations of $O(3,1)$, i.e.,

$$\mathbb{R}M_{\mu\nu}\mathbb{R}=M_{\mu\nu}. \quad (1.11)$$

The symmetry properties of the operators J_α that we shall use are thus summarized by

$$[J_\alpha,M_{\mu\nu}]=0,\quad \mathbb{R}J_\alpha\mathbb{R}=J_\alpha. \quad (1.12)$$

For the individual sets of generators of $\mathcal{D}^+$ and $\mathcal{C}$, we have only

$$[J_\alpha(+,12)\text{ or }J_\alpha(\mathcal{C},34),M_{12}\text{ or }M_{34}\text{ or }\mathbb{R}]=0. \quad (1.13)$$

Let us now establish the relations among the various Casimir operators. From the analysis in I we know that the $SU(1,1)$ Casimir operators belonging to the representations $\mathcal{D}^+$ and $\mathcal{C}$ are simple functions of M_{12} and M_{34}, respectively:

$$Q_{12}=\tfrac{1}{4}(1-M_{12}^2),\quad Q_{34}=\tfrac{1}{4}(1+M_{34}^2). \quad (1.14)$$

Turning to $O(3,1)$, it turns out as in the previous papers (Ref. 1) that we are dealing here with a special kind of representation of this group, in which one of the two independent Casimir invariants vanishes identically. Namely, from the form of $M_{\mu\nu}$ it follows that the invariant $\epsilon_{\mu\nu\rho\sigma}M^{\mu\nu}M^{\rho\sigma}=0$. The other $O(3,1)$ invariant is also the $SU(1,1)$ invariant Q for the "total" representation $\mathcal{D}^+\otimes\mathcal{C}$:

$$Q\equiv(J_1)^2+(J_2)^2-(J_0)^2=-\tfrac{1}{8}M^2\equiv-\tfrac{1}{8}M^{\mu\nu}M_{\mu\nu}. \quad (1.15)$$

We turn next to the two types of basis vectors for $\mathcal{H}$ that we wish to set up. Consider first the uncoupled basis. These vectors are direct products of basis vectors drawn from individual UIR's D_k^+ and C_q^ϵ picked out of $\mathcal{D}^+$ and $\mathcal{C}$, respectively. In setting up the basis within C_q^ϵ, we make use of the operator $\mathcal{A}_{34}$ that implements the outer automorphism τ of $SU(1,1)$, as explained in Sec. II of I. Recall that $\tau:J_0\to-J_0,J_1\to-J_1,J_2\to+J_2$ is implemented within the representation $\mathcal{C}$ acting on $\mathcal{H}(\mathcal{C},34)$ by the operator $\mathcal{A}_{34}$; thus

$$\mathcal{A}_{34}:f(x_3,x_4)\to f(x_4,x_3),$$
$$\mathcal{A}_{34}\{J_0(\mathcal{C},34),J_1(\mathcal{C},34),J_2(\mathcal{C},34)\}\mathcal{A}_{34}$$
$$=\{-J_0(\mathcal{C},34),-J_1(\mathcal{C},34),J_2(\mathcal{C},34)\}. \quad (1.16)$$

This same operator $\mathcal{A}_{34}$ has an obvious definition on $\mathcal{H}$; it is *not* a symmetry of the total generators J_α. Recall also that within the representation $\mathcal{C}$ on $\mathcal{H}(\mathcal{C},34)$, the eigenvalue $+1$ for the operator P_{34} [defined in Eq. (1.9)] selects continuous class UIR's C_q^ϵ of integral type, i.e., with $\epsilon=0$; while the eigenvalue -1 picks out the half-integral UIR's with $\epsilon=\frac{1}{2}$. In the uncoupled basis vectors for $\mathcal{H}$, then, the following mutually commuting operators should be diagonal: M_{12} (hence Q_{12}), $J_2(+,12)$; M_{34} (hence Q_{34}), $J_2(\mathcal{C},34)$, $\mathcal{A}_{34}$, and P_{34}. Because of the relation (1.10), we can replace P_{34} in this list by $\mathbb{R}$. This is helpful because $\mathbb{R}$ commutes with $O(3,1)$ generators $M_{\mu\nu}$ while P_{34} does not. The eigenvalue of $\mathbb{R}$ is simply correlated with the integral or half-integral nature of a product representation $D_k^+ \otimes C_q^\epsilon$. Since even (odd) values of M_{12} go with half-integral (integral) UIR's D_k^+ within $\mathcal{D}^+$, and because of the connection between P_{34} and ϵ noted above, Eq. (1.10) shows that in case $\mathbb{R}=-1$ the total representation of $SU(1,1)$ is of integral type, and with $\mathbb{R}=+1$ it is of half-integral type.

The coupled basis in $\mathcal{H}$ should consist of simultaneous eigenvectors for the operators M_{12} (hence Q_{12}), M_{34} (hence Q_{34}); R, M^2 (hence Q), J_2, and in case $Q>\frac{1}{4}$, also the operator implementing τ within the product representation. (Notice that the operators M_{12}, M_{34}, $\mathbb{R}$, and J_2 will be diagonal in both bases.) The first step towards construction of the coupled basis vectors is to find a complete set of $O(3,1)$-spherical harmonics in Minkowski space, which are also eigenfunctions of R; and in particular to set these up with M_{12} and M_{34} both diagonal. This will be done in the following two sections. We will conclude this section by recording the equations that express the "angular" dependences of J_α entirely in terms of the operator M^2 (or Q); the steps are identical to those leading to Eq. (1.16) of Paper II, and the results are

$$\begin{aligned} J_0 &= \tfrac{1}{4}(x^2-(1/x^2)(x\cdot\partial)^2-(2/x^2)x\cdot\partial-(4/x^2)Q),\\ J_1 &= -\tfrac{1}{4}(x^2+(1/x^2)(x\cdot\partial)^2+(2/x^2)x\cdot\partial+(4/x^2)Q),\\ J_2 &= (-i/2)(x\cdot\partial+2), \quad x\cdot\partial\equiv x^\mu\partial_\mu. \end{aligned} \tag{1.17}$$

2. O(3,1) SPHERICAL HARMONICS IN THE TIMELIKE REGION

With respect to the action of the group $O(3,1)$, Minkowski space splits up into three invariant regions: the positive timelike region V_1 in which $x^2<0$, $x_4>0$; the spacelike region V_2 where $x^2>0$; and the negative timelike region V_3 with $x^2<0$, $x_4<0$. (The light cone $x^2=0$ may be ignored since it is of lower dimensionality.) Correspondingly, each function $f(x)$ in $\mathcal{H}$ can be displayed as a column vector of three functions $f_1(x)$, $f_2(x)$, $f_3(x)$ giving the values of $f(x)$ for $x\in V_1, V_2, V_3$, respectively:

$$f=\begin{pmatrix} f_1(x)\\ f_2(x)\\ f_3(x)\end{pmatrix},$$

$$\|f\|^2=\int_{V_1}|f_1(x)|^2d^4x+\int_{V_2}|f_2(x)|^2d^4x+\int_{V_3}|f_3(x)|^2d^4x. \tag{2.1}$$

This also exhibits $\mathcal{H}$ as the direct sum of three mutually orthogonal subspaces $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ consisting of square-integrable functions that are nonvanishing in V_1, V_2, V_3, respectively:

$$\mathcal{H}=\mathcal{H}_1\oplus\mathcal{H}_2\oplus\mathcal{H}_3. \tag{2.2}$$

Each of these subspaces is invariant under $O(3,1)$; under $\mathbb{R}$, $\mathcal{H}_1$ and $\mathcal{H}_3$ get interchanged, while $\mathcal{H}_2$ is invariant. In both kinds of bases for $\mathcal{H}$ that we wish to construct, $\mathbb{R}$ will be diagonal. If $f(x)$ is an eigenfunction of $\mathbb{R}$, then it is fully specified by the pair of functions $f_1(x)$ and $f_2(x)$, with the former being unrestricted except for square-integrability and the latter being even or odd as the case may be:

$$\begin{aligned} \mathbb{R}=\pm 1 \Rightarrow f_3(x)&=\pm f_1(-x), \quad x\in V_3,\\ f_2(-x)&=\pm f_2(x), \quad x\in V_2. \end{aligned} \tag{2.3}$$

And for such elements in $\mathcal{H}$, the square of the norm is simply

$$\|f\|^2=2\int_{V_1}|f_1(x)|^2d^4x+\int_{V_2}|f_2(x)|^2d^4x. \tag{2.4}$$

We shall in the sequel deal only with eigenfunctions of R.

The problem of setting up $O(3,1)$ spherical harmonics for the regions V_1 and V_2 involves the following: First we must introduce suitable "radial" and "angular" coordinates for each region, such that only the latter are altered by the action of $O(3,1)$; next we must construct complete orthonormal bases of functions of the angular variables, such that they belong to definite UIR's of $O(3,1)$ and such that the angular dependences of elements $f_1(x)$ in $\mathcal{H}_1$ and $f_2(x)$ in $\mathcal{H}_2$ may be expanded in terms of the corresponding basis. The kinds of UIR's of $O(3,1)$ that one finds in this process are somewhat special. Recall that the UIR's of $O(3,1)$[4] can be labelled in the form $\{j_0,\rho\}$: j_0 can take on any of the values $0,1,2,\ldots$, while ρ is a real number. Except for the fact that $\{0,\rho\}$ and $\{0,-\rho\}$ denote equivalent UIR's a pair j_0, ρ labels one UIR uniquely. Now the connection between the Casimir invariants of $O(3,1)$ and the parameters j_0, ρ is the following:

$$\begin{aligned} M^2 &= M^{\mu\nu}M_{\mu\nu}=2(j_0^2-\rho^2-1),\\ \tfrac{1}{8}\epsilon_{\mu\nu\rho\sigma}M^{\mu\nu}M^{\rho\sigma} &= j_0\rho. \end{aligned} \tag{2.5}$$

But because of the special form of the $O(3,1)$ generators in Eq. (1.8), we have noted that the second invariant vanishes, so only the UIR's $\{j_0,0\}$ and $\{0,\rho\}$ will appear in our work. This is true for $\mathcal{H}_1$, as well as $\mathcal{H}_2$. [The supplementary series of UIR's of $O(3,1)$ are not relevant here.] It now turns out that in the subspace $\mathcal{H}_1$ containing functions over V_1 we encounter just the UIR's $\{0,\rho\}$ in the form of a direct integral from $\rho=0$ to $\rho=\infty$, with multiplicity one. This can be understood either from the structure of the regular representation of $O(3,1)$[2] or by the methods of integral geometry.[5] On the other hand, in the subspace of $\mathcal{H}_2$ consisting of functions with $\mathbb{R}=+1$, the UIR's $\{0,\rho\}$ appear as a direct integral with multiplicity one from $\rho=0$ to $\rho=\infty$, and in addition each of the UIR's $\{j_0,0\}$ for $j_0=2,4,\ldots$, appears once each; while if $\mathbb{R}=-1$, we have the same continuum of UIR's as before but the discrete component consists of $\{j_0,0\}$ for $j_0=1,3,5,\ldots$, once each. These results for $\mathcal{H}_2$ can be obtained using the methods of integral geometry.[3] (In all of this, we refer just to the angular dependences of functions in $\mathcal{H}_1$, and $\mathcal{H}_2$.) We must now set up spherical harmonics corresponding to these various UIR's, and labelled by the eigenvalues of M_{12} and M_{34} as well. There are two ways in which one can go about this job. One is to use the methods of integral geometry, the other is to write down differential equations obeyed by the required functions and find appropriate solutions. We will use a judicious mixture of both approaches. In particular, in dealing with the UIR's $\{0,\rho\}$ which appear both in $\mathcal{H}_1$ and in $\mathcal{H}_2$, we shall rely on the integral geometric approach,[5] since that will ensure that the corresponding basis functions transform in the same manner under $O(3,1)$, whether defined in V_1 or in V_2.

We proceed with the determination of the spherical functions for the region V_1. Since M_{12} and M_{34} are sought to be diagonal, we introduce new coordinates in V_1 as follows:

$$x_1 = r\sinh(\xi/2)\cos\psi, \quad x_2 = r\sinh(\xi/2)\sin\psi,$$
$$x_3 = r\cosh(\xi/2)\sinh\eta, \quad x_4 = r\cosh(\xi/2)\cosh\eta,$$
$$0 < r < \infty, \quad 0 \le \xi < \infty, \quad -\infty < \eta < \infty, \quad 0 \le \psi < 2\pi,$$
$$d^4x = \tfrac{1}{4} r^3\, dr \sinh\xi\, d\xi\, d\eta\, d\psi. \tag{2.6}$$

Then the operators M_{12} and M_{34} have the forms

$$M_{12} = i\frac{\partial}{\partial\psi}, \quad M_{34} = -i\frac{\partial}{\partial\eta}. \tag{2.7}$$

The operator M_{12} has exactly the same form as the corresponding operator introduced in I in order to reduce the representation $\mathcal{D}^+$ [see Eq. (2.12) of I]; similarly, M_{34} coincides with the operator S_{12} used in the analysis of the representation $\mathcal{C}$ [see Eqs. (2.33–35) of I]. To single out the particular product $D_k^+ \otimes C_q^\epsilon$ in $\mathcal{D}^+ \otimes \mathcal{C}$, where $q = \frac{1}{4} + s^2$ with $s > 0$, we must choose the eigenvalues of M_{12} and M_{34} to be $(2k-1)$ and $2s$, respectively. This determines the ψ and η dependences of the $O(3,1)$ spherical harmonic to be

$$\exp[-i(2k-1)\psi]\exp(2is\eta). \tag{2.8}$$

Now it is known that within a UIR $\{j_0, \rho\}$ of $O(3,1)$ the operators M_{12} and M_{34} form a complete commuting set, with the eigenvalues of M_{12} being all the integers from $-\infty$ to ∞, those of M_{34} being all real numbers.[6] The spherical harmonic belonging to the UIR $\{0, \rho\}$ is then just the factors in (2.8) times a definite function of ξ which will depend on k, s, and ρ as parameters. To determine this function we use the integral geometry approach.

In general, a function $f(\xi, \eta, \psi)$ can be thought of as a function on the positive timelike hyperboloid $x^2 = -1$, $x_4 > 0$. The $O(3,1)$ invariant measure on this hyperboloid is the part of the total measure d^4x in Eq. (2.6) that depends on ξ, η, and ψ. We will denote a point on this unit timelike hyperboloid by $\hat{x}$ so that we can interchangeably write $f(\hat{x})$ or $f(\xi, \eta, \psi)$. Then with each such (square-integrable) function $f(\hat{x})$ we associate a function $F(l;\rho)$, where l is a positive lightlike vector and $0 \le \rho < \infty$, according to

$$F(l;\rho) = \tfrac{1}{4}\int\int\int \sinh\xi\, d\xi\, d\eta\, d\psi\, f(\hat{x})(-\hat{x}\cdot l)^{i\rho-1}. \tag{2.9}$$

The function $F(l;\rho)$ is homogeneous in l of degree $(i\rho - 1)$, and when $f(\hat{x})$ is transformed under $O(3,1)$, $F(l;\rho)$ transforms by the UIR $\{0, \rho\}$. One can recover $f(\hat{x})$ from $F(l;\rho)$ and also express the scalar product of f with itself, in terms of F. All these properties are summarized by

$$F(al;\rho) = a^{-1+i\rho}F(l;\rho), \quad a > 0, \tag{2.10a}$$
$$f(\hat{x}) = (2\pi)^{-3}\int_0^\infty \rho^2 d\rho \int d\Omega(\hat{l})(-\hat{x}\cdot\hat{l})^{-i\rho-1}F(\hat{l};\rho), \tag{2.10b}$$
$$\tfrac{1}{4}\int\int\int \sinh\xi\, d\xi\, d\eta\, d\psi\, |f(\hat{x})|^2 = (2\pi)^{-3}\int_0^\infty \rho^2 d\rho \int d\Omega(\hat{l})\, |F(\hat{l};\rho)|^2. \tag{2.10c}$$

By $\hat{l}$ is meant a lightlike vector with $l_4 = 1$; $d\Omega(\hat{l})$ is the solid angle associated with the space-part of $\hat{l}$. By combining Eq. (2.10) with a proper choice of $F(l;\rho)$, we can obtain the spherical harmonics in the timelike region.

Now, on comparing Eqs. (1.15) and (2.5), it follows that the spherical harmonics for which the total $SU(1,1)$ Casimir operator Q has the eigenvalue $\frac{1}{4} + s'^2$ belong to the UIR $\{0, 2s'\}$ of $O(3,1)$. A convenient parametrization of l is

$$l_1 = t\cos\varphi, \quad l_2 = t\sin\varphi, \quad l_3 = t\sinh\zeta, \quad l_4 = t\cosh\zeta,$$
$$0 < t < \infty, \quad 0 \le \varphi < 2\pi, \quad -\infty < \zeta < \infty,$$
$$d\Omega(\hat{l}) = d\tanh\zeta\, d\varphi. \tag{2.11}$$

Then, in order to arrive at the $O(3,1)$ spherical harmonic belonging to the UIR $\{0, 2s'\}$ and also with $M_{12} = (2k-1)$ and $M_{34} = 2s$, we make the choice

$$F_{k,s;s'}(l;\rho) = (\sqrt{2\pi}/s')t^{-1+i\rho}\exp[-i(2k-1)\varphi] \times \exp(2is\zeta)\delta(\rho - 2s'). \tag{2.12}$$

For the associated function on the unit timelike hyperboloid we will write $\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi, \eta, \psi)$[7]; the negative sign in the superscript indicates that these functions belong to the region V_1 where $x^2 < 0$. Putting (2.12) into (2.10b) we get in the first instance

$$\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi) = \sqrt{2\pi}(s'/2\pi^3)\int_{-\infty}^\infty d\zeta \int_0^{2\pi} d\varphi \times \exp[2is\zeta - (2k-1)i\varphi] \times [\cosh(\xi/2)\cosh(\zeta-\eta) - \sinh(\xi/2)\cos(\varphi-\psi)]^{-1-2is'}. \tag{2.13}$$

By shifting ζ and φ the expected factors in (2.8) can be separated and we get

$$\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi) = \sqrt{2\pi}(s'/2\pi^3)\exp[2is\eta - i(2k-1)\psi] \times \int_{-\infty}^\infty d\zeta \int_0^{2\pi} d\varphi\, \exp[2is\zeta - i(2k-1)\varphi \times [\cosh(\xi/2)\cosh\zeta - \sinh(\xi/2)\cos\varphi]^{-1-2is'}. \tag{2.14}$$

The integrand can be expanded using the binominal theorem as

$$[\cosh(\xi/2)\cosh\zeta - \sinh(\xi/2)\cos\varphi]^{-1-2is'} = [\cosh(\xi/2)\cosh\zeta]^{-1-2is'}\sum_{n=0}^\infty \frac{(-1)^n}{n!}\frac{\Gamma(-2is')}{\Gamma(-n-2is')} \times \left(\tanh\frac{\xi}{2}\frac{\cos\varphi}{\cosh\zeta}\right)^n. \tag{2.15}$$

Interchanging the sum on n with the ζ and φ integrations, we find that these two integrations factorize, and each can be done using standard formulas.[8] The remaining sum on n is then recognized as leading to a hypergeometric function in the variable $\tanh^2(\xi/2)$, and all in all we get

$$\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi) = \sqrt{2/\pi}(s'/\pi)2^{2is'}\exp[2is\eta - i(2k-1)\psi] \times [\Gamma(k+is'+is)\Gamma(k+is'-is)/\Gamma(2k)\Gamma(1+2is')] \times [\cosh(\xi/2)]^{-2is'-2k}[\sinh(\xi/2)]^{2k-1} \times F(k+is'+is, k+is'-is; 2k; \tanh^2(\xi/2)). \tag{2.16}$$

This final result can also be recast in a form where the argument of the hypergeometric function is $-\sinh^2(\xi/2)$:

$$\begin{aligned}&\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi)\\&=-2i(2\pi)^{-3/2}2^{2is'}\exp[2is\eta-i(2k-1)\psi]\\&\quad\times[\Gamma(k+is'+is)\Gamma(k+is'-is)/\Gamma(2k)\Gamma(2is')]\\&\quad\times[\cosh(\xi/2)]^{2is}[\sinh(\xi/2)]^{2k-1}\\&\quad\times F(k+is+is',k+is-is';2k;-\sinh^2(\xi/2)).\end{aligned}\tag{2.17}$$

The normalization properties of these spherical harmonics are determined by using Eqs. (2.10c), (2.12); indeed, the factors in Eq. (2.12) were chosen so that we may have the following properties:

$$\begin{aligned}&\tfrac{1}{4}\int_0^\infty \sinh\xi\,d\xi\int_{-\infty}^\infty d\eta\int_0^{2\pi}d\psi\,\mathcal{Y}^{-(s'''\epsilon''')}_{2k'-1,2s''}(\xi,\eta,\psi)^*\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi)\\&\quad=\delta(s'''-s')\delta(s''-s)\delta_{k'k}.\end{aligned}\tag{2.18}$$

With this, the problem of constructing a complete set of $O(3,1)$ spherical harmonics for the timelike region V_1, with M^2, M_{12}, and M_{34} diagonal, is solved. Now we turn to the construction of a similar complete set for the region V_2.

3. O(3,1) SPHERICAL HARMONICS IN THE SPACELIKE REGION

This case is somewhat more complicated than the previous one, and one has to combine the two approaches involving differential equations and integral geometry. To begin with, different kinds of coordinates have to be introduced in the two regions of V_2 corresponding to $x_3^2 \gtrless x_4^2$. We define them as follows:

$$\begin{aligned}&V_2^{(1)}: x_4^2>x_3^2,\\&\quad x_1=r\cosh(\xi/2)\cos\psi,\quad x_2=r\cosh(\xi/2)\sin\psi,\\&\quad x_3=r\sinh(\xi/2)\sinh\eta,\quad x_4=r\sinh(\xi/2)\cosh\eta,\\&\quad 0<r<\infty,\quad -\infty<\xi<\infty,\quad -\infty<\eta<\infty,\quad 0\le\psi<2\pi,\\&\quad d^4x=\tfrac{1}{4}r^3\,dr\,|\sinh\xi|\,d\xi\,d\eta\,d\psi;\end{aligned}\tag{3.1a}$$

$$\begin{aligned}&V_2^{(2)}: x_3^2>x_4^2,\\&\quad x_1=r\cos(\theta/2)\cos\psi,\quad x_2=r\cos(\theta/2)\sin\psi,\\&\quad x_3=r\sin(\theta/2)\cosh\eta,\quad x_4=r\sin(\theta/2)\sinh\eta,\\&\quad 0<r<\infty,\quad -\pi\le\theta\le\pi,\quad -\infty<\eta<\infty,\quad 0\le\psi<2\pi,\\&\quad d^4x=\tfrac{1}{4}r^3dr\,|\sin\theta|\,d\theta\,d\eta\,d\psi.\end{aligned}\tag{3.1b}$$

Then, a function $f_2(x)$ defined for $x\in V_2$ has to be thought of as a pair of functions, one defined in $V_2^{(1)}$ and the other in $V_2^{(2)}$, and we have

$$\begin{aligned}&f_2(x)=\{f_2^{(1)}(r;\xi,\eta,\psi);f_2^{(2)}(r;\theta,\eta,\psi)\},\\&\int_{V_2}|f_2(x)|^2d^4x=\tfrac{1}{4}\int_{-\infty}^\infty d\eta\int_0^{2\pi}d\psi\int_0^\infty r^3\,dr\\&\qquad\times\Big\{\int_{-\infty}^\infty|\sinh\xi|\,|f_2^{(1)}(r;\xi,\eta,\psi)|^2\,d\xi\\&\qquad+\int_{-\pi}^{\pi}|\sin\theta|\,|f_2^{(2)}(r;\theta,\eta,\psi)|^2\,d\theta\Big\}.\end{aligned}\tag{3.2}$$

The $\mathbb{R}$ operation can be described in the new variables as follows:

$$\begin{aligned}\mathbb{R}:x\to-x:\ &\xi\to-\xi,\quad\psi\to\psi+\pi,\quad\eta\to\eta\ \text{ in } V_2^{(1)}\\&\theta\to-\theta,\quad\psi\to\psi+\pi,\quad\eta\to\eta\ \text{ in } V_2^{(2)}.\end{aligned}\tag{3.3}$$

If $f_2(x)$ is an eigenfunction of $\mathbb{R}$, then in place of (3.2) we have a simpler form for the squared norm of $f_2(x)$.

Now the expressions for M_{12} and M_{34} are uniformly given in both $V_2^{(1)}$ and $V_2^{(2)}$ by Eq. (2.7) again, while M^2 is given as

$$\begin{aligned}\tfrac{1}{8}M^2&=\frac{\partial^2}{\partial\xi^2}+\coth\xi\frac{\partial}{\partial\xi}+\frac{1}{\sinh^2\xi}\left\{\frac{1-\cosh\xi}{2}\frac{\partial^2}{\partial\psi^2}-\frac{1+\cosh\xi}{2}\frac{\partial^2}{\partial\eta^2}\right\},\\&=-\frac{\partial^2}{\partial\theta^2}-\cot\theta\frac{\partial}{\partial\theta}-\frac{1}{\sin^2\theta}\left\{\frac{1-\cos\theta}{2}\frac{\partial^2}{\partial\psi^2}-\frac{1+\cos\theta}{2}\frac{\partial^2}{\partial\eta^2}\right\},\quad\text{in } V_2^{(1)}\text{ and } V_2^{(2)},\text{ respectively.}\end{aligned}\tag{3.4}$$

Once again the requirement that we choose eigenfunctions of M_{12} and M_{34} with eigenvalues $(2k-1)$ and $2s$, respectively, in order to isolate the product $D_k^+\otimes C^\epsilon_{(1/4)+s^2}$ from $\mathcal{D}^+\otimes\mathcal{C}$ fixes the ψ and η dependences of the functions $f_2^{(1)}$ and $f_2^{(2)}$:

$$f_2(x)\equiv\exp[-i(2k-1)\psi-2is\eta]\{f_2^{(1)}(r;\xi);f_2^{(2)}(r;\theta)\}.\tag{3.5}$$

Now the functions $f_2^{(1)}(r;\xi)$ and $f_2^{(2)}(r;\theta)$ must be eigenfunctions of $Q=-\frac{1}{8}M^2$ with appropriate eigenvalues if they are to be $O(3,1)$ spherical harmonics and this gives us two differential equations in ξ and θ [remember that r is an $O(3,1)$ invariant] for these functions via Eq. (3.4). We shall have to use these differential equations when we want to set up $O(3,1)$ spherical harmonics belonging to the UIR $\{j_0,0\}$. But before that let us take up the comparatively simpler case of the spherical harmonics belonging to the UIR $\{0,\rho\}$ which will serve as basis vectors for the UIR $C^{\epsilon'}_{q'}$ of $SU(1,1)$ obtained in the reduction of $D^+\otimes C^\epsilon_q$. For the construction of these spherical harmonics we prefer to use the method of integral geometry (rather than solve the differential equations in the two regions $V_2^{(1)}$ and $V_2^{(2)}$ and match solutions at the boundary) not only because it is simpler but also because it gives us a way of ensuring that the spherical harmonics on the timelike and spacelike hyperboloids have identical transformation properties under $O(3,1)$. This is a necessary condition to have in our formalism.

The reduction of the space of (square-integrable) functions on the unit spacelike hyperboloid into UIR's of $O(3,1)$ have been performed by us elsewhere[3] and we shall only quote our results here.

Let $f(\hat{x})$ be a (square-integrable) function on the unit spacelike hyperboloid ($\hat{x}^2=1$). Without loss of generality we can choose it to be an eigenfunction of R with eigenvalue $\epsilon=\pm1$. With every such function $f_\epsilon(\hat{x})$ we can associate two functions: $F_\epsilon(l;\rho)$ defined on the positive light cone $l^2=0$, $l_4>0$, and transforming via the UIR $\{0,\rho\}$ of $O(3,1)$ and $F(\vec{l},\vec{b};n)$ which is a function of two mutually orthogonal unit 3-vectors $\vec{l}$ and $\vec{b}$, and transforming under $O(3,1)$ via the UIR $\{n,0\}$. ($\vec{l}$ and $\vec{b}$ are actually the space parts of a lightlike vector $\hat{l}$ and a spacelike vector b satisfying: $\hat{l}^2=0$, $\hat{l}_4=1$, $b^2=1$, $b_4=0$, and $l\cdot b=0$; see Refs. 3 and 5.) n runs over all positive even integers for $\epsilon=+1$ and over all positive odd integers for $\epsilon=-1$. These functions are defined in terms of $f_\epsilon(\hat{x})$ as follows:

$$F_\epsilon(l;\rho)=\int(dx)f_\epsilon(\hat{x})\,|\hat{x}\cdot l|^{-1+i\rho},\tag{3.6a}$$

$$F(\mathbf{l},\mathbf{b};n)=\tfrac{1}{2}\int(dx)f_\epsilon(\hat{x})\exp(in\alpha)\delta(\hat{\mathbf{x}}\cdot\hat{\mathbf{1}}-x_4).\tag{3.6b}$$

(dx) stands for the $O(3,1)$ invariant measure on the spacelike hyperboloid and the angle α in Eq. (3.6b) ($0 \leq \alpha < 2\pi$) is to be determined from the relation

$$\mathbf{x} - x_4 \mathbf{l} = R(\alpha;\mathbf{l})\mathbf{b}, \tag{3.7}$$

where $R(\alpha;\mathbf{l})$ denotes a rotation through an angle α about the direction of $\mathbf{l}$. The basic result of integral geometry is that the part of $f_\epsilon(\hat{x})$ that transforms via the UIR's $\{0,\rho\}$, $0 \leq \rho < \infty$, can be obtained (in fully reduced form) in terms of $F_\epsilon(l;\rho)$ and the part transforming via the UIR's $\{n,0\}$, $n=1,2,\cdots$, can similarly be expressed in terms of $F(\mathbf{l},\mathbf{b};n)$. Moreover, the squared norm of f_ϵ can also be written down in terms of the latter two functions:

$$f_\epsilon(\hat{x}) = (16\pi^3)^{-1}\int_0^\infty \rho^2 d\rho \int (dl) F_\epsilon(l;\rho)[\delta(\hat{x}\cdot l + 1) + \epsilon\delta(\hat{x}\cdot l - 1)] + (2/\pi^2)\sum_n n \int d\Omega(\mathbf{l}) F(\mathbf{l},\mathbf{b};n) \times \delta(\mathbf{x}\cdot\mathbf{l} - x_4), \tag{3.8a}$$

$$\int (dx) |f_\epsilon(\hat{x})|^2 = (16\pi^3)^{-1}\int_0^\infty \rho^2 d\rho \int d\Omega(\hat{\mathbf{l}}) |F_\epsilon(\hat{l};\rho)|^2 + (4/\pi^2)\sum_n n\int d\Omega(\mathbf{l})|F(\mathbf{l},\mathbf{b};n)|^2. \tag{3.8b}$$

(dl) is the $O(3,1)$ invariant measure on the light cone and $d\Omega(\hat{\mathbf{l}})$ the solid angle element associated with the direction of $\mathbf{l}$. [In Eq. (3.8b), $F_\epsilon(\hat{l};\rho)$ stands for the restriction of $F_\epsilon(l;\rho)$ to lightlike vectors $\hat{l}$ with $\hat{l}_4 = 1$.] From these equations it can be seen that in order to obtain the $O(3,1)$ spherical harmonic belonging to the UIR $\{0, 2s'\}$ in a basis where $\mathbf{R} = \epsilon$, $M_{12} = (2k-1)$, and $M_{34} = 2s$ we must set

$$F_\epsilon(l;\rho) = F_{k,s;s'}(t,\zeta,\varphi;\rho) = (\sqrt{4\pi}/s')t^{-1+2is'}\exp[2is\zeta - i(2k-1)\varphi]\delta(\rho - 2s'), \tag{3.9a}$$

$$F(\mathbf{l},\mathbf{b};n) \equiv 0. \tag{3.9b}$$

In Eq. (3.9a) we have adopted the same parameterization of l as the one given in Eq. (2.11). For the associated spherical harmonics in $V_2^{(1)}$ and $V_2^{(2)}$ we shall write $\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi)$ and $\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi)$, respectively. The superscript "+" indicates that these functions are defined on the spacelike hyperboloid $\hat{x}^2 = 1$ and for future convenience we have defined ϵ' as

$$\epsilon' = \tfrac{1}{2} \quad \text{for } \mathbf{R} = \epsilon = +1$$
$$= 0 \quad \text{for } \mathbf{R} = \epsilon = -1. \tag{3.10}$$

Before we proceed to evaluate these functions, a word about their transformation properties: $\mathcal{Y}^{+(s'\epsilon')}_{2k-1,2s}$ has the same transformation property under $O(3,1)$ as $\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}$ defined in Eq. (2.13) since both these functions are defined in a *covariant* fashion in terms of the *same* function $F(l;\rho)$ on the light cone. An $O(3,1)$ transformation on both the $\mathcal{Y}$'s can therefore be translated into the corresponding transformation acting on $F(l;\rho)$. Substituting Eq. (3.9) in Eq. (3.8a) we find:

$$\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi) = \sqrt{4\pi}(s'/4\pi^3)\int_0^\infty dt\, t^{2is'}\int_{-\infty}^\infty d\zeta \exp(2is\zeta) \times \int_0^{2\pi} d\varphi \exp[-i(2k-1)\varphi]\{\delta[t\cosh(\xi/2) \times\cos(\psi-\varphi) - t\sinh(\xi/2)\cosh(\eta-\zeta)+1] + \epsilon\delta[t\cosh(\xi/2)\cos(\psi-\varphi) - t\sinh(\xi/2) \times\cosh(\eta-\zeta)-1]\}, \tag{3.11a}$$

$$\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi) = \sqrt{4\pi}(s'/4\pi^3)\int_0^\infty dt\, t^{2is'}\int_{-\infty}^\infty d\zeta\exp(2is\zeta) \times\int_0^{2\pi} d\varphi\exp[-i(2k-1)\varphi]\{\delta[t\cos(\theta/2) \times\cos(\psi-\varphi) - t\sin(\theta/2)\sinh(\eta-\zeta)+1] + \epsilon\delta[t\cos(\theta/2)\cos(\psi-\varphi) - t\sin(\theta/2) \times\cosh(\eta-\zeta)-1]\}. \tag{3.11b}$$

By shifting ζ and φ the expected factors of η and ψ [given in Eq. (2.8)] can be extracted. After a change of variable from t to $z = 1/t$ we obtain

$$\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi) = \sqrt{4\pi}(s'/4\pi^3)\exp[2is\eta - i(2k-1)\psi] \times\int_0^\infty dz\, z^{-1-2is'}\int_{-\infty}^\infty d\zeta\exp(-2is\zeta) \times\int_0^{2\pi} d\varphi\exp[i(2k-1)\varphi]\{\delta[\cosh(\xi/2) \times\cos\varphi - \sinh(\xi/2)\cosh\zeta + z] + \epsilon\delta[\cosh(\xi/2)\cos\varphi - \sinh(\xi/2)\cosh\zeta - z]\}, \tag{3.12a}$$

$$\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi) = \sqrt{4\pi}(s'/4\pi^3)\exp[2is\eta - i(2k-1)\psi] \times\int_0^\infty dz\, z^{-1-2is'}\int_{-\infty}^\infty d\zeta\exp(-2is\zeta) \times\int_0^{2\pi} d\varphi\exp[i(2k-1)\varphi]\{\delta[\cos(\theta/2)\cos\varphi - \sin(\theta/2)\sinh\zeta + z] + \epsilon\delta[\cos(\theta/2)\cos\varphi - \sin(\theta/2)\sinh\zeta - z]\}. \tag{3.12b}$$

In order to carry out these integrations we employ the plane-wave representation for the delta function:

$$\delta(p) = (1/2\pi)\int_{-\infty}^\infty \exp(ipy)\,dy. \tag{3.13}$$

Using this representation in Eqs. (3.12a) and (3.12b) one finds that the z, ζ, and φ integrations factorize and can be done using standard formulas.[9] One is then left with a final y-integration over a product of two (generalized) Bessel functions and a power of y which can also be evaluated.[10] The final result is

$$\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi) = -\sqrt{4\pi}(s'/8\pi^3)2^{2is'}\Gamma(-2is')[\exp(-s'\pi) + \epsilon\exp(s'\pi)]\exp[2is\eta - i(2k-1)\varphi] \times\Big(\{\exp[s\pi - i(2k-1)(\pi/2)] + \epsilon\exp[-s\pi + i(2k-1)(\pi/2)]\}\frac{\Gamma(k+is+is')}{\Gamma(k-is-is')} \times\Gamma(-2is)[\sinh(\xi/2)]^{2is}[\cosh(\xi/2)]^{2k-1} \times F(k+is+is', k+is-is'; 1+2is; -\sinh^2(\xi/2)) + (s\to -s)\Big), \quad \text{for } \xi\geq 0, \tag{3.14a}$$

$$\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(-\xi,\eta,\psi) = \epsilon(-1)^{2k-1}\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi), \tag{3.14b}$$

$$\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi) = \sqrt{4\pi}(s'/8\pi^3)2^{2is'}\Gamma(-2is')[\exp(s'\pi) + \epsilon\exp(-s'\pi)]\exp[2is\eta - i(2k-1)\psi] \times\{\exp[-s\pi + i(2k-1)(\pi/2)] + \epsilon\exp[s\pi - i(2k-1)(\pi/2)]\}\Big(\frac{\Gamma(k+is+is')}{\Gamma(k-is-is')} \times[\sin(\theta/2)]^{2is}[\cos(\theta/2)]^{2k-1} \times F(k+is+is', k+is-is'; 1+2is; \sin^2(\theta/2)) + (s\to -s)\Big), \quad \text{for } \theta\geq 0, \tag{3.14c}$$

$$\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(-\theta,\eta,\psi) = \epsilon(-1)^{2k-1}\mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi). \tag{3.14d}$$

In writing Eqs. (3.14b) and (3.14d) we have made use of (3.3). The normalization of these spherical harmonics is determined by Eq. (3.8b). We have in fact chosen the factors in Eq. (3.9a) so as to have

$$\tfrac{1}{4}\int_{-\infty}^{\infty} d\eta \int_0^{2\pi} d\psi \Big[\int_{-\infty}^{\infty} d\xi\,|\sinh\xi|\,\mathcal{Y}^{+(s'''\epsilon''')1}_{2k'-1,2s'}(\xi,\eta,\psi)^* \times \mathcal{Y}^{+(s''\epsilon'')1}_{2k-1,2s}(\xi,\eta,\psi) + \int_{-\pi}^{\pi} d\theta\,|\sin\theta|\,\mathcal{Y}^{+(s'''\epsilon''')2}_{2k'-1,2s'}(\theta,\eta,\psi)^* \times \mathcal{Y}^{+(s'\epsilon'')2}_{2k-1,2s}(\theta,\eta,\psi)\Big] = \delta(s'''-s'')\delta_{\epsilon'''\epsilon}\,\delta(s'-s)\delta_{k'k}. \tag{3.15}$$

This completes the construction of the $O(3,1)$ spherical harmonics belonging to the UIR $\{0, 2s'\}$. We now turn to the problem of finding the remaining discrete set of spherical harmonics that appear on the spacelike hyperboloid.

The construction of $O(3,1)$ spherical harmonics belonging to the UIR's $\{n,0\}$ for $n=1,2,3,\cdots$ is somewhat more involved than the previous case. Here it is difficult to make direct use of the integral representation for the spherical harmonics that we get from the method of integral geometry and we will find it profitable to start instead with the differential equations for these functions. As before the η and ψ dependences are of the standard form

$$\mathcal{Y}^{+(k')1}_{2k-1,2s}(\xi,\eta,\psi) = \exp[2is\eta - i(2k-1)\psi]\,Y^{+(k')1}_{2k-1,2s}(\xi), \tag{3.16a}$$

$$\mathcal{Y}^{+(k')2}_{2k-1,2s}(\theta,\eta,\psi) = \exp[2is\eta - i(2k-1)\psi]\,Y^{+(k')2}_{2k-1,2s}(\theta), \tag{3.16b}$$

where $n=2k'-1$. The functions Y^+ are solutions of the following second-order differential equations:

$$\left(\frac{d^2}{d\xi^2} + \coth\xi\frac{d}{d\xi} + \frac{1}{4} - \frac{(2k'-1)^2}{4} + \frac{1}{\sinh^2\xi} \times (\lambda^2 + \lambda^{*2} + 2\lambda\lambda^*\cosh\xi)\right) Y^{+(k')1}_{2k-1,2s}(\xi) = 0, \tag{3.17a}$$

$$\left(\frac{d^2}{d\theta^2} + \cot\theta\frac{d}{d\theta} - \frac{1}{4} + \frac{(2k'-1)^2}{4} + \frac{1}{\sin^2\theta}(\lambda^2 + \lambda^{*2} + 2\lambda\lambda^*\cos\theta)\right) \times Y^{+(k')2}_{2k-1,2s}(\theta) = 0, \tag{3.17b}$$

where $\lambda \equiv s + i(k-\frac{1}{2})$. Further, these functions have to be normalizable and mutually orthogonal for different values of k' but fixed k and s.

Let us first consider Eq. (3.17a). If we denote by $\phi(a,b;c;z)$ a general solution of the hypergeometric equation in z with parameters a, b, and c, it is straightforward to verify that a solution of Eq. (3.17a) is of the form

$$Y^{+(k')1}_{2k-1,2s}(\xi) = [\cosh(\xi/2)]^{-(2k-1)}[\sinh^2(\xi/2)]^{-is} \times \phi(\tfrac{3}{2} - k - k' - is, \tfrac{1}{2} - k + k' - is; 1 - 2is; -\sinh^2(\xi/2)). \tag{3.18}$$

The specific solution ϕ that we must choose is determined by the condition or normalizability:

$$\int_0^\infty d\xi \sinh\xi\,|Y^{+(k')1}_{2k-1,2s}(\xi)|^2 < \infty. \tag{3.19}$$

[Here we have restricted the range of ξ taking account of the fact that for $\mathbb{R}=\epsilon$, $Y^{+(k')1}_{2k-1,2s}(-\xi) = \epsilon(-1)^{2k-1}Y^{+(k')1}_{2k-1,2s}(\xi)$.] Now it is easily checked that any two linearly independent solutions ϕ with the indicated parameters are well behaved at $\xi=0$. However, for $\xi\to\infty$, one sees from the indicial equation that one solution goes like a constant, while the other goes as $(\sinh^2(\xi/2))^{2k'-1}$. Clearly, only the former solution is acceptable and thus we know $Y^{+(k')1}$ up to a normalization constant:

$$Y^{+(k')1}_{2k-1,2s}(\xi) = N_1(\cosh(\xi/2))^{-(2k-1)} \times \left(\frac{\Gamma(2is)\Gamma(2k')}{\Gamma(k+k'-\frac{1}{2}+is)\Gamma(k'-k+\frac{1}{2}+is)}(\sinh^2(\xi/2))^{-is} \times F(\tfrac{3}{2}-k-k'-is, \tfrac{1}{2}-k+k'-is; 1-2is; -\sinh^2(\xi/2)) + (s\to -s)\right). \tag{3.20}$$

The determination of N_1 involves some calculation which we carry out in Appendix A. The result is

$$N_1 = (-1)^{2k'-1}2(2k'-1)^{1/2}\frac{\Gamma(k+k'-\frac{1}{2}+is)}{\Gamma(2k')\Gamma(k-k'+\frac{1}{2}+is)}. \tag{3.21}$$

Turning now to the case of $Y^{+(k')2}_{2k-1,2s}(\theta)$ it can be checked from Eq. (3.17b) that it must be of the form

$$Y^{+(k')2}_{2k-1,2s}(\theta) = [\cos(\theta/2)]^{-(2k-1)}[\sin^2(\theta/2)]^{-is} \times \phi(\tfrac{3}{2}-k-k'-is, \tfrac{1}{2}-k+k'-is; 1-2is; \sin^2(\theta/2)). \tag{3.22}$$

The condition of normalizability says

$$\int_0^\pi d\theta \sin\theta\,|Y^{+(k')2}_{2k-1,2s}(\theta)|^2 < \infty. \tag{3.23}$$

[As before we have restricted the range of θ using: $\mathbb{R}=\epsilon$, $Y^{+(k')2}_{2k-1,2s}(-\theta) = \epsilon(-1)^{2k-1}Y^{+(k')2}_{2k-1,2s}(\theta)$.] ϕ is necessarily well behaved at $\theta=0$ for the indicated parameters while for $\theta=\pi$ the indicial equation shows that one solution goes as a constant while the other goes like $[\cos^2(\theta/2)]^{2k-1}$. Now only the latter solution is acceptable and thus we know $Y^{+(k')2}$ up to a normalization factor:

$$Y^{+(k')2}_{2k-1,2s}(\theta) = N_2[\cos(\theta/2)]^{2k-1} \times \left(\frac{\Gamma(2is)\Gamma(2k)}{\Gamma(k+k'-\frac{1}{2}+is)\Gamma(\frac{1}{2}-k+k'+is)}[\sin^2(\theta/2)]^{-is} \times F(\tfrac{1}{2}+k-k'-is, k+k'-\tfrac{1}{2}-is; 1-2is; \sin^2(\theta/2)) + (s\to -s)\right). \tag{3.24}$$

The determination of N_2 is again relegated to Appendix A. We find

$$N_2 = (-1)^{2k'-1}2(2k'-1)^{1/2}\frac{\Gamma(k+k'-\frac{1}{2}+is)}{\Gamma(2k)\Gamma(k-k'+\frac{1}{2}+is)}. \tag{3.25}$$

Finally, a word about the normalization. The restriction of the ranges of ξ and θ using the symmetry under R and the fact that for $\mathrm{R}=+1$ we have k' half-odd integral while for $\mathrm{R}=-1$ we have k' integral imply that

$$\tfrac{1}{2}\Big[\int_0^\infty d\xi \sinh\xi\, Y^{+(k'')1}_{2k-1,2s}(\xi)^* Y^{+(k')1}_{2k-1,2s}(\xi) + \int_0^\pi d\theta \sin\theta\, Y^{+(k'')2}_{2k-1,2s}(\theta)^* Y^{+(k')2}_{2k-1,2s}(\theta)\Big] = \delta_{k''k'} \tag{3.26}$$

only if either k' and k'' are both integral or they are both half-odd integral. The above equation will not in general hold if one of k' and k'' is integral while the other is half-odd integral.

This completes the task of setting up a complete set of spherical harmonics on the spacelike hyperboloid $x^2=1$ in a basis where $\mathbb{R}=\epsilon$, $M_{12}=2k-1$, $M_{34}=2s$, and $Q=-\frac{1}{8}M^2=\frac{1}{4}+s'^2$ or $k'(1-k')$.

4. BASIS VECTORS FOR $\mathscr{H}$ AND THE C-G SERIES FOR $D^+ \otimes C$

In this section we shall obtain the two types of basis vectors that we had specified in Sec. 1. Then, from the structure of coupled basis vectors we shall directly read off the Clebsch—Gordan series for the product $D_k^+ \otimes C_q^\epsilon$.

We start with the uncoupled basis vectors Φ. For the product $D_k^+ \otimes C^\epsilon_{1/4+s^2}$ we shall use the notation $\Phi_{p\ \ p'a}^{(k+)\,(s\epsilon)}$, where $p(p')$ is the eigenvalue of $J_2(+,12)(J_2(\mathcal{C},34))$ and the eigenvalue of $\mathcal{A}_{34}$. Such a vector will be the product of a function of x_1, x_2 and another function of x_3, x_4. Using the analyses of the representations $\mathcal{D}^+$ and $\mathcal{C}$ of $SU(1,1)$ given in Sec. II of I, we see that we must introduce trigonometric variables for x_1, x_2 as in Eq. (2.12) of I and hyperbolic variables for x_3, x_4 as in Eq. (2.33) of I:

$$x_1=\rho\cos\varphi, \quad x_2=\rho\sin\varphi, \tag{4.1a}$$

$$|x_4|>|x_3|: x_3=(\operatorname{sgn}x_4)\rho'\sinh\alpha, \quad x_4=(\operatorname{sgn}x_4)\rho'\cosh\alpha, \tag{4.1b}$$

$$|x_3|>|x_4|: x_3=(\operatorname{sgn}x_3)\rho'\cosh\alpha, \quad x_4=(\operatorname{sgn}x_3)\rho'\sinh\alpha. \tag{4.1c}$$

Then apart from numerical factors Φ is given by

$$\Phi_{p\ \ p'a}^{(k+)\,(s\epsilon)} \sim \exp[-i(2k-1)\varphi](\rho)^{2ip-1} \otimes \exp(2is\alpha)(\rho')^{2ip-1}\begin{pmatrix}1\\ a\end{pmatrix}. \tag{4.2}$$

The first factor in the direct product is the above-mentioned function of x_1 and x_2, while the second factor is a column vector with the upper entry corresponding to the region $x_4>|x_3|$ and the lower one corresponding to $x_3>|x_4|$. We must now relate the parameters ρ, ρ', φ and α to the variables r, η, ψ, and ξ or θ that we have introduced in previous sections to parametrize the timelike region V_1 and the spacelike regions $V_2^{(1)}$ and $V_2^{(2)}$. We must then rewrite Φ in the form of a column vector with three entries giving its values in the above three regions. First of all from Eqs. (2.6), (3.1), and (4.1) we find the following relations between the two sets of parameters:

$$V_1: \rho=r\sinh(\xi/2), \quad \varphi=\psi, \quad \rho'=r\cosh(\xi/2), \quad \alpha=\eta, \tag{4.3a}$$

$$V_2^{(1)}: \rho=r\cosh(\xi/2), \quad \varphi=\psi, \quad \rho'=r|\sinh(\xi/2)|, \quad \alpha=\eta, \tag{4.3b}$$

$$V_2^{(2)}: \rho=r\cos(\theta/2), \quad \varphi=\psi, \quad \rho'=r|\sin(\theta/2)|, \quad \alpha=\eta. \tag{4.3c}$$

The basis vectors have to be normalized with respect to the measure

$$dx_1\,dx_2\,dx_3\,dx_4=\rho\,d\rho\,d\varphi\cdot 2\rho'\,d\rho'\,d\eta. \tag{4.4}$$

It is now straightforward to write down the properly normalized uncoupled basis vector Φ:

$$\Phi_{p\ \ p'a}^{(k+)\,(s\epsilon)}=(2\pi^2)^{-1}r^{2i(p+p')-2}\exp[2is\eta-i(2k-1)\psi] \times\begin{pmatrix}[\sinh(\xi/2)]^{2ip-1}[\cosh(\xi/2)]^{2ip'-1}\\ (1/\sqrt{2})[\cosh(\xi/2)]^{2ip-1}|\sinh(\xi/2)|^{2ip'-1}\\ (a/\sqrt{2})[\cos(\theta/2)]^{2ip-1}|\sin(\theta/2)|^{2ip'-1}\end{pmatrix}. \tag{4.5}$$

These vectors have the normalization

$$(\Phi_{p''\ \ p'''a'}^{(k'+)\,(s'\epsilon')}, \Phi_{p\ \ p'a}^{(k+)\,(s\epsilon)})=\delta_{k'k}\delta(s'-s)\delta_{\epsilon'\epsilon}\delta(p''-p)\delta(p'''-p')\delta_{a'a}. \tag{4.6}$$

Let us next take up the construction of the coupled basis vectors. Actually, most of the work involved in setting up these basis vectors has been done already in our construction of the $O(3,1)$ harmonics on the unit timelike and spacelike hyperboloids. Only the radial functions $f_1(r)$ and $f_2(r)$ for V_1 and V_2, respectively, have to be determined and these are fixed by the requirement that J_2 (and $\mathcal{A}$) be diagonal with eigenvalue p'' (and a'). We shall first obtain the coupled basis vectors for the UIR's of $SU(1,1)$ belonging to the discrete series which appear in the reduction of $D^+ \otimes C$. These are given by the $O(3,1)$ spherical harmonics on the spacelike hyperboloid that transform via the UIR's $\{2k'-1,0\}$ of $O(3,1)$. In these UIR's one finds that $Q=-(1/8)M^2=k'(1-k')$ and that the restriction of the generators J_α of Eq. (1.17) to the subspace carrying these UIR's (followed by a similarity transformation $rJ_\alpha r^{-1}$) results in the generators corresponding to the standard UIR $D_{k'}^+$ of $SU(1,1)$ given in Eq. (1.9) of Paper I. This shows that the $O(3,1)$ harmonic belonging to the UIR $\{2k'-1,0\}$ gives us the coupled basis vector corresponding to $D^+ \otimes C \to D^+$, which is

$$\Psi^{(k+)\,(s\epsilon)\,(k'+)}_{\qquad\qquad p''}=r^{2ip''-2}/\sqrt{\pi} \times\begin{pmatrix}0\\ \mathcal{Y}^{+(k')1}_{2k-1,\,2s}(\xi,\eta,\psi)\\ \mathcal{Y}^{+(k')2}_{2k-1,\,2s}(\theta,\eta,\psi)\end{pmatrix}. \tag{4.7}$$

Similarly the basis vectors that span the UIR's $\{0,2s'\}$ of $O(3,1)$ in $\mathcal{H}$ give the coupled basis vectors corresponding to the reduction $D^+ \otimes C \to C$, since in this case $Q=\frac{1}{4}+s'^2$ and the restriction of J_α to the subspace of these UIR's of $O(3,1)$ (followed by the similarity transformation $rJ_\alpha r^{-1}$) leads to the generators of the standard form for the UIR $C^\epsilon_{1/4+s'^2}$, namely the $J_\alpha(s',\epsilon')$ defined in Eq. (1.17) of Paper I. We can in addition verify that these vectors transform correctly according to the UIR (s',ϵ') in the standard form (see Appendix B).

$$\Psi^{(k+)\,(s\epsilon)\,(s'\epsilon')}_{\qquad\qquad p''a'}=r^{2ip''-2}/\sqrt{\pi} \times\begin{pmatrix}\mathcal{Y}^{-(s'\epsilon')}_{2k-1,\,2s}(\xi,\eta,\psi)\\ a'\exp[i\varphi(s'\epsilon')]\begin{pmatrix}\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,\,2s}(\xi,\eta,\psi)\\ \mathcal{Y}^{+(s'\epsilon')2}_{2k-1,\,2s}(\theta,\eta,\psi)\end{pmatrix}\end{pmatrix}. \tag{4.8}$$

$\varphi(s'\epsilon')$ is a phase angle required to ensure that $\Psi^{(k+)\,(s\epsilon)\,(s'\epsilon')}_{\qquad\qquad p''a'}$ transforms under $SU(1,1)$ according to the standard UIR $(s'\epsilon')$ set up in I. This phase has been evaluated in Appendix B and turns out to be $\varphi(s,0)=\pi/2$, $\varphi(s,\frac{1}{2})=-\pi/2$. The two types of vectors (4.7) and (4.8) are mutually orthogonal and their normalization properties are

$$\left(\Psi^{(k_1+)\,(s_1\epsilon_1)\,(k_1'+)}_{\qquad\qquad\quad p_1''}, \Psi^{(k+)\,(s\epsilon)\,(k'+)}_{\qquad\qquad p''}\right)$$

$$= \delta_{k_1 k}\delta(s_1 - s)\delta_{\epsilon_1\epsilon}\delta_{k_1' k'}\delta(p_1'' - p''), \tag{4.9a}$$

$$\left(\Psi^{(k_1+)(s_1\epsilon_1)(s_1'\epsilon_1')}_{p_1'' a_1'}, \Psi^{(k+)(s\epsilon)(s'\epsilon')}_{p'' a'}\right)$$
$$= \delta_{k_1 k}\delta(s_1 - s)\delta_{\epsilon_1\epsilon}\delta(s_1' - s)\delta_{\epsilon_1'\epsilon}\delta(p_1'' - p'')\delta_{a_1' a'}. \tag{4.9b}$$

From the completeness relations for the $O(3,1)$ spherical harmonics in the timelike and spacelike regions we can immediately write down the structure of the C–G series for the product $D^+ \otimes C$:

$$D_k^+ \otimes C^{\epsilon}_{1/4+s^2} = \sum_{k''=1 \text{ or } 3/2}^{\infty} D^+_{k''} \oplus \int_0^\infty ds' C^{\epsilon'}_{1/4+s'^2},$$

where

$$\epsilon' = 0(\tfrac12) \quad \text{and} \quad k''_{\min} = 1(\tfrac32) \quad \text{if } k+\epsilon = \text{integer (half-odd integer)}. \tag{4.10}$$

It is interesting to note that the UIR $D^+_{1/2}$ does not appear in the C–G series because the $O(3,1)$ spherical harmonic corresponding to the UIR $\{j_0 = 0, 0\}$ does not appear as a discrete summand in the reduction of functions defined on the spacelike hyperboloid. For the same reason, as we shall see later, the UIR $D^-_{1/2}$ will not appear in the reduction of $D^- \otimes C$.

5. C-G COEFFICIENTS FOR $D^+ \otimes C$ IN A CONTINUOUS BASIS

We have two types of C–G coefficients to compute; namely,

$$C(k+ s\epsilon\, \mathcal{R}\,|\, p\, p'a\, p''a') = \delta(p'' - p - p')\hat{C}(k+ s\epsilon\, \mathcal{R}\,|\, p\, p'a\, a'), \tag{5.1}$$

for $\mathcal{R} = (k', +)$ and $(s'\epsilon')$. The delta function on the right-hand side always arises out of the trivial r-integration in the scalar product and we shall drop it in future, recording only the values of $\hat{C}$.

From Eqs. (4.5) and (4.7) we have

$$\hat{C}(k+ s\epsilon\, k'+|\, p\, p'a) = \sqrt{\pi/2}\,\Big(\int_0^\infty d\xi[\sinh(\xi/2)]^{2ip}[\cosh(\xi/2)]^{2ip'} Y^{+(k')1}_{2k-1,2s}(\xi)$$
$$+ a\int_0^\pi d\theta[\sin(\theta/2)]^{2ip}[\cos(\theta/2)]^{2ip} Y^{+(k')2}_{2k-1,2s}(\theta)\Big). \tag{5.2}$$

From Eqs. (3.20) and (3.24) we see that we have essentially to evaluate the integrals

$$I_1 = \int_0^\infty d\xi[\sinh^2(\xi/2)]^{-(s+p')}[\cosh^2(\xi/2)]^{(1/2)-k-ip}$$
$$\times F(\tfrac12 - k + k' - is, \tfrac32 - k - k' - is; 1 - 2is; -\sinh^2(\xi/2)), \tag{5.3a}$$

$$I_2 = \int_0^\pi d\theta[\sin^2(\theta/2)]^{i(s+p')}[\cos^2(\theta/2)]^{k-(1/2)-ip}$$
$$\times F(\tfrac12 + k - k' - is, k + k' - \tfrac12 - is; 1 - 2is; \sin^2(\theta/2)). \tag{5.3b}$$

These integrals can be evaluated and expressed in terms of the ${}_3F_2$ function of unit argument[11]

$$I_1 = \frac{\Gamma(\frac12 - is - ip')\Gamma(k' + ip + ip')}{\Gamma(k' + \frac12 - is + ip)}$$
$$\times {}_3F_2\left(\begin{matrix} k' - k + \frac12 - is, k' + k - \frac12 - is, \frac12 - is - ip' \\ 1 - 2is, k' + \frac12 - is - ip \end{matrix}; 1\right), \tag{5.4a}$$

$$I_2 = \frac{\Gamma(\frac12 - is - ip')\Gamma(k - ip)}{\Gamma(k + \frac12 - is - ip - ip')}$$
$$\times {}_3F_2\left(\begin{matrix} k - k' + \frac12 - is, k + k' - \frac12 - is, \frac12 - is - ip' \\ 1 - 2is, k + \frac12 - is - ip - ip' \end{matrix}; 1\right). \tag{5.4b}$$

Thus we have the following result for $\hat{C}$:

$$\hat{C}(k+ s\epsilon\, k'+|\, p\, p'a)$$
$$= \sqrt{\pi/2}\, N_1\left(\frac{\Gamma(2is)\Gamma(2k')I_1}{\Gamma(k + k' - \frac12 + is)\Gamma(k' - k + \frac12 + is)} + (s \to -s)\right)$$
$$+ a\sqrt{\pi/2}\, N_2\left(\frac{\Gamma(2is)\Gamma(2k)I_2}{\Gamma(k + k' - \frac12 + is)\Gamma(k' - k + \frac12 + is)} + (s \to -s)\right). \tag{5.5}$$

Turning now to C–G coefficients of the second kind, we have from Eqs. (4.5) and (4.8)

$$\hat{C}(k+ s\epsilon\, s'\epsilon'|\, p\, p'a\, a')$$
$$= (\sqrt{\pi}/2)\Big[\int_0^\infty d\xi[\sinh(\xi/2)]^{-2ip}[\cosh(\xi/2)]^{-2ip'} Y^{-(s'\epsilon')}_{2k-1,2s}(\xi)$$
$$+ \sqrt{2}\, a' \exp[i\varphi(s'\epsilon')]\Big(\int_0^\infty d\xi[\sinh(\xi/2)]^{-2ip'}$$
$$\times[\cosh(\xi/2)]^{-2ip} Y^{+(s'\epsilon')1}_{2k-1,2s}(\xi) + a\int_0^\pi d\theta[\sin(\theta/2)]^{-2ip'}$$
$$\times[\cos(\theta/2)]^{-2ip} Y^{+(s'\epsilon')2}_{2k-1,2s}(\theta)\Big)\Big]. \tag{5.6}$$

From Eqs. (2.17), (3.14a), and (3.14c) we see that we have to evaluate the following three integrals:

$$\tilde{I}_1 = \int_0^\infty d\xi[\sinh^2(\xi/2)]^{k-1/2-ip}[\cosh^2(\xi/2)]^{i(s-p')}$$
$$\times F(k + is + is', k + is - is'; 2k; -\sinh^2(\xi/2)); \tag{5.7a}$$

$$\tilde{I}_2 = \int_0^\pi d\xi[\sinh^2(\xi/2)]^{i(s-p')}[\cosh^2(\xi/2)]^{k-1/2-ip}$$
$$\times F(k + is + is', k + is - is'; 1 + 2is; -\sinh^2(\xi/2)); \tag{5.7b}$$

$$\tilde{I}_3 = \int_0^\infty d\theta[\sin^2(\theta/2)]^{i(s-p')}[\cos^2(\theta/2)]^{k-1/2-ip}$$
$$\times F(k + is + is', k + is - is'; 1 + 2is; \sin^2(\theta/2)). \tag{5.7c}$$

These integrals can again be expressed in terms of the ${}_3F_2$ function of unit argument[11]:

$$\tilde{I}_1 = \frac{\Gamma(k - ip)\Gamma(\frac12 + is' + ip + ip')}{\Gamma(k + \frac12 + is' + ip')}$$
$$\times {}_3F_2\left(\begin{matrix} k + is + is', k - is + is', k - ip \\ 2k, k + \frac12 + is' + ip' \end{matrix}; 1\right), \tag{5.8a}$$

$$\tilde{I}_2 = \frac{\Gamma(\frac12 + is - ip')\Gamma(\frac12 + is' + ip + ip')}{\Gamma(1 + is + is' + ip)}$$
$$\times {}_3F_2\left(\begin{matrix} k + is + is', 1 - k + is + is', \frac12 + is - ip' \\ 1 + 2is, 1 + is + is' + ip \end{matrix}; 1\right), \tag{5.8b}$$

$$\tilde{I}_3 = \frac{\Gamma(\frac12 + is - ip')\Gamma(k - ip)}{\Gamma(k + \frac12 + is - ip - ip')}$$
$$\times {}_3F_2\left(\begin{matrix} k + is + is', k + is - is', \frac12 + is - ip' \\ 1 + 2is, k + \frac12 + is - ip - ip' \end{matrix}; 1\right). \tag{5.8c}$$

Hence we finally get for the C–G coefficient

$$\hat{C}(k+ s\epsilon\, s'\epsilon'|\, p\, p'a\, a') = \mathcal{J}_1 + a' \exp[i\varphi(s'\epsilon')](\mathcal{J}_2 + a\mathcal{J}_3), \tag{5.9}$$

where

$$\mathcal{G}_1 = -(i/2\sqrt{2}\pi)2^{2is'}\frac{\Gamma(k+is'+is)\Gamma(k+is'-is)}{\Gamma(2k)\Gamma(2is')}\tilde{I}_1$$

$$\mathcal{G}_2 = -\sqrt{2}(s'/8\pi^2)2^{2is'}\Gamma(-2is')[\exp(-s'\pi)+\eta_{\epsilon'}\exp(s'\pi)]$$
$$\times\Big(\{\exp[s\pi - i(2k-1)(\pi/2)] + \eta_{\epsilon'}\exp[-s\pi + i(2k-1)(\pi/2)]\}$$
$$\times\frac{\Gamma(k+is+is')}{\Gamma(k-is-is')}\Gamma(-2is)\tilde{I}_2 + (s\to -s)\Big),$$

$$\mathcal{G}_3 = \sqrt{2}(s'/8\pi^2)2^{2is'}\Gamma(-2is')[\exp(s'\pi)+\eta_{\epsilon'}\exp(-s'\pi)]$$
$$\times\{\exp[-s\pi + i(2k-1)(\pi/2)] + \eta_{\epsilon'}\exp[s\pi - i(2k-1)(\pi/2)]\}$$
$$\times\left(\frac{\Gamma(k+is+is')}{\Gamma(k-is-is')}\tilde{I}_3 + (s\to -s)\right),$$

$$\eta_{\epsilon'} \equiv \begin{cases} -1 & \text{for } \epsilon' = 0 \\ +1 & \text{for } \epsilon' = \tfrac{1}{2}. \end{cases}$$

This completes the evaluation of the C–G coefficients for $D^+ \otimes C$ in a continuous basis.

6. C-G SERIES AND COEFFICIENTS FOR $D^- \otimes C$

To obtain the C–G coefficients for $D^- \otimes C$ we can make use of the outer automorphism τ. We know that $\tau(k,\eta) = (k,-\eta)$ and $\tau(s,\epsilon) = (s,\epsilon)$. Hence if $\mathcal{R}\otimes\mathcal{R}' = \sum \mathcal{R}''$ then

$$\tau(\mathcal{R})\otimes\tau(\mathcal{R}') = \sum\tau(\mathcal{R}''). \tag{6.1}$$

Applying this to the C–G series for $D^+ \otimes C$ we get

$$D_k^- \otimes C_q^\epsilon = \sum_{k'=1 \text{ or } 3/2}^{\infty} D_{k'}^- \oplus \int_0^\infty ds' C^\epsilon_{1/4+s'^2}. \tag{6.2}$$

Turning to the C–G coefficients we observe that τ is diagonal in the UIR's C_q^ϵ which implies

$$C(\tau(\mathcal{R})\ \tau(\mathcal{R}')\ \tau(\mathcal{R}'')\gamma \mid pa\ p'b\ p''c)$$
$$= abc\sum_{\gamma'}\alpha_{\gamma\gamma'}C(\mathcal{R}\ \mathcal{R}'\ \mathcal{R}''\gamma' \mid pa\ p'b\ p''c), \tag{6.3}$$

where γ is a multiplicity index for the representation $\mathcal{R}''$ occurring in the reduction of $\mathcal{R}\otimes\mathcal{R}'$ and $\alpha_{\gamma\gamma'}$ a set of mixing coefficients. Applying this to the case of $D^+\otimes C$, we obtain

$$C(k-\ s\epsilon\ k'- \mid p\ p'a\ p'') = aC(k+\ s\epsilon\ k'+ \mid p\ p'a\ p'') \tag{6.4a}$$

$$C(k-s\epsilon\ s'\epsilon' \mid p\ p'a\ p''a') = aa'C(k+\ s\epsilon\ s'\epsilon' \mid p\ p'a\ p''a') \tag{6.4b}$$

and the same relations hold for the $\hat{C}$'s.

7. SUMMARY

Following the approach of previous papers we have related the Clebsch–Gordan problem of $SU(1,1)$ for products of the type $D_k^+ \otimes C^\epsilon_{(1/4)+s^2}$ to the properties of $O(3,1)$ spherical harmonics on the timelike and spacelike hyperboloids in an $O(2)\otimes O(1,1)$ basis.[12] We thus understand in a new way the structure of the C–G series for this case. Using these spherical harmonics we have computed the C–G coefficients in a continuous basis and these are again expressed in terms of the ${}_3F_2$ function of unit argument. There are however several terms: four in the case of $D^+\otimes C \to D^+$ and five in the case of $D^+ \otimes C \to C$. Using the properties of the representations D^+ and C under the automorphism τ we have related the C–G series and coefficients for $D^-\otimes C \to D^-$ and $D^-\otimes C \to C$ with the corresponding ones for the product $D^+ \otimes C$.

ACKNOWLEDGMENTS

The authors would like to thank Dr. V. Balakrishnan for extremely valuable discussions. One of us (N.M.) expresses his thanks to the Center for Particle Theory, University of Texas, Austin, Texas, and its Director Professor E.C.G. Sudarshan, for making possible a visit during which this paper was partly written.

APPENDIX A

In this appendix we shall be concerned with the determination of the two normalization constants N_1 and N_2 which were left undetermined in Eqs. (3.20) and (3.24). We will first find the integral representations for the functions $f_k^{(1)}(\xi)$ and $f_k^{(2)}(\theta)$ given by the method of integral geometry and then compare them with Eqs. (3.20) and (3.24). This offers a means of fixing the constants N_1 and N_2.

As stated in Sec. 3, the method of integral geometry gives the following prescription for the construction of (square-integrable) functions on the unit spacelike hyperboloid transforming via the UIR $\{n,0\}$ of $O(3,1)$ for $n=1,2,3,\ldots$. Consider a function $F(\mathbf{l},\mathbf{b};n)$ of two unit 3-vectors $\mathbf{l}$ and $\mathbf{b}$ subject to the constraint

$$F(\mathbf{l}, R(\mathbf{l},\omega)\mathbf{b};n) = \exp(in\omega)F(\mathbf{l},\mathbf{b};n), \tag{A1}$$

where $R(\mathbf{l},\omega)$ stands for a space-rotation through an angle ω about the direction of $\mathbf{l}$. Then a function $f(x)\in\mathcal{H}$ lying in the subspace of the UIR's $\{n,0\}$ can be written in fully reduced form in terms of $F(\mathbf{l},\mathbf{b};n)$. Thus

$$f(x) = (2/\pi^2)\sum_{n=1}^{\infty} n\int d\Omega(\mathbf{l})\delta(x_4 - \mathbf{x}\cdot\mathbf{l})F(\mathbf{l},\mathbf{x}-x_4\mathbf{l};n), \tag{A2}$$

with the norm of $f(x)$ given by

$$\int (dx)|f(x)|^2 = (4/\pi^2)\sum_{n=1}^{\infty} n\int d\Omega(\mathbf{l})|F(\mathbf{l},\mathbf{b};n)|^2. \tag{A3}$$

(dx) is the $O(3,1)$ invariant measure on the spacelike hyperboloid and $d\Omega(\mathbf{l})$ the solid angle element associated with the vector $\mathbf{l}$. We shall find it convenient to introduce Euler angles α,β,γ to describe $\mathbf{l}$ and $\mathbf{b}$:

$$\mathbf{l} = (\sin\beta\cos\gamma, \sin\beta\sin\gamma, \cos\beta),$$
$$\mathbf{b} = (\cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma, \cos\alpha\cos\beta\sin\gamma$$
$$+ \sin\alpha\cos\beta, -\cos\alpha\sin\beta). \tag{A4}$$

Then the α-dependence of $F(\mathbf{l},\mathbf{b};n)$ factorizes due to the constraint in Eq. (A1) if we realize

$$\exp(i\alpha) = \text{phase between 0 and } 2\pi \text{ of } -b_3 + i(\mathbf{l}\times\mathbf{b})_3. \tag{A5}$$

Then,

$$F(\mathbf{l},\mathbf{b};n) = F(\alpha,\beta,\gamma;n) = \exp(in\alpha)F(0,\beta,\gamma;n). \tag{A6}$$

In order to construct an eigenfunction of M^2, M_{12}, and M_{34} with eigenvalues $2(n_0^2-1)$, m, and a, respectively, we choose

$$F_{n_0ma}(\alpha,\beta,\gamma;n) = \delta_{nn_0}\exp(in\alpha - im\gamma)[\cot\beta/2)^{ia}/\sin\beta]. \tag{A7}$$

Then from Eq. (A3) the norm of the corresponding $f_{n_0ma}(x)$ becomes

$$\int (dx) f_{n_0'm'a'}(x)^* f_{n_0ma}(x) = 16n_0\delta_{n_0'n_0}\delta_{m'm}\,\delta(a'-a). \tag{A8}$$

Parametrizing the unit spacelike hyperboloid ($x^2=1$) according to Eq. (3.1), we have

$$f_{n_0ma}(x) = \begin{pmatrix} f^{(1)}_{n_0ma}(\xi,\eta,\psi) \\ f^{(2)}_{n_0ma}(\theta,\eta,\psi)\end{pmatrix} = \exp(ia\eta - im\psi)\begin{pmatrix} f^{(1)}_{n_0ma}(\xi) \\ f^{(2)}_{n_0ma}(\theta)\end{pmatrix}. \tag{A9}$$

Substituting Eq. (A7) in Eq. (A2) and changing the variable from β to λ by $\cot\beta/2 = \exp(\lambda)$, we find after a few simple manipulations

$$\begin{aligned} f^{(1)}_{n_0ma}(\xi) &= (2n_0/\pi^2)\int_0^{2\pi}\exp(im\gamma)\,d\gamma\int_{-\infty}^{\infty}\exp(-ia\lambda)\,d\lambda \\ &\quad\times\delta\,(\sinh(\xi/2)\cosh\lambda - \cosh(\xi/2)\cos\gamma) \\ &\quad\times(\sinh(\xi/2)\sinh\lambda + i\cosh(\xi/2)\sin\gamma)^{n_0}, \end{aligned} \tag{A10a}$$

$$\begin{aligned} f^{(2)}_{n_0ma}(\theta) &= (2n_0/\pi^2)\int_0^{2\pi}\exp(im\gamma)\,d\gamma\int_{-\infty}^{\infty}\exp(-ia\lambda)\,d\lambda \\ &\quad\times\delta\,(\sin(\theta/2)\sinh\lambda - \cos(\theta/2)\cos\gamma) \\ &\quad\times(\sin(\theta/2)\cosh\lambda + i\cos(\theta/2)\sin\gamma)^{n_0}. \end{aligned} \tag{A10b}$$

These functions have the following normalization for fixed m and a but different n:

$$\pi^2\left(\int_{-\infty}^{\infty} d\xi\,|\sinh\xi|\,f^{(1)*}_{n'ma}(\xi) f^{(1)}_{nma}(\xi) + \int_{-\pi}^{\pi} d\theta\,|\sin\theta|\,f^{(2)}_{n'ma}(\theta) f^{(2)}_{nma}(\theta)\right) = 16n\delta_{n'n}. \tag{A11}$$

Let us first concentrate on $f^{(1)}_{nma}(\xi)$ and put the integral representation for this function, Eq. (A10a), in a more convenient form. The δ function can be used to carry out the λ integration. Since $\cosh\lambda$ is an even function of λ, we have

$$\begin{aligned} &\delta(\cosh\lambda\sinh(\xi/2) - \cos\gamma\cosh(\xi/2)) \\ &\quad= [\delta(\lambda-\lambda_0) + \delta(\lambda+\lambda_0)]/\sinh\lambda_0\sinh(\xi/2), \end{aligned} \tag{A12}$$

where $\cosh\lambda_0 = \coth(\xi/2)\cos\gamma$. Moreover, since $\cosh\lambda_0 \geq 1$, we must have $\cos\gamma \geq \tanh(\xi/2)$. Therefore, we have

$$\begin{aligned} f^{(1)}_{nma}(\xi) &= (2n/\pi^2)\int_{-\gamma_0}^{\gamma_0}\exp(im\gamma)\,d\gamma[\sinh\lambda_0\sinh(\xi/2)]^{-1} \\ &\quad\times\{\exp(-ia\lambda_0)[\sinh(\xi/2)\sinh\lambda_0 + i\cosh(\xi/2)\sin\gamma]^n \\ &\quad+\exp(ia\lambda_0)[-\sinh(\xi/2)\sinh\lambda_0 + i\cosh(\xi/2)\sin\gamma]^n\}. \end{aligned} \tag{A13}$$

Changing variables to μ defined by $\sin\mu = \cosh(\xi/2)\times\sin\gamma$ (which makes the integration go from $-\pi/2$ to $\pi/2$), we get after some calculation

$$\begin{aligned} &f^{(1)}_{nma}(\xi) \\ &= (2n/\pi^2)[\cosh(\xi/2)]^{-n}\{[\sinh(\xi/2)]^{ia}\varphi(n,m,a;-\sinh^2(\xi/2)) \\ &\quad+(-1)^n[\sinh(\xi/2)]^{-ia}\varphi(-n,m,-a;-\sinh^2(\xi/2))\}, \end{aligned} \tag{A14a}$$

$$\begin{aligned} \varphi(n,m,a;z) &\equiv \int_{-\pi/2}^{\pi/2}\frac{\exp(in\mu)\,d\mu}{(\cos^2\mu - z)^{1/2}}\,(\sqrt{\cos^2\mu - z} + i\sin\mu)^m \\ &\quad\times(\sqrt{\cos^2\mu - z} + \cos\mu)^{-ia}. \end{aligned} \tag{A14b}$$

If we now equate this expression for $f^{(1)}_{nma}(\xi)$ with that obtained in Eq. (3.20) which is

$$\begin{aligned} f^{(1)}_{nma}(\xi) &= N_1[\cosh(\xi/2)]^m\Gamma(n+1) \\ &\quad\times\left[\frac{\Gamma(ia)[\sinh(\xi/2)]^{-ia}}{\Gamma[(1+m+n+ia)/2]\Gamma[(1-m+n+ia)/2]}\right. \\ &\quad\times F\left(\frac{1+m+n-ia}{2},\frac{1+m-n-ia}{2},1-ia;-\sinh^2\frac{\xi}{2}\right) \\ &\quad\left.+(a\to -a)\right], \end{aligned} \tag{A15}$$

we have an equation which can fix the value of N_1. To achieve this we multiply both sides by $(\sinh(\xi/2))^{ia}$, replace a by $a-i\epsilon$, $\epsilon>0$, and take the limit $\xi\to 0+$. Then the second term on the left-hand side [i.e., of Eq. (A15)] vanishes since

$$[\sinh(\xi/2)]^{2i(a-i\epsilon)} = [\sinh(\xi/2)]^{2ia}[\sinh(\xi/2)]^{2\epsilon}\to 0.$$

On the right-hand side [i.e., Eq. (A14)] the second φ becomes finite so that we get

$$\begin{aligned} &N_1\frac{\Gamma(n+1)\Gamma(ia)}{\Gamma[(1+m+n+ia)/2]\Gamma[(1-m+n+ia)/2]} \\ &\quad= (-1)^n\frac{2n}{\pi^2}\varphi(-n,m,-a+i\epsilon;0) + \frac{2n}{\pi^2}\lim_{\xi\to 0+} \\ &\quad\times\{[\sinh(\xi/2)]^{2ia+2\epsilon}\varphi(n,m,a-i\epsilon;-\sinh^2(\xi/2))\}. \end{aligned} \tag{A16}$$

Now,

$$\begin{aligned} \varphi(-n,m,-a+i\epsilon;0) &= 2^{ia}\int_{-\pi/2}^{\pi/2} d\mu\,\exp[i(m-n)\mu](\cos\mu)^{ia+\epsilon-1} \\ &= 2\pi\Gamma(ia)\Big/\Gamma\left(\frac{1+m-n+ia}{2}\right) \\ &\quad\times\Gamma\left(\frac{1-m+n+ia}{2}\right), \end{aligned} \tag{A17}$$

using standard results.[13]

The second term in Eq. (A16) can be shown to vanish in the limit. Thus we end up with the following value for N_1:

$$N_1 = \frac{4n}{\pi}(-1)^n\frac{\Gamma[(1+m+n+ia)/2]}{\Gamma(n+1)\Gamma[(1+m-n+ia)/2]}. \tag{A18}$$

Turning now to the case of $f^{(2)}_{nma}(\theta)$, we can again make use of the δ function in Eq. (A10b) to do the λ integration:

$$\begin{aligned} &\delta(\sin(\theta/2)\sinh\lambda - \cos(\theta/2)\cos\gamma) \\ &= \delta(\lambda-\lambda_0)/\cosh\lambda_0\sin(\theta/2), \quad \text{where } \sinh\lambda_0 = \cot(\theta/2)\cos\gamma. \end{aligned} \tag{A19}$$

From here on the calculation proceeds along similar lines as before and we finally end up with the result

$$N_2 = \frac{4n}{\pi}(-1)^n\frac{\Gamma[(1+m+n+ia)/2]}{\Gamma(1+m)\Gamma[(1+m-n+ia)/2]} \tag{A20}$$

APPENDIX B

This appendix deals with the phase angle $\varphi(s,\epsilon)$ introduced in Eq. (4.8). The need for such a phase was encountered once before in II when we considered the reduction of the product $D^+\otimes D^-$ and the situation here is analogous. We have the representation $\mathcal{D}^+\otimes\mathcal{C}$ of $SU(1,1)$ and a representation of $O(3,1)$ acting on the Hilbert space $\mathcal{H}$, the two representations commuting with each other, but sharing the same Casimir operator. Hence by reducing the $O(3,1)$ representation into UIR's we were able to simultaneously achieve the reduction of $\mathcal{D}^+\otimes\mathcal{C}$ into UIR's of $SU(1,1)$ and by suitably choosing a basis in $\mathcal{H}$ we were able to isolate the product $D^+\otimes C$

from $\mathcal{D}^+\otimes\mathcal{C}$ and thus obtained its reduction into UIR's. Now in reducing the $O(3,1)$ representation on $\mathcal{H}$ we were led to set up complete sets of spherical harmonics for the timelike and spacelike regions V_1 and V_2 of Minkowski space. The $O(3,1)$ spherical harmonics belonging to the UIR $\{0,\rho\}$ served as basis functions for the continuous series of UIR's of $SU(1,1)$ in the reduction of $\mathcal{D}^+\otimes\mathcal{C}$. Using the method of integral geometry we ensured the complete identity of $O(3,1)$ transformation properties of these spherical harmonics in V_1 and V_2.

Consider now a general vector $f\in\mathcal{H}$ belonging to the UIR (s',ϵ') of $SU(1,1)$:

$$f=\sum_{k=1/2\text{ or }1}\int_{-\infty}^{\infty}dsC_k(s)\begin{pmatrix} f_1(r)\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi,\eta,\psi)\\ \exp[i\varphi(s'\epsilon')]f_2(r)\begin{pmatrix}\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi)\\ \mathcal{Y}^{+(s'\epsilon')2}_{2k-1,2s}(\theta,\eta,\psi)\end{pmatrix}\end{pmatrix}, \tag{B1}$$

where the $C_k(s)$ are arbitrary constants.

The $O(3,1)$ transformations can in no way distinguish between the components of f belonging to V_1 and V_2 since r is invariant under $O(3,1)$ and $\mathcal{Y}^-$ and $\mathcal{Y}^+$ have been chosen to transform identically. On the other hand, $SU(1,1)$ transformations, which commute with the $O(3,1)$ ones, cannot alter the $O(3,1)$ structure of f. Hence if $h=\exp(itJ_0)f$, where J_0 is one of the generators of $\mathcal{D}^+\otimes\mathcal{C}$, both f and h will have the same $O(3,1)$ spherical harmonics [or the same set of constants $C_k(s)$] the only difference being in the radial functions. The ensure that the radial functions transform under $SU(1,1)$ according to the standard UIR (s',ϵ') set up in I, we must suitably choose the phase angle $\varphi(s'\epsilon')$ which is the only freedom allowed by the $O(3,1)$ representation.

To determine $\varphi(s'\epsilon')$ we must first discover the relation between the radial functions in h and in f. We can then adjust $\varphi(s'\epsilon')$ so that this relation is given by the kernel for the finite transformation $\exp(itJ_0)$ in the standard representation (s',ϵ') of I.

From Eq. (1.17) we have

$$J_0=\tfrac{1}{4}(x^2-\Box^2). \tag{B2}$$

Hence,[14]

$$[\exp(itJ_0)f](x)\equiv h(x;t)=\int d^4x\,\mathcal{L}(x,x';t)f(x'), \tag{B3a}$$

$$\mathcal{L}(x,x';t)=[2\pi\sin(t/2)]^{-2}\exp\{-i[(x^2+x'^2)\cos(t/2)-2x\cdot x']/2\sin(t/2)\}. \tag{B3b}$$

Let us choose $f(x)$ to be nonvanishing on V_1 (and by reflection $\mathbb{R}$ on V_3) and zero on V_2. We can get an equation for $\varphi(s'\epsilon')$ by evaluating $h(x;t)$ on V_2, say $V_2^{(1)}$. Then we must parametrize x' according to Eq. (2.6) and x according to Eq. (3.1a):

$$\begin{aligned} &x^2=r^2,\quad x'^2=r'^2,\\ &x\cdot x'=rr'[\cosh(\xi/2)\sinh(\xi'/2)\cos(\psi-\psi')-\sinh(\xi/2)\cosh(\xi'/2)\cosh(\eta-\eta')],\\ &d^4x'=\tfrac{1}{4}r'^3dr'\sinh\xi'\,d\xi'\,d\eta'\,d\psi'. \end{aligned}\tag{B4}$$

Hence we have

$$\begin{aligned} h(x;t)&=h(r,\xi,\eta,\psi;t)\\ &=\tfrac{1}{4}\int_0^\infty r'^3dr'\int_0^\infty\sinh\xi'\,d\xi'\int_{-\infty}^{\infty}d\eta'\int_0^{2\pi}d\psi'\\ &\quad\times\frac{\exp[-(i/2)(r^2-r'^2)\cot(t/2)]}{[2\pi\sin(t/2)]^2}\Bigg[\exp\left(\frac{irr'}{\sin(t/2)}\right.\\ &\quad\times[\cosh(\xi/2)\sinh(\xi'/2)\cos(\psi-\psi')-\sinh(\xi/2)\\ &\quad\times\cosh(\xi'/2)\cosh(\eta-\eta')]\Bigg)f(V_1)+\exp\Bigg(\frac{-irr'}{\sin(t/2)}\\ &\quad\times[\cosh(\xi/2)\sinh(\xi'/2)\cos(\psi-\psi')-\sinh(\xi/2)\\ &\quad\times\cosh(\xi'/2)\cosh(\eta-\eta')]\Bigg)f(V_3)\Bigg]\\ &=\tfrac{1}{4}\int_0^\infty r'^3dr'\int_0^\infty\sinh\xi'\,d\xi'\int_{-\infty}^{\infty}d\eta'\int_0^{2\pi}d\psi'\\ &\quad\times[2\pi\sin(t/2)]^{-2}\exp[-(i/2)(r^2-r'^2)\cot(t/2)]\\ &\quad\times\{\exp(i\alpha[\cosh(\xi/2)\sinh(\xi'/2)\cos(\psi-\psi')\\ &\quad-\sinh(\xi/2)\cosh(\xi'/2)\cosh(\eta-\eta')])+\eta_{\epsilon'}\\ &\quad\times\exp(-i\alpha[\cosh(\xi/2)\sinh(\xi'/2)\cos(\psi-\psi')\\ &\quad-\sinh(\xi/2)\cosh(\xi'/2)\cosh(\eta-\eta')])\}f(V_1), \end{aligned}\tag{B5}$$

where we have set $\mathbb{R}=\eta_{\epsilon'}$ and $\alpha=rr'/\sin(t/2)$. For $f(V_1)$ we must put in an expression corresponding to Eq. (B1):

$$f(V_1)=f_1(r',\xi',\eta',\psi')=\sum_k\int dsC_k(s)f_1(r')\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi',\eta',\psi').$$

The requirement that $h(x;t)$ have the same $O(3,1)$ structure as f and that the radial functions of $h(x;t)$ must transform correctly under the UIR (s',ϵ') of $SU(1,1)$ imply that

$$\exp[i\varphi(s'\epsilon')]h_2^{(1)}(r,\xi,\eta,\psi;t)=\sum_k\int dsC_k(s)\exp[i\varphi(s'\epsilon')]h_2(r;t)\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi) \tag{B6a}$$

and

$$h_2(r;t)=\int_0^\infty r'^3dr'\,L^{(s'\epsilon')}_{21}(r,r';t)f_1(r'), \tag{B6b}$$

where

$$\begin{aligned} L^{(s'\epsilon')}_{21}(r,r';t)&=[\pi\sin(t/2)]^{-1}(rr')^{-1}\exp[-(i/2)(r^2-r'^2)\cot(t/2)]\\ &\quad\times[\exp(-\pi s')-\eta_{\epsilon'}\exp(\pi s')]K_{2is}(\alpha). \end{aligned}$$

(We have taken account here of the change in measure from $r\,dr$ to r^3dr as compared to the standard UIR's defined in I.) Equating the expressions for $h(x;t)$ in Eqs. (B5) and (B6a), we clearly have an equation for $\varphi(s'\epsilon')$. Since $f_1(r')$ is arbitrary, we see that the equation must be independent of $f_1(r')$. Further, ξ is a free parameter in the equation and we can therefore set $\xi=0$ to simplify the equation. Finally, recalling that

$$\mathcal{Y}^{-(s'\epsilon')}_{2k-1,2s}(\xi',\eta',\psi')=\exp[2is\eta'-i(2k-1)\psi']Y^{-(s'\epsilon')}_{2k-1,2s}(\xi') \tag{B7a}$$

and

$$\mathcal{Y}^{+(s'\epsilon')1}_{2k-1,2s}(\xi,\eta,\psi)=\exp[2is\eta-i(2k-1)\psi]Y^{+(s'\epsilon')1}_{2k-1,2s}(\xi), \tag{B7b}$$

we finally end up with the following equation for $\varphi(s'\epsilon')$:

$$\begin{aligned} &\exp[i\varphi(s'\epsilon')]\sum_k\int dsC_k(s)[\exp(-\pi s')-\eta_{\epsilon'}\exp(\pi s')]K_{2is'}(\alpha)\\ &\quad\times Y^{+(s'\epsilon')}_{2k-1,2s}(\xi=0)=-(\alpha/16\pi)\sum_k\int dsC_k(s)\int_1^\infty d\cosh\xi'\\ &\quad\times\int_{-\infty}^{\infty}d\eta'\int_0^{2\pi}d\psi'\exp[-2is\eta'+i(2k-1)\psi']\\ &\quad\times Y^{-(s'\epsilon')}_{2k-1,2s}(\xi)\{\exp[i\alpha\sinh(\xi'/2)\cos\psi']\} \end{aligned}$$

$$+\eta_{\epsilon'}\exp[-i\alpha\sinh(\xi'/2)\cos\psi']\}$$
$$=-(\alpha/16\pi)\sum_k C_k(0)\int_1^\infty d\cosh\xi'\int_0^{2\pi}d\psi'$$
$$\times\exp[i(2k-1)\psi']Y_{2k-1,0}^{-(s'\epsilon')}(\xi')$$
$$\times\{\exp[i\alpha\sinh(\xi'/2)\cos\psi']+\eta_{\epsilon'}$$
$$\times\exp[-i\alpha\sinh(\xi'/2)\cos\psi']\}. \quad (B8)$$

Now

$$Y_{2k-1,2s}^{+(s'\epsilon')}(\xi=0)=\sqrt{4\pi}(s'/4\pi^2)\delta(s)2^{2is}\Gamma(-2is')$$
$$\times[\exp(s'\pi)+\eta_{\epsilon'}\exp(-s'\pi)]\{\exp[i(2k-1)(\pi/2)]$$
$$+\eta_{\epsilon'}\exp[-i(2k-1)(\pi/2)]\}\frac{\Gamma(k+is')}{\Gamma(k-is')}. \quad (B9)$$

Thus we see that only the $C_k(0)$ part of the equation survives on both sides. However, since $C_k(0)$ are also arbitrary numbers the equation must be independent of them. We are finally left with the following equation:

$$\exp[i\varphi(s'\epsilon')]\sqrt{4\pi}(s'/4\pi^2)2^{2is'}\Gamma(-2is')\eta_{\epsilon'}[\exp(s'\pi)-\eta_{\epsilon'}$$
$$\times\exp(-s'\pi)][\exp(s'\pi)+\eta_{\epsilon'}\exp(-s'\pi)]K_{2is'}(\alpha)\frac{\Gamma(k+is')}{\Gamma(k-is')}$$
$$=(\alpha/16)2\pi\int_1^\infty d\cosh\xi' Y_{2k-1,0}^{-(s'\epsilon')}(\xi')J_{2k-1}[\alpha\sinh(\xi'/2)]. \quad (B10)$$

The integral on the right-hand side can be evaluated using standard formulas[15] and we end up with the result

$$\varphi(s'0)=\pi/2,\quad \varphi(s'\tfrac{1}{2})=-\pi/2. \quad (B11)$$

*Present address: Center for Theoretical Studies, Indian Institute of Science, Bangalore-12, India.

[1]N. Mukunda and B. Radhakrishnan, J. Math. Phys. 15, 1320, 1332 (1974). These papers will be referred to as I and II, respectively.

[2]I. S. Shapiro, Sov. Phys.-Dokl. 1, 91 (1956); Chou Kuang-Chao and L. G. Zastavenko, Sov. Phys.-JETP 8, 990 (1959); E. Angelopoulos, Ann. Inst. Henri Poincaré A 15, 303 (1971).

[3]B. Radhakrishnan and N. Mukunda, "Space-like Representations of the Inhomogeneous Lorentz Group in a Lorentz Basis," J. Math. Phys. (to be published). See also, A. Bassetto and M. Toller, Ann. Inst. Henri Poincaré A 18, 1 (1973).

[4]I. M. Gel'fand, R. A. Minlos, and Y. A. Shaprio, *Representations of the Rotation and Lorentz Groups and their Applications* (Macmillan, New York, 1963), p. 199.

[5]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, especially Chaps. V and VI.

[6]N. Ya. Vilenkin and Ya. A. Smorodinskii, Sov. Phys.-JETP 19, 1209 (1964); E. G. Kalnins, J. Math. Phys. 13, 1304 (1972).

[7]The spherical harmonic $\mathcal{Y}$ is actually independent of its superscript ϵ' which merely indicates the eigenvalue of the reflection operator $\mathbb{R}$ connecting the positive and negative timelike hyperboloids $\epsilon'\equiv 0$ for $R=-1$ and $\epsilon'\equiv\frac{1}{2}$ for $R=+1$.

[8]*Bateman Manuscript Project* (McGraw-Hill, New York, 1954), Tables of Integral Transforms Vol. 1, formula 1.9 (5) on p. 30. We will also need the result:

$$\int_0^{2\pi}e^{im\varphi}(\cos\varphi)^n d\varphi=(2\pi/2^n)\{n!/[(n+m)/2]![(n-m)/2]!\},$$

if $n\geq m$ and $(n-m)$ is an even integer, $=0$ otherwise.

[9]*Bateman Manuscript Project* (Ref. 8, above) formula 1.3 (1) on p. 10 and formula 2.3 (1) on p. 68, (The first formula has to be regularized to hold for $Re\,\nu=1$ before it can be used.) We will also have to use certain standard representations for Bessel and Hankel functions which may be found, for example, in N. N. Lebedev, *Special Functions and their Applications* (Prentice Hall, Englewood Cliffs, N.J., 1965).

[10]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 692, formula 6.574.1 and p. 693, formula 6.576.3.

[11]I. S. Gradshteyn and I. M. Ryzhik, (Ref. 10, above), p. 849, formula 7.572.5.

[12]$O(3,1)$ spherical harmonics on the spacelike hyperboloid in an $O(3)$ basis have been constructed by Bassetto and Toller, Ref. 3.

[13]I. S. Gradshteyn and I. M. Ryzhik, (Ref. 10, above), p. 372, formula 3.631.9.

[14]Here we have made use of a result from R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), p. 63.

[15]*Bateman Manuscript Project* (McGraw-Hill, New York, 1954), Tables of Integral Transforms Vol. 2, formula 8.17 (3) on p. 81. Note that the conditions on the parameters given in this reference are those that follow from naive power counting, ignoring the oscillatory behavior of the Bessel function at infinity. A similar comment applies also to our use of the formula quoted in Ref. 11.

The Clebsch-Gordan problem and coefficients for the three-dimensional Lorentz group in a continuous basis. IV

N. Mukunda* and B. Radhakrishnan

Tata Institute of Fundamental Research, Bombay, 5, India
(Received 22 February 1973)

This is the last of four papers describing a new approach to the Clebsch–Gordan problem for the group $SU(1,1)$. Here we have related the Clebsch–Gordan series for products of the type $C \otimes C$ to properties of the group $O(2,2)$ and the structure of the series is thus seen to arise out of the properties of $O(2,2)$ spherical harmonics in an $O(1,1) \otimes O(1,1)$ basis. The Clebsch–Gordan coefficients in a continuous basis are also evaluated.

INTRODUCTION

This is the fourth and final paper in a sequence devoted to a new treatment of the Clebsch–Gordan problem for the unitary irreducible representations (UIR's) of the group $SU(1,1)$.[1] The entire work has been presented in four parts because of the characteristic differences between the four essentially distinct kinds of direct products of UIR's one can form; but the unifying element is the fact that in each case a particular four-dimensional real orthogonal or pseudo-orthogonal group acts as a symmetry group of the problem and essentially determines the structure of the C–G series. In addition, of course, we determine a normalized set of Clebsch–Gordan coefficients in an $O(1,1)$ basis in each case.

In the three previous papers, we related the direct products $D^+\otimes D^+$, $D^+\otimes D^-$ and $D^+\otimes C$ to certain representations of $O(4)$, $O(2,2)$, and $O(3,1)$, respectively. These representations were required in an $O(2)\otimes O(2)$, $O(2)\otimes O(2)$, and an $O(2)\otimes O(1,1)$ basis, respectively. In the present paper, we solve the C–G problem for products of the form $C\otimes C$ using some properties of certain representations of $O(2,2)$ in an $O(1,1)\otimes O(1,1)$ basis; this is the symmetry group for such products. The reader will notice that the treatments of the four kinds of products of UIR's of $SU(1,1)$ are very similar to one another upto the point where one recognizes the appropriate four-dimensional rotational symmetry, but naturally diverge thereafter.

In Sec. 1, we construct the unitary representation $\mathcal{C}\otimes\mathcal{C}$ of $SU(1,1)$, and display the group of transformations, G, under which it is invariant. The representation $\mathcal{C}\otimes\mathcal{C}$ acts as the source of all products of the form $C_q^\epsilon\otimes C_{q'}^{\epsilon'}$. The group G consists of its identity component which is just the group $O(2,2)$ and one more component generated by a discrete symmetry transformation of $\mathcal{C}\otimes\mathcal{C}$. The structure of G, its action on four-dimensional real space, and the choice of angular variables in this space appropriate to the present problem, are all explained in Sec. 2. Combining these results with the expression of the Plancherel theorem for $SU(1,1)$ in an $O(1,1)$ basis, we carry out in Sec. 3 the construction of a complete set of "spherical harmonics" for the group G, in four-dimensional space. All these steps are quite similar to those taken in the three previous papers of this series, only the details differ. With the help of these spherical harmonics, we set up in Sec. 4 the two types of basis vectors in the space of the representation $\mathcal{C}\otimes\mathcal{C}$ of $SU(1,1)$ from whose structure the C–G series for the product $C_q^\epsilon\otimes C_{q'}^{\epsilon'}$ is easily read off. Finally, Sec. 5 gives the expressions for the C–G coefficients in the $O(1,1)$ basis and Sec. 6 contains a summary and remarks on this work.

1. SYMMETRIES OF THE REPRESENTATION $\mathcal{C}\otimes\mathcal{C}$ OF SU(1,1)

Let us take the direct product of the unitary representation $\mathcal{C}$ of $SU(1,1)$, constructed in Sec. 2 of I, with itself. We shall write $\mathcal{H}(\mathcal{C},13)$ and $\mathcal{H}(\mathcal{C},24)$ for the two Hilbert spaces in which the individual representations $\mathcal{C}$ are defined, so that $\mathcal{C}\otimes\mathcal{C}$ acts in the space $\mathcal{H} = \mathcal{H}(\mathcal{C},13)\otimes\mathcal{H}(\mathcal{C},24)$. The numerals 1 and 3 label the variables used in constructing the generators of the first factor, 2 and 4 those of the second factor, in the product $\mathcal{C}\otimes\mathcal{C}$, in the manner of Sec. 2 of I. [The numbering is done in this particular way because then the description of the group $O(2,2)$ in terms of $O(2,1)$ can be taken over with no changes at all from II.] Elements of $\mathcal{H}$ will be functions $f(x_1,x_2,x_3,x_4)$ with the squared norm given by

$$\|f\|^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1dx_2dx_3dx_4|f(x)|^2<\infty. \tag{1.1}$$

The four independent sets of oscillator operators defined on $\mathcal{H}$ obey

$$\begin{aligned}&[a_j,a_k^+]=\delta_{jk},\quad [a_j,a_k]=[a_j^+,a_k^+]=0,\\ &a_j=-\frac{i}{\sqrt2}\left(x_j+\frac{\partial}{\partial x_j}\right),\quad a_j^+=\frac{i}{\sqrt2}\left(x_j-\frac{\partial}{\partial x_j}\right),\\ &j,k=1,2,3,4.\end{aligned} \tag{1.2}$$

The generators for $\mathcal{C}\otimes\mathcal{C}$ are the sums of the individual sets of generators, and in terms of the above oscillator operators they are

$$\begin{aligned}&J_\alpha=J_\alpha(\mathcal{C},13)+J_\alpha(\mathcal{C},24):\\ &J_0=\tfrac12(a_1^+a_1-a_3^+a_3+a_2^+a_2-a_4^+a_4),\\ &J_1=\tfrac14(a_1^+a_1^+-a_3^+a_3^++a_1a_1-a_3a_3+a_2^+a_2^+-a_4^+a_4^++a_2a_2-a_4a_4),\\ &J_2=-(i/4)(a_1^+a_1^++a_3^+a_3^+-a_1a_1-a_3a_3+a_2^+a_2^++a_4^+a_4^+-a_2a_2-a_4a_4).\end{aligned} \tag{1.3}$$

To disclose the symmetries of these generators, we switch over to new operators b_μ, b_μ^+ defined as

$$b_1=a_1,\quad b_2=a_2,\quad b_3=-a_3^+,\quad b_4=-a_4^+. \tag{1.4}$$

Using the diagonal metric tensor $g_{\mu\nu}$, with $g_{11}=g_{22}=-g_{33}=-g_{44}=+1$ for raising and lowering of Greek indices, Eq. (1.2) becomes

Copyright © 1974 American Institute of Physics

$$[b_\mu, b_\nu^+] = g_{\mu\nu}, \quad [b_\mu, b_\nu] = [b_\mu^+, b_\nu^+] = 0,$$
$$b_\mu = -\frac{i}{\sqrt{2}}(x_\mu + \partial_\mu), \quad b_\mu^+ = \frac{i}{\sqrt{2}}(x_\mu - \partial_\mu), \quad \partial_\mu \equiv \frac{\partial}{\partial x^\mu}. \tag{1.5}$$

At the same time, the J_α take on a simple appearance:

$$J_0 = \tfrac{1}{2}(g^{\mu\nu} b_\mu^+ b_\nu + 2),$$
$$J_1 = \tfrac{1}{4} g^{\mu\nu}(b_\mu^+ b_\nu^+ + b_\mu b_\nu), \tag{1.6}$$
$$J_2 = -(i/4) g^{\mu\nu}(b_\mu^+ b_\nu^+ - b_\mu b_\nu).$$

It is clear that the basic commutation relations (1.5) and the total $SU(1,1)$ generators J_α are unchanged when we perform a real linear transformation

$$x_\mu \to O_\mu{}^\nu x_\nu, \quad b_\mu \to O_\mu{}^\nu b_\nu, \quad b_\mu^+ \to O_\mu{}^\nu b_\nu^+ \tag{1.7}$$

that leaves the indefinite quadratic form $x^2 \equiv x^\mu x_\mu$ invariant. There is then a unitary representation of the group of matrices $\|O_\mu{}^\nu\|$ acting on $\mathcal{H}$ and commuting with the representation $\mathcal{C}\otimes\mathcal{C}$ of $SU(1,1)$. As in II we shall write $O(2,2)$ for the identity component of the group of matrices $\|O_\mu{}^\nu\|$; its representation on $\mathcal{H}$ is generated by the six operators $M_{\mu\nu}$:

$$M_{\mu\nu} = i(b_\mu^+ b_\nu - b_\nu^+ b_\mu) = i(x_\mu \partial_\nu - x_\nu \partial_\mu). \tag{1.8}$$

Among the improper transformations we shall make particular use of the following two:

$$P_{13}: x_1 \to -x_1, \quad x_2 \to x_2, \quad x_3 \to -x_3, \quad x_4 \to x_4, \tag{1.9}$$
$$P_{24}: x_1 \to x_1, \quad x_2 \to -x_2, \quad x_3 \to x_3, \quad x_4 \to -x_4.$$

Actually, these two transformations lie in the same component of the full group because of the relation

$$P_{24} = \exp(i\pi M_{12}) \exp(i\pi M_{34}) P_{13}. \tag{1.10}$$

We shall write G for the group made up of the identity component, $O(2,2)$, and the component containing P_{13} (and P_{24}). The relations between $O(2,2)$ and P_{13} will be worked out in the next section. We may note here that the product $P_{13}P_{24}$, corresponding to the transformation $x_\mu \to -x_\mu$ belongs to $O(2,2)$ and commutes with all elements of G.

The symmetry properties of the operators J_α under G are expressed by

$$[J_\alpha, M_{\mu\nu}] = 0, \quad P_{13} J_\alpha P_{13} = J_\alpha. \tag{1.11}$$

For the individual sets of generators we have the lesser symmetries

$$[J_\alpha(\mathcal{C},13) \text{ or } J_\alpha(\mathcal{C},24),\ M_{13} \text{ or } M_{24} \text{ or } P_{13} \text{ or } P_{24}] = 0. \tag{1.12}$$

We may remind the reader that in reducing the representation $\mathcal{C}$ on $\mathcal{H}(\mathcal{C},13)$, for instance, into UIR's, one has to simultaneously diagonalize M_{13} and P_{13}.

It is important to know, in the present kind of direct product, how the outer automorphism $\tau: J_0 \to -J_0$, $J_1 \to -J_1$, $J_2 \to J_2$ is implemented. For the first factor in the product $\mathcal{C}\otimes\mathcal{C}$, defined on $\mathcal{H}(\mathcal{C},13)$, it is implemented by the unitary operator $\mathcal{A}_{13}$:

$$(\mathcal{A}_{13} f)(x_1, x_2, x_3, x_4) = f(x_3, x_2, x_1, x_4). \tag{1.13}$$

Similarly, for the second factor, it is implemented by $\mathcal{A}_{24}$:

$$(\mathcal{A}_{24} f)(x_1, x_2, x_3, x_4) = f(x_1, x_4, x_3, x_2). \tag{1.14}$$

[Recall that $\mathcal{A}_{13}$ commutes with both M_{13} and P_{13}; on the other hand, a subspace of $\mathcal{H}(\mathcal{C},13)$ which is an eigenspace of both M_{13} and P_{13} carries just one UIR C_q^ϵ occurring in the reducible representation $\mathcal{C}$ with P_{13} determining ϵ, M_{13} determining q. Therefore, $\mathcal{A}_{13}$ acts within each UIR obtained in this way, not connecting it to another appearance of the same UIR on another subspace of $\mathcal{H}(\mathcal{C},13)$. The same holds for $\mathcal{A}_{24}$.] For the total generators J_α, the automorphism τ is implemented by

$\mathcal{A} = \mathcal{A}_{13}\mathcal{A}_{24}$:

$$(\mathcal{A} f)(x_1, x_2, x_3, x_4) = f(x_3, x_4, x_1, x_2),$$
$$\mathcal{A} J_0 \mathcal{A} = -J_0, \quad \mathcal{A} J_1 \mathcal{A} = -J_1, \quad \mathcal{A} J_2 \mathcal{A} = J_2. \tag{1.15}$$

And if we pick a particular product $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ from out of $\mathcal{C}\otimes\mathcal{C}$ on account of the above-mentioned properties of $\mathcal{A}_{13}$ and $\mathcal{A}_{24}$, we see that $\mathcal{A}$ carries this product into itself. This operator $\mathcal{A}$ will be used in the sequel.

The connections between the various Casimir operators, established in previous cases, hold good again. For the invariants associated with the two factors in the product $\mathcal{C}\otimes\mathcal{C}$, we have

$$Q_{13} = \tfrac{1}{4}(1 + M_{13}^2), \quad Q_{24} = \tfrac{1}{4}(1 + M_{24}^2). \tag{1.16}$$

For the total $SU(1,1)$ Casimir operator Q, we have

$$Q \equiv (J_1)^2 + (J_2)^2 - (J_0)^2 = -\tfrac{1}{8} M^2,$$
$$M^2 = M^{\mu\nu} M_{\mu\nu}, \tag{1.17}$$

where M^2 is one of the two $O(2,2)$ invariants. The other one, $\epsilon_{\mu\nu\rho\sigma} M^{\mu\nu} M^{\rho\sigma}$ vanishes on account of the special form of $M_{\mu\nu}$.

We shall next specify the natures of the uncoupled and coupled bases for $\mathcal{H}$, with whose help the C–G series and coefficients will be determined. In the uncoupled basis, the simultaneously diagonal operators will be M_{13}, P_{13}, $J_2(\mathcal{C},13)$, $\mathcal{A}_{13}$ and M_{24}, P_{24}, $J_2(\mathcal{C},24)$, $\mathcal{A}_{24}$; thus a particular product $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ will be singled out, the eigenvalues of M_{13} and P_{13} determining q and ϵ, those of M_{24} and P_{24} determining q' and ϵ'. In the coupled basis, M_{13}, P_{13}, M_{24}, and P_{24} will again be diagonal, and to these will be added Q and J_2. In addition, when $Q > \frac{1}{4}$, the operator $\mathcal{A}$ and one other operator will be diagonal; this latter one is necessary because of the double appearance of each continuous class UIR $C_{q''}^{\epsilon''}$ in a product $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$.

In the following sections, we shall describe the construction of a complete set of $O(2,2)$ "spherical harmonics" in a basis in which M_{13}, M_{24}, and M^2 are diagonal, and then by including the action of P_{13} (and P_{24}) extend them to bases for certain UIR's of the larger group G. We conclude this section by recording the equations that isolate the "angular dependences" of the J_α in the operator Q; the method is exactly the same as in the previous cases. We have

$$J_0 = \tfrac{1}{4}\left(x^2 - \frac{1}{x^2}(x\cdot\partial)^2 - \frac{2}{x^2}\, x\cdot\partial - \frac{4}{x^2} Q\right),$$
$$J_1 = -\tfrac{1}{4}\left(x^2 + \frac{1}{x^2}(x\cdot\partial)^2 + \frac{2}{x^2}\, x\cdot\partial + \frac{4}{x^2} Q\right),$$

$$J_2 = -(i/2)(x\cdot\partial + 2), \quad x\cdot\partial \equiv x^\mu\partial_\mu. \tag{1.18}$$

2. STRUCTURE OF G AND CHOICE OF ANGULAR VARIABLES

Let us divide the four-dimensional space into two regions, V_+ and V_- accordingly as the indefinite form x^2 is either positive or negative. The surface $x^2=0$ can be ignored. A function $f(x)\in\mathcal{H}$ can be written as a column vector with two entries, $f_-(x)$ and $f_+(x)$, giving its values in V_- and V_+, respectively:[2]

$$f\in\mathcal{H}\quad f=\begin{pmatrix} f_-(x)\\ f_+(x)\end{pmatrix},$$

$$\|f\|^2=\int_{V_-} d^4x\,|f_-(x)|^2+\int_{V_+} d^4x\,|f_+(x)|^2. \tag{2.1}$$

We shall write $\mathcal{H}_\pm$ for the two subspace corresponding to $f_\mp=0$ respectively; so $\mathcal{H}=\mathcal{H}_+\oplus\mathcal{H}_-$. We shall make use of the mapping P defined by

$$P(x_1,x_2,x_3,x_4)=(x_3,x_4,x_1,x_2), \tag{2.2}$$

We shall write $\mathcal{H}_\pm$ for the two subspaces corresponding to given in Eq. (1.15) and implements the automorphism τ on the total generators J_α can be rewritten using P:

$$(\mathcal{A}f)\,(x)=f(Px). \tag{2.3}$$

Note that P is not contained in the symmetry group G, and is not related to P_{13} and P_{24}.

We need a convenient description of G and its action on x_μ in a form that will suggest a natural choice of the angular variables and spherical harmonics. Let us begin with the structure of $O(2,2)$. We recall from Sec. 2 of II, we shall make each direction within V_+ correspond which can be uniquely parametrized using four real variables (a_1,a_2,a_3,a_4), as

$$g(a)=a_1-ia_2\sigma_3-a_3\sigma_1-a_4\sigma_2,$$
$$a_1^2+a_2^2-a_3^2-a_4^2=1, \tag{2.4}$$

there are two special transformations $L(g)$ and $R(g)$ of $O(2,2)$:

$$L(g)=\begin{pmatrix} a_1 & -a_2 & a_3 & a_4\\ a_2 & a_1 & a_4 & -a_3\\ a_3 & a_4 & a_1 & -a_2\\ a_4 & -a_3 & a_2 & a_1\end{pmatrix},\quad R(g)=\begin{pmatrix} a_1 & a_2 & -a_3 & -a_4\\ -a_2 & a_1 & a_4 & -a_3\\ -a_3 & a_4 & a_1 & -a_2\\ -a_4 & -a_3 & a_2 & a_1\end{pmatrix}. \tag{2.5}$$

These transformations obey

$$L(g')L(g)=L(g'g),\quad R(g')R(g)=R(g'g),$$
$$L(g')R(g)=R(g)\,L(g'), \tag{2.6}$$

and the general element in the identity component of $O(2,2)$ can be written as $L(g)R(g')$, g and g' being independent $SU(1,1)$ elements. In this way, the $SU(1,1)\otimes SU(1,1)$ structure of $O(2,2)$ (locally) is made manifest. The structure of G is specified by stating the following relations, which are easy to verify, in addition to (2.6):

$$P_{13}L(g)P_{13}=P_{24}L(g)P_{24}=L(\tau(g)),$$
$$P_{13}R(g)P_{13}=P_{24}R(g)P_{24}=R(\tau(g));$$
$$P_{13}P_{24}=P_{24}P_{13}=-\mathbf{1},\quad P_{13}{}^2=P_{24}{}^2=\mathbf{1}. \tag{2.7}$$

For convenience, the properties of P_{24} have been written down, though they follow from those of P_{13} and Eq. (1.10). The change $g\to\tau(g)$ corresponds to a change in sign of the parameters a_2 and a_4 in Eq. (2.4).

In order to describe the action of G on x_μ as in Sec. 2 of II, we shall make each direction within V_+ correspond to one element of $SU(1,1)$ in a reversible way, and similarly for V_-. So we define

$$x\in V_+: x^2=r^2,\quad r>0,$$
$$a(x)=r^{-1}(x_1-ix_2\sigma_3-x_3\sigma_1-x_4\sigma_2),$$
$$x\in V_-: x^2=-r^2,\quad r>0,$$
$$a(x)=r^{-1}(x_3-ix_4\sigma_3-x_1\sigma_1-x_2\sigma_2),$$
$$a(Px)=a(x). \tag{2.8}$$

Then the transformations of G applied to x_μ can be described in this way:

$$a(L(g)x)=g\,a(x),\quad \text{all } x,$$
$$a(R(g)x)=a(x)g^{-1}\ \text{ if } x\in V_+,\quad a(x)\tau(g)^{-1}\ \text{ if } x\in V_-, \tag{2.9a}$$

$$a(P_{13}x)=-\mathbf{1}\cdot\tau(a(x)),\quad \text{all } x,$$
$$a(P_{24}x)=\tau(a(x)),\ \text{all } x. \tag{2.9b}$$

The replacement of x_μ by $\{r, a(x)\}$ is to be viewed as the passage from Cartesian to radial and (generalized) angular variables; within each of the regions V_+ and V_-, the $SU(1,1)$ element is the "angle". Once a definite parameterization for $SU(1,1)$ is adopted, those parameters become ordinary "angular" variables. Each of the functions $f_\pm(x)$ in Eq. (2.1) can be written as a function $f_\pm(r;a(x))$ and the squared norm of f is

$$\|f\|^2=2\pi^2\int_0^\infty r^3dr\int_{su(1,1)} dg(|f_-(r;g)|^2+|f_+(r;g)|^2). \tag{2.10}$$

Here, dg is the invariant volume element on $SU(1,1)$, written down in Eq. (2.20) of II.

The parametrization that we seek for $SU(1,1)$ must be such as to make M_{13}, M_{24} particularly simple. Since $g_{11}=-g_{33}$ and $g_{22}=-g_{44}$ each of these operators generates hyperbolic, and not Euclidean, rotations. The appropriate parametrization for $SU(1,1)$, in these circumstances, is the $O(1,1)$-parametrization,[3] and not the Bargmann parametrization which was used in II in decomposing the representation $\mathcal{D}^+\otimes\mathcal{D}^-$. This parametrization of $SU(1,1)$ can be explained briefly as follows.[4] For the moment, denote the generators of the defining two-dimensional representation of $SU(1,1)$ by J_0, J_1, J_2 according to

$$J_0=\tfrac{1}{2}\sigma_3,\quad J_1=(i/2)\sigma_2,\quad J_2=-(i/2)\sigma_1. \tag{2.11}$$

Then, barring a set of measure zero, every element $g(a)$ in $SU(1,1)$ can be written uniquely in the form

$$g(a)=\exp(i\zeta J_2)\,X(a)\exp(i\zeta' J_2), \tag{2.12}$$

where $X(a)$ is a suitable and simple element of the group. Depending on the signs of the expressions $a_1^2-a_3{}^2$, $a_2^2-a_4^2$ etc., [cf., Eq. (2.4)], the group space can be divided into five disjoint regions and in each of them a particular form of $X(a)$ is to be used. The definitions of these regions, and the decomposition in the above fashion to be used in each of them, are as follows:

(i) Region R: $a_1^2 > a_3^2$, $a_2^2 > a_4^2$,

$g(a) = \exp(i\zeta J_2)\exp(i\mu J_0)\exp(i\zeta' J_2)$, $-\infty < \zeta, \zeta' < \infty$, $-2\pi \leqslant \mu \leqslant 2\pi$;

(ii) Region S_0: $a_1^2 > a_3^2$, $a_2^2 < a_4^2$, $a_1 \geqslant 1$,

$g(a) = \exp(i\zeta J_2)\exp(i\nu J_1)\exp(i\zeta' J_2)$, $-\infty < \zeta, \zeta', \nu < \infty$;

(iii) Region S_1: $a_1^2 < a_3^2$, $a_2^2 > a_4^2$, $a_2 \leqslant -1$,

$g(a) = \exp(i\zeta J_2)\exp(i\nu J_1)\exp(i\pi J_0)\exp(i\zeta' J_2)$,

$-\infty < \zeta, \zeta', \nu < \infty$;

(iv) Region S_2: $a_1^2 > a_3^2$, $a_2^2 < a_4^2$, $a_1 \leqslant -1$,

$g(a) = \exp(i\zeta J_2)\exp(i\nu J_1)\exp(2\pi i J_0)\exp(i\zeta' J_2)$,

$-\infty < \zeta, \zeta', \nu < \infty$;

(v) Region S_3: $a_1^2 < a_3^2$, $a_2^2 > a_4^2$, $a_2 \geqslant 1$,

$g(a) = \exp(i\zeta J_2)\exp(i\nu J_1)\exp(3\pi i J_0)\exp(i\zeta' J_2)$,

$\infty < \zeta, \zeta', \nu < \infty$. (2.13)

Using Eq. (2.11), one can express the a's in each region in terms of the new parameters, but we will not do that here since the related expressions for the x_μ will be given.

Now, on the basis of the above parametrization for $SU(1,1)$, we divide the region V_+ into five disjoint subregions, based on the nature of $a(x)$. Thus, the points $x \in V_+$ for which $a(x) \in R \subset SU(1,1)$ will comprise the region $V_{+,R} \subset V_+$, those for which $a(x) \in S_0 \subset SU(1,1)$ give the region $V_{+,S_0} \subset V_+$ and so on. So then the passage from the Cartesian variables x_μ to new radial and "angular" ones, in V_+, is obtained by combining the first line of Eq. (2.8) with Eq. (2.13) and gives the following results:

$V_{+,R}$: $x_1^2 > x_3^2$, $x_2^2 > x_4^2$,

$x_1 = r\cos(\mu/2)\cosh(\zeta_+/2)$, $x_2 = -r\sin(\mu/2)\cosh(\zeta_-/2)$,

$x_3 = -r\cos(\mu/2)\sinh(\zeta_+/2)$, $x_4 = r\sin(\mu/2)\sinh(\zeta_-/2)$;

V_{+,S_0}: $x_1^2 > x_3^2$, $x_2^2 < x_4^2$, $x_1 \geqslant r$,

$x_1 = r\cosh(\nu/2)\cosh(\zeta_+/2)$, $x_2 = -r\sinh(\nu/2)\sinh(\zeta_-/2)$,

$x_3 = -r\cosh(\nu/2)\sinh(\zeta_+/2)$, $x_4 = r\sinh(\nu/2)\cosh(\zeta_-/2)$;

V_{+,S_1}: $x_1^2 < x_3^2$, $x_2^2 > x_4^2$, $x_2 \leqslant -r$,

$x_1 = r\sinh(\nu/2)\sinh(\zeta_+/2)$, $x_2 = -r\cosh(\nu/2)\cosh(\zeta_-/2)$,

$x_3 = -r\sinh(\nu/2)\cosh(\zeta_+/2)$, $x_4 = r\cosh(\nu/2)\sinh(\zeta_-/2)$;

V_{+,S_2}: $x_1^2 > x_3^2$, $x_2^2 < x_4^2$, $x_1 \leqslant -r$,

$x_1 = -r\cosh(\nu/2)\cosh(\zeta_+/2)$, $x_2 = r\sinh(\nu/2)\sinh(\zeta_-/2)$,

$x_3 = r\cosh(\nu/2)\sinh(\zeta_+/2)$, $x_4 = -r\sinh(\nu/2)\cosh(\zeta_-/2)$;

V_{+,S_3}: $x_1^2 < x_3^2$, $x_2^2 > x_4^2$, $x_2 \geqslant r$,

$x_1 = -r\sinh(\nu/2)\sinh(\zeta_+/2)$, $x_2 = r\cosh(\nu/2)\cosh(\zeta_-/2)$,

$x_3 = r\sinh(\nu/2)\cosh(\zeta_+/2)$, $x_4 = -r\cosh(\nu/2)\sinh(\zeta_-/2)$;

$\zeta_\pm = \zeta' \pm \zeta$. (2.14)

To introduce new variables in V_-, we just make use of the mapping P in Eq. (2.2). This only involves interchanging x_1 and x_3, x_2 and x_4 in the above. The mapping P applied to $V_{+,R}$, V_{+,S_0}, ..., V_{+,S_3} generates $V_{-,R}$, V_{-,S_0}, ..., V_{-,S_3}, respectively.

Let us now spell out the way in which elements of $\mathcal{H}$ are to be described. Going back to Eq. (2.1), we replace $f_+(x)$ by a collection of five functions in this way:

$$x \in V_{+,R}:\ f(x) = f_+(r;\zeta\mu\zeta'),\qquad x \in V_{+,S_n}:\ f(x) = f_{+,n}(r;\zeta\nu\zeta'),\quad n = 0,1,2,3. \tag{2.15}$$

In an exactly similar fashion, $f_-(x)$ is replaced by $f_-(r;\zeta\mu\zeta')$ and $f_{-,n}(r;\zeta\nu\zeta')$. So a general element $f \in \mathcal{H}$ consists now of a column vector with ten entries:

$$f = \begin{pmatrix} f_-(r;a(x)) \\ \\ f_+(r;a(x)) \end{pmatrix} = \begin{pmatrix} f_-(r;\zeta\mu\zeta') \\ f_{-,0}(r;\zeta\nu\zeta') \\ \vdots \\ f_{-,3}(r;\zeta\nu\zeta') \\ f_+(r;\zeta\mu\zeta') \\ f_{+,0}(r;\zeta\nu\zeta') \\ \vdots \\ f_{+,3}(r;\zeta\nu\zeta') \end{pmatrix},\qquad \begin{array}{l} -\infty < \zeta, \zeta', \nu < \infty \\ -2\pi \leqslant \mu \leqslant 2\pi. \end{array} \tag{2.16}$$

And after evaluating the $SU(1,1)$-volume element in the new parameters, Eq. (2.10) becomes

$$\|f\|^2 = \int_0^\infty 2\pi^2 r^3 dr (32\pi^2)^{-1} \int_{-\infty}^{\infty} d\zeta' \int_{-\infty}^{\infty} d\zeta \Big[\int_{-2\pi}^{2\pi} |\sin\mu|\, d\mu (|f_-(r;\zeta\mu\zeta')|^2 + |f_+(r;\zeta\mu\zeta')|^2) + \sum_{n=0}^{3} \int_{-\infty}^{\infty} |\sinh\nu|\, d\nu (|f_{-,n}(r;\zeta\nu\zeta')|^2 + |f_{+,n}(r;\zeta\nu\zeta')|^2)\Big]. \tag{2.17}$$

This way of describing elements of $\mathcal{H}$ appears rather cumbersome, and can be simplified under certain conditions. In practice we will always need to deal with eigenfunctions of the two operators P_{13}, P_{24} in Eq. (1.9). Thus, to deal with the direct product $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ within $\mathcal{C} \otimes \mathcal{C}$ and its reduction, we need to work only with eigenfunctions of P_{13} and P_{24} with eigenvalues η_ϵ, $\eta_{\epsilon'}$, respectively, and the specification of such an element is simpler than Eq. (2.16). (Here, $\eta_0 = +1$ and $\eta_{1/2} = -1$.) Starting with a general element f as in Eq. (2.16), and imposing the two conditions

$$P_{13} f = \eta_\epsilon f,\quad P_{24} f = \eta_{\epsilon'} f, \tag{2.18}$$

We get the following consequences on f_+:

$$\begin{aligned} f_+(r;\zeta\mu\zeta') &= \eta_\epsilon f_+(r;\zeta, 2\pi-\mu, \zeta') = \eta_{\epsilon'} f_+(r;\zeta, -\mu, \zeta'), \\ f_{+,0}(r;\zeta\nu\zeta') &= \eta_\epsilon f_{+,2}(r;\zeta, -\nu, \zeta') = \eta_{\epsilon'} f_{+,0}(r;\zeta, -\nu, \zeta'), \\ f_{+,1}(r;\zeta\nu\zeta') &= \eta_\epsilon f_{+,1}(r;\zeta, -\nu, \zeta') \\ &= \eta_{\epsilon'} f_{+,3}(r;\zeta, -\nu, \zeta'). \end{aligned} \tag{2.19}$$

Exactly similar equations hold with f_- in place of f_+. So such an f is completely determined by knowing the functions $f_\pm(r;\zeta\mu\zeta')$ for $-\pi \leqslant \mu \leqslant 0$, $f_{\pm,0}(r;\zeta\nu\zeta')$ for $\nu \geqslant 0$, and $f_{\pm,3}(r;\zeta\nu\zeta')$ for $\nu \geqslant 0$. (This choice of independent "parts" of f is made since it corresponds to covering the regions $x_3 > |x_1|$, $x_1 > |x_3|$ in the $x_1 - x_3$ variables, and $x_4 > |x_2|$, $x_2 > |x_4|$ in the $x_2 - x_4$ variables). With the understanding then that we are speaking of an eigenfunction of P_{13} and P_{24}, we can replace Eq. (2.16) by something simpler, namely

$$f=\begin{pmatrix} f_-(r;\zeta\mu\zeta') \\ f_{-,0}(r;\zeta\nu\zeta') \\ f_{-,3}(r;\zeta\nu\zeta') \\ f_+(r;\zeta\mu\zeta') \\ f_{+,0}(r;\zeta\nu\zeta') \\ f_{+,3}(r,\zeta\nu\zeta') \end{pmatrix}, \quad -\pi\leqslant\mu\leqslant 0, \quad 0\leqslant\nu<\infty. \tag{2.20}$$

And if f' is another eigenfunction of P_{13}, P_{24} with the *same* eigenvalues as f, then the scalar product of f' with f reads

$$(f',f)=\int_0^\infty 2\pi^2 r^3\, dr(8\pi^2)^{-1}\int_{-\infty}^\infty d\zeta' \int_{-\infty}^\infty d\zeta\Big(\int_{-\pi}^0(-\sin\mu)\, d\mu$$
$$\times(f'_-(r;\zeta\mu\zeta')^* f_-(r;\zeta\mu\zeta')+f'_+(r;\zeta\mu\zeta')^* f_+(r;\zeta\mu\zeta'))$$
$$+\int_0^\infty \sinh\nu d\nu \sum_{n=0,3}(f'_{-,n}(r;\zeta\nu\zeta')^* f_{-,n}(r;\zeta\nu\zeta') \tag{2.21}$$
$$+f'_{+,n}(r;\zeta\nu\zeta')^* f_{+,n}(r;\zeta\nu\zeta'))\Big).$$

In constructing the uncoupled and coupled basis vectors later on, we shall always display them in the form (2.20).

Finally, in concluding this section, let us give the forms for M_{13}, M_{24}, and comment on the others. Of course, the representation of $O(2,2)$ we are dealing with at present is just the same as the one encountered in II but only expressed in a different basis; so we can take over from II the steps by which the $M_{\mu\nu}$ are replaced by suitable linear combinations which make up two mutually commuting $SU(1,1)$ Lie algebras, and also the result that these two independent algebras share the same Casimir operator which in turn coincides with Q. The combinations $M_{13}\pm M_{24}$ belong to these two commuting $SU(1,1)$ Lie algebras. Each of the $M_{\mu\nu}$, in each of the regions $V_{-,R}$, V_{-,S_n}, $V_{+,R}$, V_{+,S_n}, is a partial differential operator in the angular variables appropriate to that region (ζ, μ, ζ' or ζ, ν, ζ' as the case may be). M_{13} and M_{24} have, in all regions, the forms

$$M_{13}=i\Big(\frac{\partial}{\partial\zeta'}+\frac{\partial}{\partial\zeta}\Big), \quad M_{24}=i\Big(\frac{\partial}{\partial\zeta'}-\frac{\partial}{\partial\zeta}\Big). \tag{2.22}$$

For the rest, the use of the mapping P shows that M_{34}, M_{41}, M_{12} and M_{32} in any subregion of V_- have the same expression as M_{21}, M_{32}, M_{43}, and M_{41}, respectively, in the corresponding subregion of V_+. But we do not need these explicit expressions in our work.

3. CONSTRUCTION OF SPHERICAL HARMONICS FOR THE GROUP G

With the preparation of the previous section, we are now in a position to construct complete sets of functions of the "angular" variables, both for V_+ and for V_-, which will form bases for certain irreducible representations of G. The interesting point will be to ascertain what representations of G appear in $\mathcal{H}_+$, and which ones in $\mathcal{H}_-$. In general, a UIR of G may be composed of several UIR's of the subgroup $O(2,2)$, though sometimes it may remain irreducible under this subgroup. We will first construct complete sets of functions forming bases for UIR's of $O(2,2)$, and then see how to form bases for UIR's of G.

The construction of a complete set of $O(2,2)$ harmonics, in V_+ or in V_- involves just the Plancherel theorem for $SU(1,1)$, exactly as in II. The only difference is that now the theorem needs to be stated in the $O(1,1)$ basis. Using the notation introduced in Ref. 4(b), the "matrix" representing the element g of $SU(1,1)$ in the UIR $\mathcal{R}$ and in a basis with J_2 diagonal is written as

$$\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g). \tag{3.1}$$

Here, p' and p are the eigenvalues of J_2 in the two states between which the matrix element is being evaluated; and the subscripts b and a, which are present only if $\mathcal{R}=C_q^\epsilon$, are the eigenvalues of the operator implementing τ. For g in each of the five regions R, S_n of $SU(1,1)$, we have a special form for $\mathcal{D}(g)$. Namely, using the parametrization of Eq. (2.13), we have

$$g\in R:\ \mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g)=\exp[i(\zeta p'+\zeta' p)]\,\mathcal{G}^{(\mathcal{R})}_{ba}(p',p;\mu);$$
$$g\in S_n,\ n=0,1,2,3:\ \mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g)=\exp[i(\zeta p'+\zeta' p)]$$
$$\times\mathcal{J}^{(\mathcal{R})}_{ba}(p',p;\nu;n). \tag{3.2}$$

The explicit expressions for the $\mathcal{G}$'s and $\mathcal{J}$'s may be found in Ref. 4. An important property of these representation matrices is that under the automorphism $g\to\tau(g)$

$$\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(\tau(g))=ba\,\mathcal{D}^{(\tau(\mathcal{R}))}_{p'b,pa}(g). \tag{3.3}$$

Of course, $\tau(k,\eta)=(k,-\eta)$ and $\tau(s,\epsilon)=(s,\epsilon)$. The effect of τ on the various regions in $SU(1,1)$ is $\tau(R)=R$, $\tau(S_0)=S_0$, $\tau(S_1)=S_3$, $\tau(S_2)=S_2$.

The orthogonality and completeness properties of the $\mathcal{D}$'s can be summarized thus:

$$\int_{SU(1,1)} dg\,\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g)^*\,\mathcal{D}^{(\mathcal{R}')}_{p'''d,p''c}(g)$$
$$=\delta(\mathcal{R},\mathcal{R}')\,\delta(p'''-p')\,\delta(p''-p)$$
$$\times\frac{\delta_{bd}\,\delta_{ac}}{\mu(\mathcal{R})\,\mu(\mathcal{R}')}, \tag{3.4a}$$

$$\int d\mathcal{R}\,\mu^2(\mathcal{R})\int_{-\infty}^\infty dp'\int_{-\infty}^\infty dp\sum_{b,a}\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g)\,\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(g')^*$$
$$=\delta(g,g'). \tag{3.4b}$$

The meaning of integration over $SU(1,1)$ in the new parametrization can be understood from a comparison of Eqs. (2.10) and (2.17); while the process of integration over the UIR's $\mathcal{R}$ as well as the weight factor $\mu(\mathcal{R})$ and the symbol $\delta(\mathcal{R},\mathcal{R}')$ are all explained in Sec. 3 of II.

The $O(2,2)$ spherical harmonics for the region V_+ can be defined as follows:

$$\mathcal{Y}^{+(\mathcal{R})}_{(p'b,pa)}(x)=\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(a(x)), \quad x\in V_+. \tag{3.5}$$

All the labels $p'b$, pa collectively form a composite index like the "m" in the three-dimensional spherical harmonics Y^l_m; and $\mathcal{R}$ goes over all those UIR's of $SU(1,1)$ that appear in the Plancherel formula. The manner in which the above set of functions transforms under $O(2,2)$, for any fixed $\mathcal{R}$, can be obtained by using Eq. (2.9a):

$$\mathcal{Y}^{+(\mathcal{R})}_{(p'b,pa)}(L(g_1)R(g_2)x)=\int dp'''\int dp''\sum_{c,d}\mathcal{D}^{(\mathcal{R})}_{p'b,p'''d}(g_1) \tag{3.6}$$
$$\times\mathcal{D}^{(\mathcal{R})}_{pa,p''c}(g_2)^*\,\mathcal{Y}^{+(\mathcal{R})}_{(p'''d,p''c)}(x).$$

From here we can read off the UIR of $O(2,2)$ for which the functions (3.5) form a basis: it is the UIR $(\mathcal{R}, \tau(\mathcal{R}))$ exactly as we found in II. [Since at least locally $O(2,2)$ is the direct product of two independent $SU(1,1)$ subgroups, namely of the subgroup consisting of $L(g)$ and the one consisting of $R(g)$, a UIR of $O(2,2)$ is the direct product of one UIR for each subgroup and so is denoted $(\mathcal{R}_1, \mathcal{R}_2)$.] Therefore, the UIR's $((k,+),(k,-))$ and $((k,-),(k,+))$ of $O(2,2)$ appear once each in $\mathcal{H}_+$ for $k=1, \frac{3}{2}, \ldots$, while the UIR's $((s,\epsilon),(s,\epsilon))$ appear once each for $s\geq 0$ and $\epsilon=0, \frac{1}{2}$. Next, to discover what UIR's of the larger group G appear in $\mathcal{H}_+$ we must use the behavior of the functions $\mathcal{Y}^{+(\mathcal{R})}(x)$ under P_{13} and P_{24}. Combining Eqs. (2.9b) and (3.3), we find

$$\mathcal{Y}^{+(\mathcal{R})}_{(p'b,pa)}(P_{13}x)=\eta_{\mathcal{R}}\, ba\, \mathcal{Y}^{+(\tau(\mathcal{R}))}_{(p'b,pa)}(x),$$
$$\mathcal{Y}^{+(\mathcal{R})}_{(p'b,pa)}(P_{24}x)=ba\, \mathcal{Y}^{+(\tau(\mathcal{R}))}_{(p'b,pa)}(x). \tag{3.7}$$

Here, $\eta_{\mathcal{R}}$ is $+1$ or -1 according as $\mathcal{R}$ is an integral or half-integral UIR of $SU(1,1)$. We see that, for each k, the operations P_{13} and P_{24} in G mix the two UIR's $((k,+),(k,-))$ and $((k,-),(k,+))$ of $O(2,2)$, so the corresponding two sets of basis functions combine to form the basis for one UIR of G. We shall refer to this discrete sequence of UIR's of G by G_k^+; the superscript indicates the subspace $\mathcal{H}_+$ wherein they occur. If, on the other hand, we set $\mathcal{R}=(s,\epsilon)$ in (3.7), we see that P_{13} and P_{24} carry the basis functions of the UIR $((s,\epsilon),(s,\epsilon))$ of $O(2,2)$ into themselves, not mixing them up with any other UIR of $O(2,2)$. So the corresponding set of basis functions forms, with no extension, the basis for a UIR of G as well; we shall write $G_{s,\epsilon}$ for this UIR. No superscript indicating the subspace $\mathcal{H}_+$ is necessary in this case, as we will soon see. To summarize the situation in $\mathcal{H}_+$: We have a discrete set of UIR's G_k^+ of G for $k=1, \frac{3}{2}, 2, \ldots$, and then a continuum $G_{s,\epsilon}$ for $s\geq 0$, $\epsilon=0, \frac{1}{2}$. Each G_k^+ is reducible under $O(2,2)$, containing the two UIR's $((k,+),(k,-))$ and $((k,-),(k,+))$ of the subgroup; each $G_{s,\epsilon}$ remains irreducible under $O(2,2)$, yielding the single UIR $((s,\epsilon),(s,\epsilon))$ of the subgroup.

Turning to the region V_- let us define the $O(2,2)$ spherical harmonics as

$$\mathcal{Y}^{-(\mathcal{R})}_{(p'b,pa)}(x)=a\mathcal{D}^{(\mathcal{R})}_{p'b,pa}(a(x)),\quad x\in V_-. \tag{3.8}$$

Then Eqs. (3.6) and (3.7) are replaced by the following:

$$\mathcal{Y}^{-(\mathcal{R})}_{(p'b,pa)}(L(g_1)R(g_2)x)=\int dp'''\int dp''\sum_{c,d}\mathcal{D}^{(\mathcal{R})}_{p'b,p'''d}(g_1)\times\mathcal{D}^{\tau(\mathcal{R})}_{pa,p''c}(g_2)^*\,\mathcal{Y}^{-(\mathcal{R})}_{(p'''d,p''c)}(x), \tag{3.9a}$$

$$\mathcal{Y}^{-(\mathcal{R})}_{(p'b,pa)}(P_{13}x)=\eta_{\mathcal{R}}\, ba\, \mathcal{Y}^{-(\tau(\mathcal{R}))}_{(p'b,pa)}(x),$$
$$\mathcal{Y}^{-(\mathcal{R})}_{(p'b,pa)}(P_{24}x)=ba\, \mathcal{Y}^{-(\tau(\mathcal{R}))}_{(p'b,pa)}(x). \tag{3.9b}$$

From Eq. (3.9a) we see that the functions $\mathcal{Y}^{-(\mathcal{R})}(x)$ form a basis for the UIR $((k,+),(k,+))$ of $O(2,2)$ when $\mathcal{R}=(k,+)$; for the UIR $((k,-),(k,-))$ when $\mathcal{R}=(k,-)$; and for $((s,\epsilon),(s,\epsilon))$ when $\mathcal{R}=(s,\epsilon)$. Moving up to G, Eq. (3.9b) shows that P_{13} and P_{24} mix the two UIR's $((k,+),(k,+))$ and $((k,-),(k,-))$ of $O(2,2)$, so these two sets of basis functions combine to form a basis for one UIR of G. We shall write G_k^- for this discrete sequence of UIR's of G. The important point is that each G_k^- is distinct from (i.e., not equivalent to) each G_k^+, so we have really different discrete UIR's of G on $\mathcal{H}_+$ and $\mathcal{H}_-$; this is evident from the fact that when reduced with respect to $O(2,2)$, G_k^- and G_k^+ yield different results. If, on the other hand, we set $\mathcal{R}=(s,\epsilon)$ in Eq. (3.9b), we see that both P_{13} and P_{24} act within the UIR $((s,\epsilon),(s,\epsilon))$ of $O(2,2)$; in fact, they act in exactly the same way as they did in the case of $\mathcal{H}_+$. So the basis functions for this UIR of $O(2,2)$ form a basis for a UIR of G as well, and this is just the UIR $G_{s,\epsilon}$ encountered in $\mathcal{H}_+$. To summarize: We have a discrete sequence of UIR's G_k^- of G for $k=1, \frac{3}{2}, 2, \ldots$, and then a continuum $G_{s,\epsilon}$ for $s\geq 0$, $\epsilon=0, \frac{1}{2}$, in $\mathcal{H}_-$. Each G_k^- contains the two UIR's $((k,+),(k,+))$ and $((k,-),(k,-))$ of $O(2,2)$, and so is inequivalent to the UIR G_k^+ of G appearing in $\mathcal{H}_+$. (Of course if $k\neq k'$, $G_k^{\pm}$ and $G_{k'}^{\pm}$ are obviously inequivalent.) The UIR $G_{s,\epsilon}$ of G appears once in $\mathcal{H}_-$ as it did in $\mathcal{H}_+$, and is irreducible under $O(2,2)$ as well.

Since the group G commutes with the transformations $\mathcal{C}\otimes\mathcal{C}$ of $SU(1,1)$, we conclude: Elements of $\mathcal{H}$ in $\mathcal{H}_+$ and belonging to the UIR G_k^+ of G retain these properties when acted on by $SU(1,1)$; similarly for elements in $\mathcal{H}_-$ and belonging to G_k^-; while elements in $\mathcal{H}_+$ or $\mathcal{H}_-$ belonging to $G_{s,\epsilon}$ get mixed into one another under $SU(1,1)$. We should now write down the general forms of such elements of $\mathcal{H}$. In each case, the dependence on the angular variables is fixed by the appropriate UIR of G, the radial functions are arbitrary. The spherical functions $\mathcal{Y}^{\pm(\mathcal{R})}(x)$ as defined in Eqs. (3.5, 3.8) are eigenfunctions of M_{13}, M_{24}, and Q; since we want to have P_{13}, P_{24} diagonal as well, we must form suitable linear combinations of these functions, based on Eqs. (3.7), (3.9b). This is of course the case only if $\mathcal{R}=(k,\eta)$; if $\mathcal{R}=(s,\epsilon)$, P_{13} and P_{24} are already diagonal. We may also note that both $\mathcal{Y}^{\pm(\mathcal{R})}(x)$ are eigenfunctions of the product $P_{13}P_{24}$ with eigenvalue $\eta_{\mathcal{R}}$ as follows from Eqs. (3.7), (3.9b). If therefore we are looking for an element of $\mathcal{H}$ with $P_{13}=\eta_\epsilon$, $P_{24}=\eta_{\epsilon'}$, and belonging to a definite UIR of G, then we must have $\eta_\epsilon\eta_{\epsilon'}=\eta_{\mathcal{R}}$ which restricts the possible UIR's of G that can be associated with chosen eigenvalues for P_{13} and P_{24}. This is just a reflection of the fact that in the reduction of the product $C_q^0\otimes C_{q'}^0$, we can get only integral type UIR's of $SU(1,1)$, etc.

Let us now go through the list of UIR's of G encountered in $\mathcal{H}$, and construct the general forms of vectors belonging to each. This will then make the construction of the coupled basis vectors in the following section straightforward. Since we are concerned with the reduction of the product $C_q^\epsilon\otimes C_{q'}^{\epsilon'}$, where $q=\frac{1}{4}+s^2$ and $q'=\frac{1}{4}+s'^2$, we shall work with eigenfunctions of M_{13} and M_{24} with the eigenvalues $2s$, $2s'$, respectively [cf. Eq. (1.16)]. For the present, let us indicate elements of $\mathcal{H}$ in the form of Eq. (2.1), using Eq. (2.10) for computing scalar products. Then, an element f belonging to G_k^+ with $(-1)^{2k}=+1$, and having $P_{13}=P_{24}=\pm 1$ (and furthermore with $M_{13}=2s$, $M_{24}=2s'$—this will be understood in all the following) has this form:

$$G_k^+,\quad (-1)^{2k}=+1,\quad P_{13}=P_{24}=\pm 1,$$

$$f=f_+(r)\begin{pmatrix}0\\ \mathcal{D}^{(k,+)}_{s'-s,-s'-s}(a(x))\pm\mathcal{D}^{(k,-)}_{s'-s,-s'-s}(a(x))\end{pmatrix}. \tag{3.10}$$

If $P_{13}=P_{24}=+1$, which goes with the upper (+) sign in the column vector so that we have a sum of two $\mathcal{D}$ functions, f lies in the subspace of $\mathcal{H}$ carrying a product $C_q^0\otimes C_{q'}^0$; if $P_{13}=P_{24}=-1$, going with the minus sign on the right-hand side, it belongs to a subspace carrying $C_q^{1/2}\otimes C_{q'}^{1/2}$. And now if f' is another vector of the above form, but with the replacements $f_+(r)\to f_+'(r)$, $k\to k'$, $s\to s''$, $s'\to s'''$ and both f and f' have the same eigenvalues for P_{13}, P_{24}, then from Eqs. (2.10), (3.4a) we get

$$(f',f)=[\delta_{kk'}/\mu(k)^2]\,\delta(s'''-s')\,\delta(s''-s)\int_0^\infty 2\pi^2r^3dr\,f_+'(r)^*f_+(r). \qquad (3.11)$$

Next, if f belongs to G_k^+ with $(-1)^{2k}=-1$, then $P_{13}=-P_{24}=\pm 1$,

$$G_k^+,\quad (-1)^{2k}=-1,\quad P_{13}=-P_{24}=\pm 1,$$

$$f=f_+(r)\begin{pmatrix}0\\ \mathcal{D}^{(k,+)}_{s'-s,-s'-s}(a(x))\mp\mathcal{D}^{(k,-)}_{s'-s,-s'-s}(a(x))\end{pmatrix}. \qquad (3.12)$$

The upper signs go with products $C_q^0\otimes C_{q'}^{1/2}$ the lower signs with $C_q^{1/2}\otimes C_{q'}^0$. Equations analogous to (3.11) can be easily worked out and need not be stated any further. We deal next with the UIR's G_k^-,

$$G_k^-,\quad (-1)^{2k}=+1,\quad P_{13}=P_{24}=\pm 1,$$

$$f=f_-(r)\begin{pmatrix}\mathcal{D}^{(k,+)}_{s'-s,-s'-s}(a(x))\pm\mathcal{D}^{(k,-)}_{s'-s,-s'-s}(a(x))\\ 0\end{pmatrix}. \qquad (3.13)$$

The upper signs are associated with products $C_q^0\otimes C_{q'}^0$, the lower ones with $C_q^{1/2}\otimes C_{q'}^{1/2}$. For $(-1)^{2k}=-1$, we have,

$$G_k^-,\quad (-1)^{2k}=-1,\quad P_{13}=-P_{24}=\pm 1,$$

$$f=f_-(r)\begin{pmatrix}\mathcal{D}^{(k,+)}_{s'-s,-s'-s}(a(x))\mp\mathcal{D}^{(k,-)}_{s'-s,-s'-s}(a(x))\\ 0\end{pmatrix}, \qquad (3.14)$$

with the upper signs belonging to products $C_q^0\otimes C_{q'}^{1/2}$, the lower ones to $C_q^{1/2}\otimes C_{q'}^0$.

If the representation $\mathcal{C}\otimes\mathcal{C}$ of $SU(1,1)$ is made to act on a vector which has one of the above forms in Eqs. (3.10)–(3.14), the only change will be in the radial wavefunction $f_\pm(r)$. In other words, acting on the above types of vectors, the generators J_α of Eqs. (1.6), (1.18) become just differential operators in the variable r. So, starting with Eq. (1.18) and setting $Q=k(1-k)$ and $x^2=+r^2$ therein, we see that on restriction to vectors of any of the types (3.10), (3.12) the generators $rJ_\alpha r^{-1}$ take up the form $J_\alpha(k,+)$ associated with the UIR D_k^+ of $SU(1,1)$ (see Sec. 1 of I). This signals the presence of all the UIR's D_k^+ for $k\geq 1$, once each, in the reduction of the product $C_q^\epsilon\otimes C_{q'}^{\epsilon'}$—whether k runs over the integers or the half-odd integers depends on ϵ and ϵ'. Similarly, the restrictions of $rJ_\alpha r^{-1}$ to vectors of either of the types (3.13), (3.14) is obtained by setting $Q=k(1-k)$, $x^2=-r^2$ in Eq. (1.18); and this gives us the standard form $J_\alpha(k,-)$ for the UIR D_k^- (see Sec. 1 of I). This shows that all the UIR's D_k^- for $k\geq 1$ and appropriate parity for $2k$ appear once each in any product $C_q^\epsilon\otimes C_{q'}^{\epsilon'}$.

Let us consider now vectors in $\mathcal{H}$ belonging to the UIR $G_{s'',\epsilon''}$ of G, which is present once in $\mathcal{H}_+$ and once in $\mathcal{H}_-$. Now, the action of the operator $\mathcal{A}$ of Eq. (1.5) is relevant. It is generally given by

$$\mathcal{A}\begin{pmatrix}f_-(r;a(x))\\ f_+(r;a(x))\end{pmatrix}=\begin{pmatrix}f_+(r;a(Px))\\ f_-(r;a(Px))\end{pmatrix}. \qquad (3.15)$$

We also need to make use of the property

$$x\in V_-:\quad \mathcal{Y}^{+(\mathcal{R})}_{(p'b,pa)}(Px)=a\mathcal{Y}^{-(\mathcal{R})}_{(p'b,pa)}(x). \qquad (3.16)$$

Further, the specification now of the UIR $G_{s'',\epsilon''}$ of G and the eigenvalues of M_{13}, M_{24}, P_{13}, P_{24} does not determine the form of f apart from one radial function in $\mathcal{H}_-$ and another in $\mathcal{H}_+$; this is because of the presence of the quantum numbers b, a in both spherical harmonics $\mathcal{Y}^{\pm(\mathcal{R})}_{(p'b,pa)}(x)$ when $\mathcal{R}=(s'',\epsilon'')$. These quantum numbers constitute of course part of the state labels within the UIR $G_{s'',\epsilon''}$ of G, so they too are preserved under the action of $\mathcal{C}\otimes\mathcal{C}$ just as p', p are. [Here it is important that the spherical harmonics for the region V_- were so defined in Eq. (3.8) that when $\mathcal{R}=(s,\epsilon)$ the transformation laws (3.9) were *identical* to the transformations laws (3.6), (3.7).] So in writing down the form of f in various cases, the values of b and a must also be stated. We now take up the cases one by one. Let f be a vector belonging to the representation $G_{s'',0}$ of G, and have the eigenvalue $+1$ for both P_{13} and P_{24} (and, of course, $M_{13}=2s$, $M_{24}=2s'$); we are concerned then with the occurrence of the UIR $C_{q''}^0$ of $SU(1,1)$ in the reduction of the product $C_q^0\otimes C_{q'}^0$. Then f is of one of two possible forms, corresponding to the labels b, a in the spherical harmonics obeying either $b=a=+$ or $b=a=-$. [From Eqs. (3.7), (3.9a), the eigenvalue of P_{24} determines the product ba.] So the possibilities are,

$$G_{s'',0},\quad P_{13}=P_{24}=+1,\quad b=a=\pm,$$

$$f=\begin{pmatrix}f_-(r)\,\mathcal{Y}^{-(s'',0)}_{(s'-s)\pm,(-s'-s)\pm}(x)\\ \pm f_+(r)\mathcal{Y}^{+(s'',0)}_{(s'-s)\pm,(-s'-s)\pm}(x)\end{pmatrix},$$

$$\mathcal{A}: f_-(r)\to f_+(r),\ f_+(r)\to f_-(r). \qquad (3.17)$$

A vector of the first type, corresponding to the upper signs throughout, is automatically orthogonal to one of the second type, corresponding to the lower sign throughout. They both lie in a subspace of $\mathcal{H}$ carrying the product $C_q^0\otimes C_{q'}^0$, and each preserves its form under the action of the representation $\mathcal{C}\otimes\mathcal{C}$; that is to say, under this action only the radial functions change. In particular, a vector of the first type does not get mixed into one of the second type. Thus we see in a natural way how each UIR $C_{q''}^0$ appears *twice* in the decomposition of any product $C_q^0\otimes C_{q'}^0$. Next, keeping the UIR of G unchanged, we consider the choice $P_{13}=P_{24}=-1$, so that the corresponding vectors lie in a subspace of $\mathcal{H}$ carrying the product $C_q^{1/2}\otimes C_{q'}^{1/2}$; now, $ba=-$, so the two types of vectors are,

$$G_{s'',0},\quad P_{13}=P_{24}=-1,\ b=-a=\pm,$$

$$f=\begin{pmatrix}f_-(r)\mathcal{Y}^{-(s'',0)}_{(s'-s)\pm,(-s'-s)\mp}(x)\\ \mp f_+(r)\ \mathcal{Y}^{+(s'',0)}_{(s'-s)\pm,(-s'-s)\mp}(x)\end{pmatrix},$$

$$\mathcal{A}: f_-(r)\to f_+(r),\quad f_+(r)\to f_-(r). \qquad (3.18)$$

The possibility of having these two types, again, signals, the double appearance of $C_{q''}^0$ in $C_q^{1/2}\otimes C_{q'}^{1/2}$. Switching now to the UIR $G_{s'',1/2}$ of G, we must have $P_{13}=-P_{24}$; if

$P_{13}=+1$, $P_{24}=-1$, then $ba=-$ and the two types of vectors are,

$$G_{s'',1/2}, \quad P_{13}=-P_{24}=+1, \quad b=-a=\pm,$$

$$f=\begin{pmatrix} f_-(r)\, \mathcal{Y}^{-(s'',1/2)}_{(s'-s)\pm,(-s'-s)\mp}(x) \\ \mp f_+(r)\, \mathcal{Y}^{+(s'',1/2)}_{(s'-s)\pm,(-s'-s)\mp}(x) \end{pmatrix},$$

$$\mathcal{A}: f_-(r)\rightarrow f_+(r), \quad f_+(r)\rightarrow f_-(r). \tag{3.19}$$

These two types correspond to the double occurrence of $C^{1/2}_{q''}$ in $C^0_q \otimes C^{1/2}_{q'}$. And finally, in a similar fashion, the double occurrence of $C^{1/2}_{q''}$ in $C^{1/2}_q \otimes C^0_{q'}$ is described by,

$$G_{s'',1/2}, \quad P_{13}=-P_{24}=-1, \quad b=a=\pm,$$

$$f=\begin{pmatrix} f_-(r)\, \mathcal{Y}^{-(s'',1/2)}_{(s'-s)\pm,(-s'-s)\pm}(x) \\ \pm f_+(r)\, \mathcal{Y}^{+(s'',1/2)}_{(s'-s)\pm,(-s'-s)\pm}(x) \end{pmatrix},$$

$$\mathcal{A}: f_-(r)\rightarrow f_+(r), \quad f_+(r)\rightarrow f_-(r). \tag{3.20}$$

The eight types of vectors appearing in Eqs. (3.17)—(3.20) are pairwise orthogonal, and as stated earlier each of them suffers a change in the radial functions $f_\mp(r)$ alone under the action of $\mathcal{C}\otimes\mathcal{C}$. The restrictions of the total generators J_α of Eqs. (1.6), (1.18) to any one of these eight types of vectors are just 2×2 matrices with differential operators in r as entries. These restrictions are easily obtained by starting with Eq. (1.18), replacing Q by $\frac{1}{4}+(s'')^2$, and x^2 by $\mp r^2$ accordingly as J_α acts on either $f_-(r)$ or $f_+(r)$. In this way one easily sees that on restriction to vectors of any one of the 8 forms in Eqs. (3.17)—(3.20), the generators $rJ_\alpha r^{-1}$ have the standard appearance of the generators $J_\alpha(s'', \epsilon'')$ associated with the UIR $C^{\epsilon''}_{q''}$, as set up in Sec. 1 of I. At the same time we note that we have taken care to define the radial functions $f_\mp(r)$ in such a way that in all cases the outer automorphism operator $\mathcal{A}$ just has the effect of interchanging $f_-(r)$ and $f_+(r)$; this is the standard form of $\mathcal{A}$ as well, as given in Eq. (1.19) of I. Finally, the value of ϵ'' is directly correlated with the product $P_{13}P_2$ Therefore, the interpretations given above for the vectors of types (3.17)—(3.20) are unambiguous.

4. THE BASIS VECTORS FOR $\mathcal{H}$ AND THE C-G SERIES FOR $\mathcal{C}\otimes\mathcal{C}$

We have found in the previous section the forms of vectors in $\mathcal{H}$ that are eigenfunctions of P_{13}, P_{24}, M_{13}, M_{24} and also belong to definite UIR's of G. Now we will obtain the two types of basis vectors for $\mathcal{H}$ described in Sec. 1. These vectors will be set up in the simplified form of Eq. (2.20) which is an adequate representation of eigenvectors of P_{13} and P_{24}. However, one must bear in mind the fact that the eigenvalues of P_{13} and P_{24} do not appear explicitly in the representation (2.20), but must be stated or understood separately.

Let us start with the uncoupled basis vectors Φ. We must here use the analysis of the representation $\mathcal{C}$ of $SU(1,1)$ given in Sec. 2 of I. The vector Φ belonging to the product $C^\epsilon_q \otimes C^{\epsilon'}_{q'}$ can be displayed as

$$\Phi^{(s\epsilon)\,(s'\epsilon')}_{pa\;\;p'a'} \tag{4.1}$$

with $p(p')$ being the eigenvalue of $J_2(\mathcal{C},13)(J_2(\mathcal{C},24))$, and $a(a')$ the eigenvalue of $\mathcal{A}_{13}(\mathcal{A}_{24})$. Such a vector is the product of a function of the variables x_1, x_3 and another function of x_2, x_4. It is necessary to specify the former only in the regions $x_3>|x_1|$, $x_1>|x_3|$ and the latter only in $x_4>|x_2|$, $x_2>|x_4|$. Suppose we had introduced hyperbolic variables separately for the pairs x_1, x_3 and x_2, x_4 on the lines of Eq. (2.23) of I, namely,

$$\rho\exp(\pm\eta)=\begin{cases} x_3\pm x_1 & \text{if } x_3>|x_1| \\ x_1\pm x_3 & \text{if } x_1>|x_3| \end{cases},$$

$$\rho'\exp(\pm\eta')=\begin{cases} x_4\pm x_2 & \text{if } x_4>|x_2| \\ x_2\pm x_4 & \text{if } x_2>|x_4|. \end{cases} \tag{4.2}$$

Then apart from numerical factors the vector, (4.1) is given by

$$\Phi^{(s\epsilon)\,(s'\epsilon')}_{pa\;\;p'a'} \sim \exp(2is\eta)(p)^{2ip-1}\begin{pmatrix}1\\a\end{pmatrix}\otimes\exp(2is'\eta')(p')^{2ip'-1} \times\begin{pmatrix}1\\a'\end{pmatrix}. \tag{4.3}$$

The first column vector is the above-mentioned function of x_1 and x_3 with the upper entry corresponding to the region $x_3>|x_1|$ and the lower one to $x_1>|x_3|$. Similarly, the second column vector is the function of x_2 and x_4, with entries corresponding to $x_4>|x_2|$, $x_2>|x_4|$, respectively. To put Φ into the form of Eq. (2.20), we must relate ρ, η, ρ', η' to r, ζ, ζ' and μ or ν appropriately in each region. We also note that in Eq. (2.20), the first entry gives f in the region $x_3>|x_1|$, $x_4>|x_2|$; the second and sixth give f in $x_3>|x_1|$, $x_2>|x_4|$ and correspond to $x^\mu x_\mu \gtrless 0$; the third and fifth cover $x_1>|x_3|$, $x_4>|x_2|$, and $x^\mu x_\mu \gtrless 0$, respectively; while the fourth entry gives f in $x_1>|x_3|$, $x_2>|x_4|$. Identifying the variables appropriately and putting in the normalization factors, we have

$$\Phi^{(s\epsilon)\,(s'\epsilon')}_{pa\;\;p'a'} = (2\pi)^{-2}\exp[i(s'-s)\zeta - i(s'+s)\zeta']\, r^{2i(p+p')-2}$$

$$\times\begin{pmatrix} (\cos\mu/2)^{2ip-1}(-\sin\mu/2)^{2ip'-1} \\ a'(\cosh\nu/2)^{2ip-1}(\sinh\nu/2)^{2ip'-1} \\ a(\sinh\nu/2)^{2ip-1}(\cosh\nu/2)^{2ip'-1} \\ aa'(\cos\mu/2)^{2ip-1}(-\sin\mu/2)^{2ip'-1} \\ a(\cosh\nu/2)^{2ip-1}(\sinh\nu/2)^{2ip'-1} \\ a'(\sinh\nu/2)^{2ip-1}(\cosh\nu/2)^{2ip'-1} \end{pmatrix}. \tag{4.4}$$

[This is in the notation of Eq. (2.20).] Fortunately, this six-component column vector, and the later ones corresponding to the coupled basis for $\mathcal{H}$, can be written as the direct product of a two-component vector by a three-component one, which makes the writing a little easier:

$$\Phi^{(s\epsilon)\,(s'\epsilon')}_{pa\;\;p'a'} = (2\pi)^{-2}\exp[i(s'-s)\zeta - i(s'+s)\zeta']\, r^{2i(p+p')-2}$$

$$\times\begin{pmatrix}1\\aa'\end{pmatrix}\otimes\begin{pmatrix} (\cos\mu/2)^{2ip-1}(-\sin\mu/2)^{2ip'-1} \\ a'(\cosh\nu/2)^{2ip-1}(\sinh\nu/2)^{2ip'-1} \\ a(\sinh\nu/2)^{2ip-1}(\cosh\nu/2)^{2ip'-1} \end{pmatrix}. \tag{4.5}$$

These vectors are normalized according to

$$\left(\Phi^{(s_1\epsilon_1)(s_1'\epsilon_1')}_{p_1 a\ \ p_1' a_1'},\ \Phi^{(s\epsilon)(s'\epsilon')}_{pa\ \ p'a'}\right) = \delta(s_1 - s)\,\delta(s_1' - s')\,\delta_{\epsilon_1\epsilon}\,\delta_{\epsilon_1'\epsilon'} \times \delta(p_1 - p)\,\delta(p_1' - p')\,\delta_{a_1 a}\,\delta_{a_1' a'}. \tag{4.6}$$

Turning now to the coupled basis vectors Ψ, there are essentially three types to be constructed, corresponding to the single occurrence of each UIR D_k^+ and each UIR D_k^- and the double occurrence of each UIR $C_{q''}^{\epsilon''}$ in a product $C_q^{\epsilon}\otimes C_{q'}^{\epsilon'}$. Most of the work in constructing the Ψ's has been done in arriving at the general forms in Eqs. (3.10), (3.12)–(3.14), (3.17)–(3.20) for eigenvectors of M_{13}, P_{13}, M_{24}, P_{24}, and Q. Only the radial functions are to be chosen so as to obtain eigenfunctions of J_2 and $\mathcal{A}$. The coupled vectors Ψ corresponding to the final UIR being D_k^+ or D_k^- can be written in unified form thus:

$$\Psi^{(s\epsilon)(s'\epsilon')(k+)+}_{p''}\Big)_{p''} = (4\pi^3)^{-1/2}\,\mu(k)\exp[i(s'-s)\zeta - i(s'+s)\zeta'] \times r^{2ip''-2}\begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix} \mathcal{G}^{(k+)}(s'-s,-s'-s;\mu) + \eta_{\epsilon'}\mathcal{G}^{(k-)}(s'-s,-s'-s;\mu) \\ \mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;0) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;0) \\ \mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;3) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;3) \end{pmatrix} \tag{4.7a}$$

$$\Psi^{(s\epsilon)(s'\epsilon')(k-)}_{p''} = (4\pi^3)^{-1/2}\,\mu(k)\exp[i(s'-s)\zeta - i(s'+s)\zeta'] \times r^{2ip''-2}\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix} \mathcal{G}^{(k+)}(s'-s,-s'-s;\mu) + \eta_{\epsilon'}\mathcal{G}^{(k-)}(s'-s,-s'-s;\mu) \\ \mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;0) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;0) \\ \mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;3) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;3) \end{pmatrix} \tag{4.7b}$$

Just like Φ in Eq. (4.5), these are shorthand expressions for six-component vectors as in Eq. (2.20); thus when written out in six-component form the first three entries of the vector (4.7a) vanish, while (4.7b) has its last three entries vanishing. Turning to the third type of coupled basis vector, a form valid for all cases can be given:

$$\Psi^{(s\epsilon)(s'\epsilon')(s''\epsilon'')b}_{p''a''} = (4\pi^3)^{-1/2}\,\mu(s''\epsilon'')\exp[i(s'-s)\zeta - i(s'+s)\zeta']\,r^{2ip''-2} \times \begin{pmatrix}1\\a''\end{pmatrix}\otimes\begin{pmatrix} \mathcal{G}^{(s''\epsilon'')}_{ba}(s'-s,-s'-s;\mu) \\ \mathcal{J}^{(s''\epsilon'')}_{ba}(s'-s,-s'-s;\nu;0) \\ \mathcal{J}^{(s''\epsilon'')}_{ba}(s'-s,-s'-s;\nu;3) \end{pmatrix},$$
$$a = b\eta_{\epsilon'}. \tag{4.8}$$

The value of ϵ'' is determined in the natural way by ϵ and ϵ'. p'' and a'' are the eigenvalues of J_2 and $\mathcal{A}$, respectively. The two values $b=\pm$ distinguish the two occurrences of $C_{q''}^{\epsilon''}$ in $C_q^{\epsilon}\otimes C_{q'}^{\epsilon'}$.

Vectors of type (4.7a) are orthogonal to those of types (4.7b) and (4.8), and so are the latter two. Among the vectors of each type the factors have been chosen so as to have

$$\left(\Psi^{(s_1\epsilon_1)(s_1'\epsilon_1')(k_1\pm)}_{p_1''},\ \Psi^{(s\epsilon)(s'\epsilon')(k\pm)}_{p''}\right) = \delta(s_1-s)\,\delta(s_1'-s') \times \delta_{\epsilon_1\epsilon}\,\delta_{\epsilon_1'\epsilon'}\,\delta_{k_1 k}\,\delta(p_1''-p''), \tag{4.9a}$$

$$\left(\Psi^{(s_1\epsilon_1)(s_1'\epsilon_1')(s_1''\epsilon_1'')b_1}_{p_1''a_1''},\ \Psi^{(s\epsilon)(s'\epsilon')(s''\epsilon'')b}_{p''a''}\right) \tag{4.9b}$$
$$= \delta(s_1-s)\,\delta(s_1'-s')\,\delta_{\epsilon_1\epsilon}\,\delta_{\epsilon_1'\epsilon'}\ \delta(s_1''-s'')\,\delta_{\epsilon_1''\epsilon''}\,\delta_{b_1 b}\,\delta(p_1''-p'') \times \delta_{a_1''a''}.$$

From the existence, orthogonality and completeness of the coupled basis vectors Ψ there follows the structure of the C–G series

$$C_q^{\epsilon}\otimes C_{q'}^{\epsilon'} = \sum_{k\geq 1 \text{ or } 3/2}^{\infty} D_k^+ \oplus \sum_{k\geq 1\text{ or } 3/2}^{\infty} D_k^- \oplus 2\int_{1/4}^{\infty} dq''\, C_{q''}^{\epsilon''}. \tag{4.10}$$

It is interesting to point out the following two features: (i) the UIR's $D_{1/2}^{\pm}$ do not appear in the C–G series here *because* they do not appear in the Plancherel theorem for $SU(1,1)$; this was also the reason for their absence in the reduction of $D_k^+\otimes D_{k'}^-$; (ii) the appearance of $C_{q''}^{\epsilon''}$ with multiplicity two is here related to the fact that within the continuous class UIR's of $SU(1,1)$ each eigenvalue of J_2 also appears twice, leading to the necessity of using extra labels such as b, a.

5. C–G COEFFICIENTS IN A CONTINUOUS BASIS

It now remains only to compute the three distinct kinds of C–G coefficients, namely,

$$C(s\epsilon\ s'\epsilon'\ \mathcal{R} b\,|\,pa\ p'a'\ p''a'') = \delta(p''-p-p')\,\hat{C}(s\epsilon\ s'\epsilon'\ \mathcal{R} b\,|\,pa\ p'a'\ a''),$$

for $\mathcal{R}=(k,+)$, $(k,-)$, and (s'',ϵ''). Here $b=\pm$ is a multiplicity index labelling the double appearance of the UIR (s'',ϵ'') in the reduction of $C^{\epsilon}{}_{(1/4+s^2)}\otimes C^{\epsilon'}{}_{(1/4+s'^2)}$. It is to be omitted if $\mathcal{R}=(k,+)$ or $(k,-)$.

From Eqs. (4.5) and (4.7a) we find

$$\hat{C}(s\epsilon\ s'\epsilon'\ k+|pa\ p'a') = (\pi^{3/2}/4)\,\mu(k)\Big(aa'\int_{-\pi}^{0}(-\sin\mu)d\mu \times(\cos\mu/2)^{-2ip-1}(-\sin\mu/2)^{-2ip'-1}[\mathcal{G}^{(k+)}(s'-s,-s'-s;\mu) + \eta_{\epsilon'}\mathcal{G}^{(k-)}(s'-s,-s'-s;\mu)]$$
$$+ a\int_0^{\infty}(\sinh\nu)\,d\nu\,(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}[\mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;0) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;0)]$$
$$+ a'\int_0^{\infty}(\sinh\nu)d\nu\,(\sinh\nu/2)^{-2ip-1}(\cosh\nu/2)^{-2ip'-1} \times[\mathcal{J}^{(k+)}(s'-s,-s'-s;\nu;3) + \eta_{\epsilon'}\mathcal{J}^{(k-)}(s'-s,-s'-s;\nu;3)]\Big). \tag{5.1}$$

The trivial integrations over ζ, ζ' and r have been performed and the last of these integrations yields the factor $\delta(p''-p-p')$, dropping which we arrive at $\hat{C}$. In order to carry out the remaining integrations over μ and ν we need the relevant expressions for $\mathcal{G}^{(k\pm)}$ and $\mathcal{J}^{(k\pm)}$ which we quote from Ref. 4.

$$\mathcal{G}^{(k\pm)}(s'-s,-s'-s;\mu) = (2\pi)^{-1}\exp(\mp s'\pi)\,\exp\{-i[\eta_k(s'-s)-\eta_k(-s'-s)]\} \times\Gamma(-2is')\,\psi_1(k;s'-s,-s'-s;\mu) + (2\pi)^{-1}\exp(\pm s'\pi) \times\exp\{i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\Gamma(2is')\,\psi_2(k;s'-s,-s'-s;\mu),$$
$$\text{for } -\pi\leq\mu\leq 0, \tag{5.2}$$

where

$$\psi_1(k;s'-s,-s'-s;\mu) = (\cos^2\mu/2)^{-is}(\sin^2\mu/2)^{is'}\,{}_2F_1(k-is+is',\ 1-k-is+is';\ 1+2is';\ \sin^2\mu/2),$$

$$\psi_2(k;s'-s,-s'-s;\mu)=\psi_1(k;-s'-s,s'-s;\mu), \qquad (5.3)$$

and $\eta_k(x)\equiv \arg\Gamma(k-ix)$,

$$\mathcal{J}^{(k\pm)}(s'-s,-s'-s;\nu;0)=(2\pi)^{-1}\exp(\mp s'\pi)\ \exp\{-i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\ \Gamma(-2is')$$
$$\times\phi_1(k;s'-s,-s'-s;\nu)+\exp\{i[\eta_k(s'-s)-\eta_k(-s'-s)]\}$$
$$\Gamma(2is')\,\phi_2(k;s'-s,-s'-s;\nu)\ ,\quad \text{for } 0\leq\nu<\infty, \qquad (5.4)$$

where

$$\phi_1(k;s'-s,-s'-s;\nu)=(\cosh^2\nu/2)^{-is}(\sinh^2\nu/2)^{is'}\ {}_2F_1(k-is+is',\ 1-k-is+is';\ 1+2is';\ -\sinh^2\nu/2),$$
$$\phi_2(k;s'-s,-s'-s;\nu)=\phi_1(k;-s'-s,\ s'-s;\nu), \qquad (5.5)$$

$$\mathcal{J}^{(k\pm)}(s'-s,-s'-s;\nu;3)=(-1)^{2k}\exp(\pm i\pi k)\,\mathcal{J}^{(k\pm)}(s'-s,s'+s;\nu;0). \qquad (5.6)$$

From (5.1)—(5.6) we see that we have essentially the following integrals to evaluate:

$$L(s,s';p,p';k)\equiv\int_{-\pi}^{0}(-\sin\mu)\,d\mu\,(\cos\mu/2)^{-2ip-1}(-\sin\mu/2)^{-2ip'-1}\times\psi_1(k;s'-s,-s'-s;\mu), \qquad (5.7a)$$

and

$$M(s,s';p,p';k)\equiv\int_0^{\infty}(\sinh\nu)\,d\nu\,(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}\times\phi_1(k;s'-s,-s'-s;\nu). \qquad (5.7b)$$

These integrations can be carried out using the method employed in I, II, and III. We omit the details and simply quote the results:[5]

$$L(s,s';p,p';k)=2\,\frac{\Gamma(\frac12-is-ip)\,\Gamma(\frac12+is'-ip')}{\Gamma(1-is+is'-ip-ip')}\times{}_3F_2\left(\begin{matrix}k-is+is',\ 1-k-is+is',\ \frac12+is'-ip'\\ 1+2is',\ 1-is+is'-ip-ip'\end{matrix};1\right), \qquad (5.8a)$$

$$M(s,s';p,p';k)=2\,\frac{\Gamma(k+ip+ip')\,\Gamma(\frac12+is'-ip')}{\Gamma(k+\frac12+is'+ip)}\times{}_3F_2\left(\begin{matrix}k-is+is',\ k+is+is',\ \frac12+is'-ip'\\ 1+2is',\ k+\frac12+is'+ip\end{matrix};1\right). \qquad (5.8b)$$

It is easy to see that

$$\int_{-\pi}^{0}(-\sin\mu)\,d\mu\,(\cos\mu/2)^{-2ip-1}(-\sin\mu/2)^{-2ip'-1}\psi_2(k;s'-s,-s'-s;\mu)=L(s,-s';p,p';k)$$
$$\int_0^{\infty}(\sinh\nu)\,d\nu\,(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}\times\phi_2(k;s'-s,-s'-s;\nu)=M(s,-s';p,p';k). \qquad (5.9)$$

In the integration involving $\mathcal{J}^{(k\pm)}(s'-s,-s'-s;\nu;3)$, we will also need

$$\int_0^{\infty}(\sinh\nu)\,d\nu\,(\sinh\nu/2)^{-2ip-1}(\cosh\nu/2)^{-2ip'-1}\times\phi_{\substack{1\\2}}(k;s'-s,s'+s;\nu)=M(-s',\mp s;p',p;\nu). \qquad (5.10)$$

Putting all this together we finally obtain

$$\hat{C}(s\epsilon\ s'\epsilon'\ k+|pa\ p'a')=(\pi^{3/2}/4)\,\mu(k)[aa'(G(k+)+\eta_{\epsilon'}G(k-))+a(F(k+;0)+\eta_{\epsilon'}F(k-;0))+a'(F(k+;3)+\eta_{\epsilon'}F(k-;3))],$$

where we have set (5.11)

$$G(k\pm)=(2\pi)^{-1}\exp(\mp s'\pi)\exp\{-i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\Gamma(-2is')L(s,s';p,p';k)$$
$$+(2\pi)^{-1}\exp(\pm s'\pi)\exp\{i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\Gamma(2is')L(s,-s';p,p';k), \qquad (5.12a)$$

$$F(k\pm;0)=(2\pi)^{-1}\exp(\mp s'\pi)\Big[\exp\{-i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\Gamma(-2is')M(s,s';p,p';k)$$
$$+\exp\{i[\eta_k(s'-s)-\eta_k(-s'-s)]\}\Gamma(2is')M(s,-s';p,p';k)\Big], \qquad (5.12b)$$

and

$$F(k\pm;3)=(2\pi)^{-1}(-1)^{2k}\exp[\pm(s+ik)\pi]\Big[\exp\{i[\eta_k(s'+s)-\eta_k(s'-s)]\}\Gamma(2is)$$
$$\times M(-s',-s;p',p;k)+\exp\{-i[\eta_k(s'+s)-\eta_k(s'-s)]\}\Gamma(-2is)M(-s',s;p',p;k)\Big]. \qquad (5.12c)$$

By an identical procedure, but using (4.76) instead of (4.7a) we get

$$\hat{C}(s\epsilon\ s'\epsilon'\ k-|pa\ p'a')=(\pi^{3/2}/4)\ \mu(k)[(G(k+)+\eta_{\epsilon'}G(k-))+a'(F(k+;0)+\eta_{\epsilon'}F(k-;0))+a(F(k+;3)+\eta_{\epsilon'}F(k-;3))]. \qquad (5.13)$$

Turning now to the third (and final) case, we have

$$\hat{C}(s\epsilon\ s'\epsilon'\ s''\epsilon''\ b|pa\ p'a'\ a'')=(\pi^{3/2}/4)\,\mu(s''\epsilon'')\Big((1+aa'a'')\times\int_{-\pi}^{0}(-\sin\mu)\,d\mu\,(\cos\mu/2)^{-2ip-1}(-\sin\mu/2)^{-2ip'-1}$$
$$\times\mathcal{G}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\mu)$$
$$+(a'+aa'')\int_0^{\infty}(\sinh\nu)d\nu(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}\times\mathcal{J}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\nu;0)$$
$$+(a+a'a'')\int_0^{\infty}(\sinh\nu)d\nu\,(\sinh\nu/2)^{-2ip-1}(\cosh\nu/2)^{-2ip'-1}\times\mathcal{J}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\nu;3)\Big), \qquad (5.14)$$

where

$$c\equiv b\eta_{\epsilon'}.$$

We again quote the relevant expressions for $\mathcal{G}_{bc}^{(s''\epsilon'')}$ and $\mathcal{J}_{bc}^{(s''\epsilon'')}$ from Ref. 4.

$$\mathcal{G}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\mu)=(2\pi)^{-2}\,\Gamma(\tfrac12+is+is'-is'')\times\Gamma(\tfrac12-is+is'+is'')$$
$$\times\Gamma(-2is')[\cosh(s+s'-s'')\pi+bc\cosh(s-s'-s'')\pi-i\eta_{\epsilon''}c\sinh 2s'\pi]$$
$$\times\psi_1{}^*(s'';-s'-s,s'-s;\mu)+(2\pi)^{-2}\,\Gamma(\tfrac12-is-is'-is'')\times\Gamma(\tfrac12+is-is'+is'')$$
$$\times\Gamma(2is')[\cosh(s-s'+s'')\pi+bc\cosh(s+s'+s'')\pi+i\eta_{\epsilon''}b\sinh 2s'\pi]$$

$\times \psi_2{}^*(s''; -s'-s, s'-s;\mu)$, for $-\pi \leqslant \mu \leqslant 0$. (5.15)

The $\psi_{\frac{1}{2}}{}^*(s''; -s'-s, s'-s;\mu)$ are obtained form $\psi_{\frac{1}{2}}(k; -s'-s, s'-s;\mu)$ defined in Eq. (5.3) by making the replacement $k \to \frac{1}{2}+is''$ and then taking the complex conjugate.

$$\mathcal{J}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\nu;0) = (2\pi)^{-2}\,\Gamma(\tfrac{1}{2}-is+is'+is'')$$
$$\times\,\Gamma(\tfrac{1}{2}+is+is'-is'')$$
$$\times\,\Gamma(-2is')\,[\cosh(s+s'-s'')\pi + bc\,\cosh(s-s'-s'')\pi$$
$$-\,ib\,\sinh 2s'\pi]$$
$$\times\,\phi_1(s'';s'-s,-s'-s;\nu) + (2\pi)^{-2}\,\Gamma(\tfrac{1}{2}+is-is'+is'')$$
$$\times\,\Gamma(\tfrac{1}{2}-is-is'-is'')$$
$$\times\,\Gamma(2is')\,[\cosh(s+s'+s'')\pi + bc\,\cosh(s-s'+s'')\pi$$
$$+\,i\eta_{\epsilon''}\,b\,\sinh 2s'\pi]$$
$$\times\,\phi_2(s'';s'-s,-s'-s;\nu), \quad \text{for } 0 \leqslant \nu < \infty. \tag{5.16}$$

Here again $\phi_{\frac{1}{2}}(s'';s'-s,-s'-s;\nu)$ are obatined by making the replacement $k \to \frac{1}{2}+is''$ in Eq. (5.5):

$$\mathcal{J}_{bc}^{(s''\epsilon'')}(s'-s,-s'-s;\nu;3) = \eta_{\epsilon''}\,c\,\mathcal{J}_{b,\eta_{\epsilon''}c}^{(s''\epsilon'')}(s'-s,s'+s;\nu;0). \tag{5.17}$$

Referring now to Eq. (5.14) we see that we again need the analogs of L and M defined in Eqs. (5.7a) and (5.7b). We define these as follows:

$$L(s,s';p,p';s'') = \int_{-\pi}^{0}(-\sin\mu)d\mu(\cos\mu/2)^{-2ip-1}(-\sin\mu/2)^{-2ip'-1}$$
$$\times\,\psi_1{}^*(s'';-s'-s,s'-s;\mu), \tag{5.18a}$$

$$M(s,s';p,p';s'') = \int_0^\infty \sinh\nu\,d\nu\,(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}$$
$$\times\,\phi_1(s'';s'-s,-s'-s;\nu). \tag{5.18b}$$

The evaluation of these integrals proceeds along the same lines as before and we get

$$L(s,s';p,p';s'') = 2\,\frac{\Gamma(\frac{1}{2}+is'-ip')\,\Gamma(\frac{1}{2}+is-ip)}{\Gamma(1+is+is'-ip-ip')}$$
$$\times\,{}_3F_2\left(\begin{matrix}\frac{1}{2}+is+is'-is'',\ \frac{1}{2}+is+is'+is'',\ \frac{1}{2}+is'-ip' \\ 1+2is',\ 1+is+is'-ip-ip'\end{matrix};1\right), \tag{5.19a}$$

$$M(s,s';p,p';s'') = 2\,\frac{\Gamma(\frac{1}{2}+is''+ip+ip')\,\Gamma(\frac{1}{2}+is'-ip')}{\Gamma(1+is'+is''+ip)}$$
$$\times\,{}_3F_2\left(\begin{matrix}\frac{1}{2}-is+is'+is'',\ \frac{1}{2}+is+is'+is'',\ \frac{1}{2}+is'-ip' \\ 1+2is',\ 1+is'+is''+ip\end{matrix};1\right). \tag{5.19b}$$

And we also have

$$\int_{-\pi}^{0}(-\sin\mu)\,d\mu\,(\cos\mu/2)^{-2ip-1}(\sin\mu/2)^{-2ip'-1}\,\psi_2{}^*(s'';s'-s,-s'-s;\mu)$$
$$= L(s,-s';p,p';s''), \tag{5.20a}$$

and

$$\int_0^\infty(\sinh\nu)d\nu(\cosh\nu/2)^{-2ip-1}(\sinh\nu/2)^{-2ip'-1}$$
$$\phi_2(s'';s'-s,-s'-s;\nu)$$
$$= M(s,-s';p,p';s'').$$

In evaluating the third integral in (5.14) we also need

$$\int_0^\infty(\sinh\nu)d\nu(\sinh\nu/2)^{-2ip-1}(\cosh\nu/2)^{-2ip'-1}$$
$$\phi_{\frac{1}{2}}(s'';s'-s,s'+s;\nu)$$
$$= M(-s',\mp s;p',p;s''). \tag{5.21}$$

We can now write down the C–G coefficient:

$$\hat{C}(s\epsilon\ s'\epsilon'\ s''\epsilon''b\,|\,pa\ p'a'\ a'') = (\pi^{3/2}/4)\,\mu(s''\epsilon'')$$
$$\times[(1+aa'a'')\,G_{bc}(s''\epsilon'') + (a'+aa'')\,F_{bc}(s''\epsilon'';0)$$
$$+(a+a'a'')\,F_{bc}(s''\epsilon'';3)], \tag{5.22}$$

where

$$G_{bc}(s''\epsilon'') = (2\pi)^{-2}\,\Gamma(\tfrac{1}{2}+is+is'-is'')\,\Gamma(\tfrac{1}{2}-is+is'+is'')$$
$$\times\,\Gamma(-2is')$$
$$\times[\cosh(s+s'-s'')\pi + bc\,\cosh(s-s'-s'')\pi$$
$$-\,i\eta_{\epsilon''}\,c\,\sinh 2s'\pi]\,L(s,s';p,p';s'')$$
$$+(2\pi)^{-2}\,\Gamma(\tfrac{1}{2}-is-is'-is'')\ \Gamma(\tfrac{1}{2}+is-is'+is'')\,\Gamma(2is')$$
$$\times[\cosh(s-s'+s'')\pi + bc\,\cosh(s+s'+s'')\pi$$
$$-\,i\eta_{\epsilon''}b\,\sinh 2s'\pi]\,L(s,-s';p,p';s''), \tag{5.23a}$$

$$F_{bc}(s''\epsilon'';0) = (2\pi)^{-2}\,\Gamma(\tfrac{1}{2}-is+is'+is'')\,\Gamma(\tfrac{1}{2}+is+is'-is'')$$
$$\times\,\Gamma(-2is')$$
$$\times[\cosh(s+s'-s'')\pi + bc\,\cosh(s-s'-s'')\pi - ib\,\sinh 2s'\pi]$$
$$\times\,M(s,s';p,p';s'')$$
$$+(2\pi)^{-2}\,\Gamma(\tfrac{1}{2}+is-is'+is'')\,\Gamma(\tfrac{1}{2}-is-is'-is'')\,\Gamma(2is')$$
$$\times[\cosh(s+s'+s'')\pi + bc\,\cosh(s-s'+s'')\pi + i\eta_{\epsilon''}b\sinh 2s'\pi]$$
$$\times\,M(s,-s';p,p';s''), \tag{5.23b}$$

and

$$F_{bc}(s''\epsilon'';3)$$
$$= \eta_{\epsilon''}\,c\,\{(2\pi)^{-2}\,\Gamma(\tfrac{1}{2}-is+is'+is'')\,\Gamma(\tfrac{1}{2}-is-is'-is'')\,\Gamma(2is)$$
$$\times[\cosh(s+s'+s'')\pi + \eta_{\epsilon''}\ bc\,\cosh(s-s'-s'')\pi + ib\,\sinh 2s\pi]$$
$$\times\,M(-s',-s;p',p;s'')$$
$$+(2\pi)^{-2}\,\Gamma(\tfrac{1}{2}+is-is'+is'')\,\Gamma(\tfrac{1}{2}+is+is'-is'')\,\Gamma(-2is)$$
$$+\eta_{\epsilon''}bc\,\cosh(s-s'+s'')\pi - i\eta_{\epsilon''}b\,\sinh 2s\pi]$$
$$\times\,M(-s',s;p',p;s''). \tag{5.23c}$$

This completes the evaluation of the C–G coefficients in the continuous basis.

SUMMARY

We first make a few comments on the present paper, and then on the entire series of which this forms the concluding part. Following the approach of the previous papers, we have established a connection between the Clebsch–Gordan problem of $SU(1,1)$ for products of the form $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ and the structure of spherical harmonics for the group $O(2,2)$ in an $O(1,1)\otimes O(1,1)$ basis. By exploiting this connection, we have explained the form of the C–G series for this case in a new way, and have also obtained the C–G coefficients in an $O(1,1)$ basis. As in the previous cases, these coefficients are expressible in terms of the generalized hypergeometric function ${}_3F_2$ with unit argument. However, the actual expression for the C–G coefficient corresponding to the case $C_q^\epsilon \otimes C_{q'}^{\epsilon'} \to C_{q''}^{\epsilon''}$, say, is rather lengthy and involves several terms, in comparison with the cases $D^\pm \otimes D^\pm \to D^\pm$ for example. As in the case of the products $D_k^+ \otimes D_{k'}^-$ treated in II, the reason why the representations $D_{1/2}^\pm$ never make an appearance in the reduction of $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ is understood satisfactorily; it is directly related to their absence in the Plancherel theorem for $SU(1,1)$. But the absence of $D_{1/2}^\pm$ in the products $D_k^\pm \otimes C_q^\epsilon$ had a somewhat different explanation, being related to the structure of spherical harmonics with respect to the group $O(3,1)$. This is explained in III. We have also explained the double occurrence of each $C_{q''}^{\epsilon''}$ in $C_q^\epsilon \otimes C_{q'}^{\epsilon'}$ (for appropriate choice of ϵ'') in a new way: It happens because within a continuous series UIR of $SU(1,1)$ each eigenvalue of a noncompact $O(1,1)$ generator appears twice.

The calculations of the C–G coefficients that we have performed in the four different cases of products, and the expressions that we have given for them, are mutually consistent in the following sense. At the beginning of the investigation we set up standard forms for each of the UIR's of $SU(1,1)$ that were of interest, namely $D_k^\pm$ and C_q^ϵ for $q \geq \frac{1}{4}$. And we made sure that any such UIR whether present as a factor in a direct product $\mathcal{R}\otimes\mathcal{R}'$ or as a summand in the direct sum decomposition of a product was always, exhibited in the standard form. Further, the choice of the $O(1,1)$-basis vectors within each UIR of $SU(1,1)$ was specified completely, with no uncertain phase factors, in I, and this choice was adhered to throughout in setting up the uncoupled and coupled basis vectors for each product $\mathcal{R}\otimes\mathcal{R}'$. However, the question arises as to the form of the $SU(1,1)$ representation matrices in the $O(1,1)$ basis that must be taken with our expressions for the C–G coefficients. Let us write $\tilde{D}^{(\mathcal{R})}_{p'b,pa}(g)$ for these matrices. They are such that they obey, along with the C–G coefficients that we have calculated, the following equation:

$$\tilde{D}^{(\mathcal{R})}_{p'c,pa}(g)\,\tilde{D}^{(\mathcal{R}')}_{p'''d,p'b}(g)=\int d\mathcal{R}''\sum_{\gamma ef}\int_{-\infty}^{\infty}dp_1\int_{-\infty}^{\infty}dp_1'$$
$$\times\,\Delta(\mathcal{R},\mathcal{R}';\mathcal{R}'')$$
$$\times C(\mathcal{R}\;\mathcal{R}'\;\mathcal{R}''\;\gamma\,|\,p''c\;p'''d\;p_1'f)\;\tilde{D}^{(\mathcal{R}'')}_{p_1'f,p_1e}(g) \qquad (6.1)$$
$$\times C(\mathcal{R}\;\mathcal{R}'\;\mathcal{R}''\;\gamma\,|\,pa\;p'b\;p_1e)^*.$$

The form of this equation follows from the way the various states have been normalized, as set forth in Sec. 4 of I. The precise definition of the measure $d\mathcal{R}''$ is given in Sec. 3 of II; and the function $\Delta(\mathcal{R},\mathcal{R}';\mathcal{R}'')$ is unity if $\mathcal{R}''$ occurs in the decomposition of $\mathcal{R}\otimes\mathcal{R}'$ and vanishes otherwise. So for given $\mathcal{R}$ and $\mathcal{R}'$, the range of the $\mathcal{R}''$ integration is determined by the appropriate C–G series listed in Sec. 2 of I. Now the representation matrices $\tilde{D}^{(\mathcal{R})}(g)$ are of course completely determined by the standard forms that we have set up in Sec. 1 of I for the various UIR's. On the other hand, in Ref. 4, a calculation of the representation matrices in an $O(1,1)$ basis has been carried out for all the UIR's of $SU(1,1)$, using a different description of the UIR's. These matrices, $D^{(\mathcal{R})}_{p'b,pa}(g)$, are just the ones we have used in Sec. 3 in setting up the $O(2,2)$ spherical harmonics in an $O(1,1)\otimes O(1,1)$ basis. Equation (6.1) will not be valid if we were to replace $\tilde{D}$ by D everywhere. Instead of calculating $\tilde{D}^{(\mathcal{R})}(g)$, it is enough to relate them to $D^{(\mathcal{R})}(g)$; this relation must necessarily be of the form

$$\tilde{D}^{(\mathcal{R})}_{p'b,pa}(g)=\exp[i\varphi(\mathcal{R};p',b)-i\varphi(\mathcal{R};p,a)]\,D^{(\mathcal{R})}_{p'b,pa}(g), \qquad (6.2)$$

with $\varphi(\mathcal{R};p,a)$ a real quantity. By making $\varphi(\mathcal{R};p,a) = \varphi(\tau(\mathcal{R});p,a)$, we secure the property

$$\tilde{D}^{(\mathcal{R})}_{p'b,pa}(\tau(g))=ba\,\tilde{D}^{(\tau(\mathcal{R}))}_{p'b,pa}(g) \qquad (6.3)$$

for $\tilde{D}$, analogous to Eq. (3.3) for D; and then this makes Eq. (6.1) above and Eq. (5.4) of I, describing the effect of τ on the C–G coefficients, mutually consistent. The values of $\varphi(\mathcal{R};p,a)$ turn out to be[6]

$$\varphi(k\pm;p)=-p\ln 2+\arg\,\Gamma(k-ip), \qquad (6.4a)$$

$$\varphi(s\epsilon;p,a)=-p\ln 2-\arg\Gamma(\tfrac{1}{2}+is+ip)$$
$$-\arctan[\eta_\epsilon\,a\exp(-(s+p)\pi)]. \qquad (6.4b)$$

We emphasize once more that a mutually consistent set of $O(1,1)$-basis representation matrices and C–G coefficients, in the sense of the validity of Eq. (6.1), is given by $\tilde{D}^{(\mathcal{R})}(g)$ and the C–G coefficients as calculated by us, and *not* by $D^{(\mathcal{R})}(g)$ and these coefficients.

It is unfortunate that in our analysis of the Clebsch-Gordan problem for $SU(1,1)$ we had to exclude the UIR's of the exceptional interval, C_q^0 for $0<q<\frac{1}{4}$ in the formation of direct products. This was because a simple construction of these UIR's in terms of oscillator operators is not possible, while on the other hand it is just such constructions of the other UIR's that led to the higher symmetries that we have exploited. It would be interesting to extend our analysis to include the exceptional UIR's and to discover corresponding symmetries in the problem.

A useful byproduct of our work has been the construction of complete sets of spherical harmonics in four-dimensional real space with respect to the various groups $O(p,q)$ for $p+q=4$, $p\geq q$. Of course the fact that the $O(3)$ representation matrices $D^j_{mm'}(R)$ form a complete set of spherical functions on the unit sphere in four dimensional Euclidean space is very well known. We have made explicit the analogous connection between the $SU(1,1)$ representation matrices and the $O(2,2)$ "spherical harmonics". In all cases, our constructions keep a maximal commuting subset of the $O(p,q)$ generators diagonal. In the third paper of this series, we had to deal with the group $O(3,1)$, which is the one case that does not simplify to a lower-dimensional group. Here we had to construct spherical functions for both the "timelike" and "spacelike" regions. The former are reasonably straightforward, while the latter are more involved. Our expressions for the $O(3, 1)$ spherical harmonics in the spacelike region, in an $O(2)\otimes O(1, 1)$ basis, are new and do not exist in the previous literature. These may be of use in various problems involving the $(3+1)$ Lorentz group. The $O(3, 1)$ spherical harmonics for the same region but in a basis in which the $O(3)$ subgroup of $O(3, 1)$ is "diagonal," have been presented in the literature.[7]

The concept and use of the generating representations $\mathcal{D}^{\pm}$, $\mathcal{C}$ of $SU(1, 1)$ seems to us to be quite novel and in principle capable of extension to other groups. It gives a convenient and elegant way of dealing with a large number of UIR's of a chosen noncompact group, G_1 say, in a unified manner. The decomposition of a generating representation of G_1 into UIR's of G_1 would be accomplished by looking for a sufficiently large group of symmetries of this representation. And if the direct product of two generating unitary representations of G_1 has a symmetry group larger than the direct product of the individual symmetry groups, then from the representation structure of the symmetry groups we can learn something about the C–G series for G_1. For groups of larger dimension than $SU(1, 1)$, the analog of the construction of spherical harmonics for appropriate symmetry groups may be quite involved, and not practical. Nevertheless, this general method can be useful in that it may explain the structure of certain C–G series for G_1, though the calculation of the C–G coefficients may be much harder. In the present work we have been particularly lucky since the problems of dealing with the symmetry groups $O(4)$ and $O(2, 2)$ were greatly simplified by the relationships of $O(4) \approx O(3)\otimes O(3)$ and $O(2, 2)\approx O(2, 1)\otimes O(2, 1)$. We hope to examine these questions elsewhere.

ACKNOWLEDGMENT

One of the authors (N. M.) expresses his thanks to the Center for Particle Theory, University of Texas, Austin, Texas, and its Director Professor E. C. G. Sudarshan, for making possible a visit during which his paper was partly written.

*Present address: Centre for Theoretical Studies, Indian Institute of Science, Bangalore-12, India.

[1]N. Mukunda and B. Radhakrishnan, J. Math. Phys. 15, 1320, 1332, 1643 (1974). These papers will be referred to as (I), (II), and (III), respectively.

[2]We prefer to list f_- first and f_+ second so that the UIR's $C_q^{\epsilon''}$ found in the reduction of $C_q^{\epsilon} \bigcirc C_{q'}^{\epsilon'}$ are obtained in the standard form of (I).

[3]N. Mukunda, J. Math. Phys. 8, 2210 (1967).

[4](a) N. Mukunda, J. Math. Phys. 11, 2086, 2092 (1969); (b) N. Mukunda, "Matrices of Finite Lorentz Transformations in a Concompact Basis-III. Completeness Relation for $O(2,1)$," J. Math. Phys. (to be published).

[5]In evaluating these integrals use has been made of I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 849, formual 7.512.5.

[6]The details of this calculation may be found in B. Radhakrishnan; Ph. D. thesis (under preparation).

[7]A. Bassetto and M. Toller, Ann. Inst. Henri Poincaré A 15, 1 (1973).

Multiregion criticality in general geometries

O. J. Sheaks

Department of Chemical Engineering, University of Maryland, College Park, Maryland 20742

T. W. Schnatz

Middle South Services, New Orleans, Louisiana

(Received 12 February 1973)

A general transform technique is developed for multiregion critical problems. The equivalence of a replication procedure and a derived boundary condition approach is demonstrated for the general multiregion geometries. An exact representation for the particle density may be obtained using this approach in the form of singular integral equations or equivalent Fredholm equations for expansion coefficients which arise from the superposition of the normal modes representing the particle density. The method is specifically demonstrated in the determination of solutions to the two-region critical cylinder problem.

I. INTRODUCTION

Since the application by Case[1] of the singular eigenfunction expansion technique to solving the Boltzmann transport equation for neutron distributions in plane, homogeneous, isotropically scattering media, extensive investigations have been made to include more realistic assumptions of anisotropic scattering, time-dependence, multigroup, and energy dependence for distributions of neutrons and photons. Several investigations have been made to extend the method to nonplanar geometries using various techniques applicable to isolated cases.[2,3] One procedure, referred to as the transform approach, has evolved which can be employed in a general manner to a variety of transport problems, greatly extending the class of problems solvable using the other techniques.

Leonard and Mullikan[4] were the first to suggest the idea in an application to neutron transport in spheres. Mitsis[5] extended the transform concept to include critical problems in single region, one-dimensional slabs, infinite cylinders and spheres. In a classical mathematical approach Gibbs[6] demonstrated general applicability in three-dimensional, arbitrary convex bodies consisting of a single homogeneous material, where arbitrary source distributions were permitted.

The techniques employed by Mitsis and Gibbs were similar in philosophy. Basically each consisted of reducing the integral form of the transport equations to an equation for a transform function which could be solved using Case's method. The particle density could then be represented by a simple integral of the transform function. The two methods differed radically, however, in the method employed to derive equations which must be satisfied by the expansion coefficients for the normal modes comprising the transform function. Mitsis's procedure was somethat formal in that "boundary conditions" were derived from the definition of the transform function, and no mathematical evidence was given to assure validity of the transform procedure. On the other hand, Gibb's method made use of the replication properties of the transform eigensolutions and assured a consistent mathematical foundation for his transform procedure. To demonstrate the equivalence of the two methods Gibbs showed his singular integral equations for the expansion coefficients were identical to those obtained by Mitsis with his boundary condition approach. Thus, for single-region problems, the mathematical rigor of the more easily applied Mitsis procedure could be inferred from the work of Gibbs.

The extension of the transform concept to bodies containing regions of differing multiplication has been accomplished by Smith and Siewert[7] and Sheaks.[8] The former paper determined solutions in two-region spheres; the latter presented solutions for the particle density in an annular region surrounding a central black cavity. In each case a procedure analogous to that of Mitsis was employed. However, the technique lacked generality, being dependent on the specific problem analyzed. Also, the mathematical rigor of the method was not proven since the analogy to Gibb's single-region analysis was no longer applicable.

The purpose of this paper is two-fold: we demonstrate that the boundary condition method and the replication method are in general equivalent for multiregion problems; in addition, we extend the transform technique to multiregion cylinders by determining solutions for the two-region critical cylinder problem.

II. MULTIREGION TRANSFORM TECHNIQUE

We consider the general form of the equation[9] describing particle transport in a convex region, V,

$$n(\mathbf{r}) = \int_V d\mathbf{r}'[c(\mathbf{r}')n(\mathbf{r}') + S(\mathbf{r}')]K(|\mathbf{r}-\mathbf{r}|), \quad \mathbf{r}\in V, \tag{2.1}$$

where $n(\mathbf{r})$ is the particle density, $c(\mathbf{r})$ is the mean number of secondaries per collision, $S(\mathbf{r})$ includes contributions from a flux incident on V and from distributed sources within V, and distance is measured in units of the total mean free path. We assume isotropic scattering and an invariant total mean free path throughout V, and that V can be subdivided into N subregions, V_i, where $c(\mathbf{r})$ has constant values, c_i.

The transfer kernel under these assumptions becomes

$$K(|\mathbf{r}|) = e^{-|\mathbf{r}|}/4\pi|\mathbf{r}|^2. \tag{2.2}$$

Noting that we can write

$$K(|\mathbf{r}|) = \int_0^1 \frac{d\mu}{\mu^2}\frac{e^{-|\mathbf{r}|/\mu}}{4\pi|\mathbf{r}|} = \int_0^1 \frac{d\mu}{\mu^2} G(|\mathbf{r}|, \mu) \tag{2.3}$$

and that $G(\mathbf{r}, \mu)$ satisfies

$$(-\nabla^2 + 1/\mu^2)G(\mathbf{r}-\mathbf{r}', \mu) = \delta(\mathbf{r}-\mathbf{r}'), \tag{2.4}$$

Copyright © 1974 American Institute of Physics

we proceed in the manner of Gibbs to write

$$n(\mathbf{r}) = \int_0^1 \frac{d\mu}{\mu^2} F(\mathbf{r}, \mu), \quad \mathbf{r} \in V, \tag{2.5}$$

where

$$F(\mathbf{r}, \mu) = \int_V d\mathbf{r}'[c(\mathbf{r}')n(\mathbf{r}') + S(\mathbf{r}')]G(|\mathbf{r}-\mathbf{r}'|, \mu). \tag{2.6}$$

Equations (2.5) and (2.6) are thus considered a transform pair, and $F(\mathbf{r}, \mu)$ can be shown by substitution to satisfy

$$\left(-\nabla^2 + \frac{1}{\mu^2}\right) F(\mathbf{r}, \mu) - c(\mathbf{r}) \int_0^1 \frac{d\mu}{\mu^2} F(\mathbf{r}, \mu) = S(\mathbf{r}). \tag{2.7}$$

The basis of the transform procedure consists of constructing solutions to Eq. (2.7) such that Eqs. (2.5) and (2.6) are mutually consistent. We note that while the single-region transform procedure has been followed to this point, Eq. (2.7) is not separable over $\mathbf{r}$ and μ and a new technique must now be employed.

We continue with a procedure followed extensively in multiregion reactor physics analysis, i.e., in lieu of determining the general eigensolutions of Eq. (2.7), we find solutions for each subregion, separately. Thus we consider N-subregions for which the transform function, $F_i(\mathbf{r}, \mu)$, defined in that region satisfies

$$\left(-\nabla^2 + \frac{1}{\mu^2}\right) F_i(\mathbf{r}, \mu) + c_i \int_0^1 \frac{d\mu}{\mu^2} F_i(\mathbf{r}, \mu) = S(\mathbf{r}), \qquad i = 1, N. \tag{2.8}$$

The homogeneous solutions of Eq. (2.8) have been presented in detail by Mitsis and Gibbs are presented here for completeness of presentation. Assuming separation of variables we find

$$F_i(\mathbf{r}, \mu) = \int d\nu f_i(\nu, \mu) R_i(\mathbf{r}, \nu) \tag{2.9}$$

where $R_i(\mathbf{r}, \nu)$ is a solution of

$$(-\nabla^2 + 1/\nu^2) R_i(\mathbf{r}, \nu) = 0, \quad \mathbf{r} \in V_i, \tag{2.10}$$

$$f_i(\nu, \mu) = c_i \frac{P\nu^2 \mu^2}{\nu^2 - \mu^2} + \nu^2 \lambda_i(\nu)[\delta(\nu - \mu) + \delta(\nu + \mu)] \nu \in [0,1], \tag{2.11}$$

$$\lambda_i(\nu) = 1 - c_i \nu \tanh^{-1} \nu, \tag{2.12}$$

$$f_i(\nu_{0i}, \mu) = \frac{c_i \nu_{0i}^2 \mu^2}{\nu_{0i}^2 - \mu^2}, \tag{2.13}$$

in which $\delta(x)$ denotes the Dirac delta-function, P indicates a Cauchy principal value integration, and ν_{0i} is the positive root of the dispersion function

$$\Lambda(z) = 1 - c_i z \tanh^{-1}(1/z). \tag{2.14}$$

The function $R_i(\mathbf{r}, \nu)$ can be constructed from a linear superposition of an appropriate basis set of the null space of the operator $(-\nabla^2 + 1/\nu^2)$. Choosing a basis sufficient for the compatibility of Eqs. (2.5) and (2.6), we can write

$$F_i(\mathbf{r}, \mu) = \int d\nu f_i(\nu, \mu) \sum_n A_i^n(\nu) R_i^n(\mathbf{r}, \nu), \tag{2.15}$$

where the integral sign is used mnemonically to include the discrete eigenvalue and the values of $\nu \in \mathrm{Re}[0,1]$.

For regions containing distributed sources or when particles are incident on the external boundary, particular solutions to Eq. (2.8) may be constructed by a variety of techniques. In a general manner, the classic Green's function is readily applied thus representing the particular solution in terms of the homogeneous functions derived above. In this paper, however, we consider only multiregion critical problems and thus simplify the notation considerably.

III. EQUIVALENCE OF TRANSFORMS

Having determined a general solution for the transform function $F(\mathbf{r}, \mu)$, we now have a choice of two techniques to derive equations which must be satisfied by the expansion coefficients, $A_i^n(\nu)$. The extension of the Mitsis technique would consist of deriving pseudo-boundary conditions directly from the definition of the transform function, Eq. (2.6). Using the Gibbs analogy, Eq. (2.15) is substituted into Eq. (2.5), the resulting expression for the density is inserted into Eq. (2.6), and the replication properties of the transform functions are employed to obtain necessary conditions on $A_i^n(\nu)$ to cause agreement between the resulting expression for $F(\mathbf{r}, \mu)$ and that of Eq. (2.15).

Since the boundary condition approach is more easily facilitated, it is useful to establish the equivalence of this method to the more fundamental and rigorous replication method. We begin with the replication approach to a one-dimensional multiregion system with N homogeneous subregions with dependent variable r, where planar, cylindrical, and spherical geometries are included. For simplicity of presentation we consider only critical problems, thus eliminating particles incident on the outside boundary and distributed external sources.

Because of the above assumptions a sufficient basis set for the spatial functions, $\{R_i^n(\nu, \mathbf{r})\}$, consist of a single pair of linearly independent functions. For simplicity of notation the following analysis is presented for only one of the functions without loss of generality. Thus we write

$$F_i(r, \mu) = \int d\nu f_i(\nu, \mu) A_i(\nu) R_i(\nu, r) \tag{3.1}$$

and

$$n_i(r) = \int d\nu A_i(\nu) R_i(\nu, r), \tag{3.2}$$

since

$$\int_0^1 d\mu f_i(\nu, \mu) = 1.$$

From the definition of the transform function Eq. (2.6),

$$F_i(r, \mu) = \sum_{j=1}^N \int_{V_j} dr' c_j n_j(r') G(|\mathbf{r}-\mathbf{r}'|, \mu), \quad \mathbf{r} \in V_i. \tag{3.3}$$

Substituting Eq. (3.2) into Eq. (3.3) we find the following integral which can be evaluated analytically:

$$Y_{ij} \equiv \int_{V_j} d\mathbf{r}' G(|\mathbf{r}-\mathbf{r}'|, \mu) R_j(\nu, r'), \quad r \in V_i. \tag{3.4}$$

The evaluation procedure is classic: Eq. (2.4) is multiplied by $R_j(\nu, r')$, Eq. (2.10) by $G(|\mathbf{r}-\mathbf{r}'|, \mu)$, the resulting equations integrated over V_j and subtracted. The result is

$$Y_{i,j}=f_j(\nu,\mu)/c_j(R_j(\nu,r)\delta_{i,j}+\int_{V_j}d\mathbf{r}'\nabla'\cdot[R_j(\nu,r')\nabla G \times(|\mathbf{r}-\mathbf{r}'|,\mu)-G(|\mathbf{r}-\mathbf{r}'|,\mu)\nabla R_j(\nu,r')]) \tag{3.5}$$

where δ is the Kronicker delta.

The second term in brackets may be evaluated by Gauss's theorem and a representation of $G(|\mathbf{r}-\mathbf{r}'|,\mu)$ of the form[10]

$$G(|\mathbf{r}-\mathbf{r}'|,\mu)=\frac{g\pm(r',\mu)g\mp(r,\mu)}{W[g_+(r',\mu)g_-(r',\mu)]},\quad r\gtrless r', \tag{3.6}$$

where $g_-(r,\mu)$ and $g_+(r,\mu)$ denote the Sturm–Liouville solutions regular at the origin (or at $-\infty$ for planar geometry) and at $+\infty$, respectively, and W denotes the wronskian.

We find

$$Y_{ij}=\frac{f_j(\nu,\mu)}{c_j}\{R_j(\nu,r)\delta_{i,j}+[I^{\pm}_{j,j}(\nu,\mu)g_{\pm}(r,\mu) - I^{\pm}_{j,j-1}(\nu,\mu)g(\pm r,\mu)]\}, \tag{3.7}$$

where

$$I^{\pm}_{j,k}(\nu,\mu)=S_k\frac{W[R_j(\nu,r_k),g\mp(r_k,\mu)]}{W[g_+(r_k,\mu),g_-(r_k,\mu)]}. \tag{3.8}$$

Here S_{j-1} and S_j designate the surface areas of the inner and outer boundaries, $r=r_{j-1}$, and $r=r_j$, respectively, of the jth subregion; the top sign is applicable for $r>r_j$, the bottom for $r<r_j$. Also, for spheres and cylinders $I^+_{1,0}(\nu,\mu)\equiv 0$.

Inserting Eqs. (3.7) and (3.8) into Eq. (3.3), we find

$$\begin{aligned}F_i(r,\mu)=&\int d\nu A_i(\nu)f_i(\nu,\mu)R_i(\nu,r)\\&-\Big(\sum_{j=1}^{i-1}\int d\nu A_j(\nu)f_j(\nu,\mu)[I^-_{j,j}(\nu,\mu)-I^-_{j,j-1}(\nu,\mu)]g_-(r,\mu)\\&+\sum_{j=i+1}^{N}\int d\nu A_j(\nu)f_j(\nu,\mu)[I^+_{j,j}(\nu,\mu)-I^+_{j,j-1}(\nu,\mu)]g_+(r,\mu)\\&+\int d\nu f_i(\nu,\mu)A_i(\nu)I^+_{i,i}(\nu,\mu)g_+(r,\mu)-\int d\nu f_i(\nu,\mu)\\&\times A_i(\nu)I^-_{i,i-1}(\nu,\mu)g_-(r,\mu)\Big).\end{aligned} \tag{3.9}$$

Thus, by comparing Eq. (3.9) with Eq. (3.1), the bracketed term must necessarily be set to zero. Also, since $g_+(r,\mu)$ and $g_-(r,\mu)$ are linearly independent the coefficients of these functions must independently equal zero. The resulting expressions yield $2N$ equations for the expansion coefficients, $A_i(\nu)$, as i is varied from 1 to N.

The precise equations obtained above using the replication properties can be derived from a boundary condition technique analogous to that of Mitsis. Formally, we can use the definition of the transform function, Eq. (3.3), to derive the following conditions:

$$F_i(r_{i-1},\mu)=F_{i-1}(r_{i-1},\mu),\quad i=2,N, \tag{3.10a}$$

$$\nabla F_i(r_{i-1},\mu)=\nabla F_{i-1}(r_{i-1},\mu),\quad i=2,N, \tag{3.10b}$$

$$F_N(r_N,\mu)\nabla G(r_N-r,\mu)-G(r_N-r,\mu)\nabla F_N(r_N,\mu)=0, \tag{3.10c}$$

where Eq. (3.6) is used to specifically derive Eq. (3.10c). As in the replication method, Eqs. (3.10) represent equations which can be solved for the expansion coefficients.

The equivalence of the two sets of equations can be seen by examining successively the coefficients of $g_+(r,\mu)$ in Eq. (3.9) beginning with $i=N$. We find, first,

$$\int d\nu f_N(\nu,\mu)A_N(\nu)I^+_{N,N}(\nu,\mu)=0. \tag{3.11}$$

By explicitly writing $I^+_N(\nu,\mu)$ from Eq. (3.8) it is easily seen that Eq. (3.10c) is identical with Eq. (3.11).

Next, letting $i=N-1$, we find

$$\int d\nu f_N(\nu,\mu)A_N(\nu)I^+_{N,N}(\nu,\mu)-\int d\nu f_N(\nu,\mu)A_N(\nu)I^+_{N,N-1}(\nu,\mu) + \int d\nu f_{N-1}(\nu,\mu)A_{N-1}(\nu)I^+_{N-1}(\nu,\mu)=0. \tag{3.12}$$

The first term is zero by Eq. (3.11); the equivalent of the remainder of Eq. (3.12) can easily be derived from Eqs. (3.10a) and (3.10b) with $i=N$ by multiplying Eq. (3.10a) by $\nabla g_+(r_{N-1},\mu)$, Eq. (3.10b) by $g_-(r_{N-1},\mu)$, and subtracting the results. If the process is continued with the remaining values of i, the equivalence of the two sets of equations is readily demonstrated. We find each new successive equation will contain only two nonzero terms which can be shown to be equivalent to the boundary conditions Eqs. (3.10a) and (3.10b).

We note that we have used only the replication equations associated with $g_+(r,\mu)$. However, the equations associated with $g_-(r,\mu)$ are easily seen to be equivalent to Eqs. (3.10) using a procedure similar to that described above.

Thus the conditions derived from the transform function definitions are sufficient to determine the expansion coefficients which will satisfy the necessary replication equations. In the next section we apply the formalism presented in this section to determine solutions to the two-region critical cylinder problem.

IV. THE TWO-REGION CRITICAL CYLINDER

We seek solutions for the particle density in a two-region infinite cylinder consisting of a central region, radius R_1, with a multplication constant c_1, surrounded by a concentric outer region, outside radius R_2 with multiplication c_2. The critical problem assumes no external sources or particles incident on the outside boundary. The specific form of Eq. (2.1) applicable to this geometry may be written

$$n_1(r)=\int_0^1\frac{d\mu}{\mu^2}F_1(r,\mu), \tag{4.1}$$

where

$$\begin{aligned}F_1(r,\mu)=&\int_0^r dt\,t\,n_1(t)K_0(r/\mu)I_0(t/\mu)\\&+\int_r^{R_1}dt\,t\,n_1(t)I_0(r/\mu)K_0(t/\mu)\\&+\int_{R_1}^{R_2}dt\,t\,n_2(t)I_0(r/\mu)K_0(t/\mu),\quad r\in[0,R_1],\end{aligned} \tag{4.2}$$

$$n_2(r)=\int_0^1\frac{d\mu}{\mu^2}F_2(r,\mu), \tag{4.3}$$

$$F_2(r,\mu)=\int_0^{R_1}dt\,t n_1(t)K_0(r/\mu)I_0(t/\mu) + \int_{R_1}^{r}dt\,t n_1(t)K_0(r/\mu)I_0(t/\mu) + \int_{R}^{R_2}dt\,t n_2(t)I_0(r/\mu)K_0(t/\mu),\quad r\in[R_1,R_2], \tag{4.4}$$

where $I_0(x)$ and $K_0(x)$ are modified Bessel functions of zeroth order.

Using the above definitions the transform functions can be shown to satisfy

$$\nabla^2F_i(r,\mu)-\frac{1}{\mu^2}F_i(r,\mu)+c_i\int_0^1\frac{d\mu'}{\mu'^2}F_i(r,\mu')=0,\quad i=1,2. \tag{4.5}$$

The separation of variables technique presented in Sec. 2 leads to solutions in the form

$$F_1(r,\mu)=A_1f_1(\nu_{01},\mu)I_0(r/\nu_{01})+\int_0^1 d\nu A_1(\nu)f_1(\nu,\mu)I_0(r/\nu),\quad \mu\in[0,1],\ r\in[0,R_1], \tag{4.6}$$

$$F_2(r,\mu)=f_2(\nu_{02},\mu)[A_2I_0(r/\nu_{02})+B_2K_0(r/\nu_{02})]+\int_0^1 d\nu f_2(\nu,\mu)\times[A_2(\nu)I_0(r/\nu)+B_2(\nu)K_0(r/\nu)],\quad \mu\in[0,1],\ r\in[R_1,R_2]. \tag{4.7}$$

The appropriate conditions derivable from Eqs. (4.2) and (4.4) which must be satisfied by the transform functions are

(i) $F_1(0,\mu)$ is finite,

(ii) $K_0(R_2/\mu)\dfrac{\partial F_2(R_2,\mu)}{\partial r}+\dfrac{1}{\mu}K_1(R_2/\mu)F_2(R_2/\mu)=0$,

(iii) $F_1(R_1,\mu)=F_2(R_1,\mu)$,

(iv) $\dfrac{\partial F_1(R_1,\mu)}{\partial r}=\dfrac{\partial F_2(R_1,\mu)}{\partial r}$.

The choice of the $R(\nu,r)$ function in Eq. (4.6) insures that (i) is satisfied. The application of (ii) leads to the following equation:

$$A_2f_2(\nu_{02},\mu)g_2(\nu_{02},\mu)+\int_0^1 d\nu A_2(\nu)f_2(\nu,\mu)g_2(\nu,\mu) = B_2f_2(\nu_{02},\mu)h_2(\nu_{02},\mu)+\int_0^1 d\nu B_2(\nu)f_2(\nu,\mu)h_2(\nu,\mu), \tag{4.8}$$

where

$$q_i(x,\mu)=R_2\left(\frac{1}{x}K_0(R_2/\mu)I_1(R_2/x)+\frac{1}{\mu}K_1(R_2/\mu)I_0(R_2/x)\right),$$

$$q_i(x,x)=1,\quad i=1\text{ or }2,$$

and

$$h_2(x,\mu)=R_2\left(K_0(R_2/\mu)\frac{K_1(R_2/x)}{x}-\frac{K_1(R_2/\mu)}{\mu}K_0(R_2/x)\right).$$

The standard techniques of Vekua[11] are now employed to explicitly solve for A_2 and $A_2(\nu)$. We isolate the singular part of Eq. (4.8) to write

$$A_2\phi_+^{(2)}(\mu)+\int_0^1 d\nu A_2(\nu)\phi_\nu^{(2)}(\mu)=\Phi'(\mu)+F(\mu), \tag{4.9}$$

where the following definitions are employed:

$$\Phi'(\mu)=-A_2\phi_-^{(2)}(\mu)q_2(\nu_{02},\mu)-\frac{c_2}{2}\int_0^1 d\nu\frac{\nu A_2(\nu)[2\nu Q_2(\nu,\mu)+1]}{\nu+\mu}$$

$$F(\mu)=\frac{1}{\mu^2}\Big(B_2f_2(\nu_{02},\mu)h_2(\nu_{02},\mu)+\int_0^1 d\nu B_2(\nu)f_2(\nu,\mu)h_2(\nu,\mu)\Big),$$

$$Q_i(\nu,\mu)=\frac{q_i(\nu,\mu)-q_i(\nu,\nu)}{\nu-\mu},$$

$$\phi_\pm^{(i)}(\nu)=\frac{c_i}{2}\frac{\nu_{0i}}{\nu_{0i}-\mu},$$

$$\phi_\nu^{(i)}(\mu)=P\frac{c_2\nu}{2}/\nu-\mu+\lambda_i(\nu)\delta(\nu-\mu).$$

We recognize the functions $\phi_\nu^{(i)}(\mu)$ to be those obtained by case in slab geometry. We choose to utilize the half-range orthogonality properties[12] of these functions to find

$$A_2=S(\nu_{02})+\int_0^1 d\nu' A_2(\nu')K^{(2)}(\nu_{02},\nu'), \tag{4.10}$$

$$A_2(\nu)[1+f(\nu)]=S(\nu)+\int_0^1 d\nu' A_2(\nu')K^{(2)}(\nu,\nu'), \tag{4.11}$$

where

$$K^{(2)}(\eta,\nu')=\frac{1}{N(\eta)}\int_0^1 d\mu\,\phi_\eta^{(2)}(\mu)\phi_{\nu'}^{(2)}(-\mu)W_2(\mu)[2\nu'Q_2(\nu',\mu)+1],\qquad \eta=\nu_{02}\text{ or }\nu,$$

$$f(\nu)=\frac{1}{N(\nu)}\int_0^1 d\mu\,\phi_+^{(2)}(\mu)[\phi_+^{(2)}(\mu)+\phi_-^{(2)}(\mu)]q_2(\nu_{02},\mu),$$

$$N(\eta)=\begin{cases}\int_0^1 d\mu\,\phi_+^{(2)}(\mu)[\phi_+^{(2)}(\mu)+\phi_-^{(2)}(\mu)]q_2(\nu_{02},\mu), & \eta=\nu_{02}\\[4pt] \dfrac{g(c_2,\nu)}{W_2(\nu)}, & \eta=\nu\end{cases}$$

$$W_i(\mu)=(\nu_{0i}-\mu)\gamma_i(\mu),$$

$$\gamma_i(\mu)=c_i\mu/2\,x_i(-\mu)(\nu_{0i}^2-\mu^2),\quad i=1\text{ or }2,$$

$$S(\eta)=\frac{1}{N_2(\eta)}\int_0^1 d\mu\,W_2(\mu)F(\mu)\phi_\eta^{(2)}(\mu).$$

The functions $g(c_i,\nu)$ and $x_i(\nu)$ are the familiar functions found in slab geometry applications and have been tabulated by several authors.[13,14]

Equations (4.10) and (4.11) comprise one set of a "constraint" and Fredholm integral equation pair, common to critical problems. A second pair may be developed by applying boundary conditions (iii) and (iv). We follow a procedure similar to the one employed above to eliminate A_1 and $A_1(\nu)$. Now, however, we use the orthogonality and completeness of the functions $f_1(\nu,\mu)$.[5]

The application of condition (iii) and the appropriate orthogonality properties leads to

$$A_1=\frac{1}{\nu_{01}N_1+I_0(R_1/\nu_{01})}([A_2I_0(R_1/\nu_{02})+B_2K_0(R_1/\nu_{02})]$$

$$\times Z(\nu_{02}, \nu_{01}) + \int_0^1 d\nu [A_2(\nu) I_0(R_1/\nu) + B_2(\nu) K_0(R_1/\nu)] Z(\nu, \nu_{01})),$$

(4.12)

$$A_1(\nu) = \frac{g(c_1, \nu)}{\nu^2 I_0(R_1/\nu)} ([A_2 I_0(R_1/\nu_{02}) + B_2 K_0(R_1/\nu_{02})] Z(\nu_{02}, \nu) + \int_0^1 d\nu' [A_2(\nu') I_0(R_1/\nu') + B_2(\nu') K_0(R_1/\nu')] Z(\nu', \nu)), \quad (4.13)$$

where

$$N_{1+} = \frac{\nu_{01}^2}{g(c_1, \nu_{01})}$$

and

$$\int_0^1 d\mu f_2(\eta, \mu) f_1(\eta' \mu) = Z(\eta, \eta') = (c_2 - c_1),$$

$$(\eta\eta')^2 \frac{P}{\eta^2 - \eta'^2} + \eta^2 [\lambda_1(\eta)\lambda_2(\eta) + \pi^2 c_1 c_2 \eta^2] \delta(\eta - \eta').$$

Applying condition (iv) leads to equations similar in form to Eqs. (4.12) and (4.13), isolating the coefficients A_1 and $A_1(\nu)$. Thus by a simple subtraction A_1 and $A_1(\nu)$ may be eliminated. We find a constraint equation to be

$$Z(\nu_{02}, \nu_{01})[A_2 P_1(\nu_{01}, \nu_{02}) + B_2 q_1(\nu_{01}, \nu_{02})] + \int_0^1 d\nu\, Z(\nu, \nu_{01}) \times [A_2(\nu) P_1(\nu_{01}, \nu) + B_2(\nu) q_1(\nu_{01}, \nu)] = 0, \quad (4.14)$$

and a singular integral equation

$$Z(\nu_{02}, \nu)[A_2 P_1(\nu, \nu_{02}) + B_2 q_1(\nu, \nu_{02})] + \int_0^1 d\nu' Z(\nu', \nu) \times [A_2(\nu') P_1(\nu, \nu') + B_2(\nu') q_1(\nu, \nu')] = 0, \quad (4.15)$$

where

$$P_1(\eta, \eta') = R_1 \left(\frac{I_0(R_1/\eta') I_1(R_1/\eta)}{\eta} - \frac{I_0(R_1/\eta) I_1(R_1/\eta')}{\eta'} \right),$$

$$P_1(\eta, \eta') = 0.$$

Equation (4.15) can be reduced to a Fredholm equation by methods similar to those employed by Kuszell[15] in the multiregion slab problem. We write Eq. (4.15) as

$$B_2 q_1(\nu, \nu_{02}) Z(\nu_{02}, \nu) + \int_0^1 d\nu' Z(\nu', \nu) B_2(\nu') q_1(\nu, \nu') = \phi'(\nu), \quad (4.16)$$

where

$$\phi'(\nu) = -(A_2 P_1(\nu, \nu_{02}) Z(\nu_{02}, \nu) + \int_0^1 d\nu\, Z(\nu', \nu) A_2(\nu') P_1(\nu, \nu')).$$

The singular kernel, $Z(\nu', \nu)$ is separated into singular and regular parts using a partial fraction technique to yield

$$M(\nu) B_2(\nu) + \frac{c_2 - c_1}{2} \int_0^1 d\nu' \frac{P}{\nu' - \nu} B_2(\nu') = \phi(\nu), \quad (4.17)$$

and the following definitions have been employed:

$$\phi(\nu) = (\phi'(\nu) - B_2 q_1(\nu, \nu_{02}) Z(\nu_{02}, \nu)/\nu^2 - \frac{(c_2 - c_1)}{2} \times \int_0^1 d\nu' \frac{\nu'}{\nu' + \nu} B_2(\nu')[1 + 2\nu Q_1(\nu, \nu')],$$

$$M(\nu) = \lambda_1(\nu)\lambda_2(\nu) + \pi^2 c_1 c_2 \nu^2.$$

We recognize Eq. (4.17) as a form analogous to that obtained by Kuszell.[15] Following a similar procedure of solution we obtain

$$B_2(\nu) = \left[1 \Big/ \left(M^2(\nu) + \frac{c_2 - c_1}{2} \nu\pi \right)^2 \gamma_0(\nu) \right] \times \int_0^1 d\nu' \gamma_0(\nu') T(\nu, \nu') \phi(\nu'), \quad (4.18)$$

where

$$T(\nu, \nu') = \frac{(c_2 - c_1)}{2} \nu \frac{P}{\nu - \nu'} + M(\nu)\delta(\nu - \nu'),$$

and $\gamma_0(\nu)$ is the appropriate half-range weight function used by Kuszell.

Thus Eqs. (4.18) and (4.11) constitute a pair of coupled Fredholm equations to be solved for $A_2(\nu)$ and $B_2(\nu)$ subject to the constraint conditions, Eqs. (4.14) and (4.10). Because the equations are homogeneous in expansion coefficients we can arbitrarily set one constant, B_2 say, equal to -1. Once these equations have been solved, the coefficients A_1 and $A_1(\nu)$ follow straightforwardly from Eqs. (4.12) and (4.13). Finally, the particle density in terms of these coefficients is determined from Eqs. (4.1) and (4.3),

$$n_1(r) = A_1 I_0(r/\nu_{01}) + \int_0^1 d\nu\, A_1(\nu) I_0(r/\nu), \quad (4.19)$$

$$n_2(r) = A_2 I_0(r/\nu_{02}) + B_2 K_0(r/\nu_{02}) + \int_0^1 d\nu\, A_1(\nu) I_0(r/\nu) + \int_0^1 d\nu\, B_2(\nu) K_0(r/\nu). \quad (4.20)$$

As with all but the most highly idealized problems which employ techniques similar to those of case, the expansion coefficients are solutions to integral equations involving analytically complex Fredholm or singular kernels. Numerical procedures have been successful in solving similar equations, such as the Neumann series method employed by Mitsis[5] in solving critical single-region problems and those of Doctor[16] in two-region spheres. Alternative procedures such as discrete ordinates and that employed by Bareiss and Neumann[17] to obtain the expansion coefficients from the singular integral equations would be applicable should the iterative Neumann series technique fail to converge. These numerical techniques would not be available for the direct numerical solution of $n(\mathbf{r})$ from Eq. (2.1). In addition, the transform solution provides considerable insight into the analytical structure of the solution for the particle density.

In summary we have demonstrated the equivalence of the replication and boundary condition transform techniques for multiregion critical problems in one-dimensional planar, cylindrical or spherical geometry. The specific application to a two-region cylinder presented herein can be extended in a straightforward manner to multiregion cylinders, spheres, and slabs. The logical extension to general multiregion geometries including nonsymmetric sources and incident particles should follow from the procedures presented in this analysis.

[1]K. M. Case, Ann. Phys. (N.Y.) **9,** 1 (1960).
[2]R. C. Erdmann and C. E. Siewart, J. Math. Phys. **9,** 81 (1968).
[3]T. W. Schnatz and C. E. Siewert, J. Math. Phys. **11,** 766 (1970).
[4]A. Leonard and T. W. Mullikan, Proc. Natl. Acad. Sci. USA **52,** 683 (1964).
[5]G. J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problems," Report No. ANL-6787, Argonne National Laboratory, Argonne, Illinois, 1963.
[6]A. G. Gibbs, J. Math. Phys. **10,** 875 (1969).
[7]O. J. Smith and C. E. Siewert, J. Nucl. Energy **22,** 579 (1968).
[8]O. J. Sheaks, J. Math. Phys. **13,** 203 (1972).
[9]B. Davison and J. R. Sykes, *Neutron Transport Theory* (Oxford U.P., London, 1958).
[10]P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill, New York, 1953).
[11]N. P. Vekua, *Systems of Singular Integral Equations* (P. Noordhoff, Groningen, The Netherlands, 1967).
[12]I. Kuščer, N. J. McCormack, and G. C. Summerfield, Ann. Phys. (Leipz.) **30,** 411 (1964).
[13]K. M. Case, F. de Hoffman, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (Los Alamos Scientific Laboratory, Los Alamos, New Mexico, 1953), Vol. I.
[14]K. Kowalska, "Tables of Functions $X(c_1, \nu)$ and $X^{\pm}(c_1, c_2)$, Report No. 630/IX-A/PR, Institute of Nuclear Research,,Warsaw, Poland, 1965.
[15]A. Kuszell, Acta Phys. Pol. **20,** 567 (1961).
[16]D. Doctor, Trans. Am. Nucl. Soc. **11,** 553 (1968).
[17]E. H. Bareiss and C. P. Neuman, "Singular Integrals and Singular Equations with a Cauchy Kernel and the Method of Symmetric Paring," Report No. ANL-6988, Argonne National Laboratory, Argonne, Illinois, 1965.

Lattice gas with nearest-neighbor interaction in one dimension with arbitrary statistics

C. K. Lai

Physics Department, University of Utah, Salt Lake City, Utah 84112
(Received 19 April 1974)

We define a quantum lattice gas with arbitrary statistics. For a one-dimensional system with nearest-neighbor interaction, we show that the problem is exactly soluble by use of Bethe's hypothesis when the interaction $\Delta=\pm 1$. The ground state energy is then obtained for the fermions of spin 1/2. Two phases are found in the case $\Delta=-1$.

INTRODUCTION

Many years ago, Matsubara and Matsuda[1] considered a *Bose* gas moving on a lattice as a model of critical phenomena in liquid helium. In the language of Ref. 2, this model considers a system of *bosons* interacting through a potential energy which has two parts. First there is a hard core to forbid any two particles from occupying the same site. In addition there is a nearest neighbor interaction equal to -2Δ. One further replaces the usual kinetic energy operator ∇^2 by a double difference. That is, the potential is

$$V=\begin{cases}-2\Delta & \text{for nearest neighbor}\\ \infty & \text{for the hard core,}\end{cases} \tag{1}$$

and the kinetic energy operator is

$$\langle x|\text{ kinetic energy }|\psi\rangle=-\langle x+1|\psi\rangle-\langle x-1|x\rangle+2\langle x|\psi\rangle. \tag{2}$$

The model Hamiltonian then becomes

$$H=\sum(-\text{double difference operator})+\sum_{\text{N.B.}}V \tag{3}$$

(3) is applicable to system in three, two, or one dimension.

In this paper, we first define a generalization of (3) into the case where one could have particles with arbitrary statistics instead of Bose statistics alone. This generalized Hamiltonian is

$$H=\sum(-\text{double difference operator})+\sum_{i<j}V\frac{(1+P_{ij})}{2}, \tag{4}$$

where P_{ij} is the permutation operator. The use of such permutation operator is well known in nuclear physics and was fashionable in the 1930's. For Bose gas, (4) reduces to (3) (as $P_{ij}=1$ for the totally symmetric wavefunction). Next, we establish that for a one-dimensional system, (4) is exactly soluble for any statistics by use of Bethe's hypothesis when $\Delta=\pm 1$. Subsequently we obtain the ground state energy for the fermions of spin $\frac{1}{2}$ in the limit of an infinite system at fixed densities and *two phases will occur* when $\Delta=-1$.

BETHE'S HYPOTHESIS FOR ONE-DIMENSIONAL DIMENSIONAL SYSTEM

Yang[3] has used permutation operators in his solution of the δ-function interaction problem with arbitrary statistics. The same method is applicable to the present problem. We assume that the wavefunction takes the form in Bethe's hypothesis:

$$\psi=\sum_{P}[Q,P]\exp\{i[p_{P1}x_{Q1}+\cdots+p_{PN}x_{QN}]\} \tag{5}$$

for $0<x_{Q1}<\cdots<x_{QN}<L$. $P=[P1,P2\cdots PN]$ and $Q=[Q1,Q2\cdots QN]$ are two permutations of $1,2,\ldots,N$. By adopting the same notations as in Yang's paper,[3] (3) yields the equation

$$\xi_{\ldots ij\ldots}=Y_{ji}^{34}\,\xi_{\ldots ji\ldots},$$

where

$$Y_{ji}^{34}=(y_{ji}^{-1}{}_{-1})+y_{ji}^{-1}P_{34} \tag{6}$$

and

$$y_{ji}^{-1}=\Delta(\exp ip-\exp iq)/\{\exp[i(p+q)]+1-2\Delta\exp iq\} \tag{7}$$

and the energy is given by

$$E=-2\sum_{j}(\cos p_j-1). \tag{8}$$

Now it turns out that identities (Y7), (Y8) hold only for $\Delta=\pm 1$. Thus we only deal with these cases in the following.

SPECIAL CASE OF FERMIONS WITH SPIN ½

As in Ref. 3 the special case of fermions with spin $\frac{1}{2}$ is solved through specializing to a particular representation of the permutation group of the coordinate indices. We discuss cases $\Delta=\pm 1$ separately.

Case $\Delta=-1$

If we make the transformation

$$\alpha_i=\tfrac{1}{2}\tan(p_i/2) \tag{9}$$

then

$$x_{ji}=-i(\alpha_i-\alpha_j)^{-1} \tag{10}$$

and the procedure leading to (Y20), (Y21) can be applied here. Hence for a collection of N fermions with M spin down, one obtains the algebraic equations

$$\exp(ip_jL)=\prod_{l}\frac{i\alpha_j-i\Lambda_l+\frac{1}{2}}{i\alpha_j-i\Lambda_l-\frac{1}{2}}, \tag{11a}$$

$$-\prod_{j}\frac{i\alpha_j-i\Lambda_l+\frac{1}{2}}{i\alpha_j-i\Lambda_l-\frac{1}{2}}=\prod_{l'}\frac{-i\Lambda_{l'}+i\Lambda_l+1}{-i\Lambda_{l'}+i\Lambda_l-1}, \tag{11b}$$

where L is the number of sites. By taking p and Λ to be real, the logarithms of (11a), (11b) give

$$pL=2\pi(I_p+\tfrac{1}{2})+\sum_{\Lambda}\theta(2\alpha-2\Lambda), \tag{12a}$$

$$\sum_{\alpha}\theta(2\Lambda-2\alpha)=2\pi J_\Lambda+\sum_{\Lambda'}\theta(\Lambda-\Lambda'), \tag{12b}$$

where $\theta(x)=2\tan^{-1}x$ and I_p, J_Λ are integers (we take

Copyright © 1974 American Institute of Physics

$M=\text{odd}$, $N=\text{even}$). In the limit $L, N, M\to\infty$ proportionally, one obtains

$$\frac{4}{1+4\alpha^2}=2\pi\rho(\alpha)+\int_{-B}^{B}\frac{4\sigma(\Lambda)\,d\Lambda}{1+4(\alpha-\Lambda)^2}, \tag{13a}$$

$$2\pi\sigma(\Lambda)=\int_{-Q}^{Q}\frac{4\rho(\alpha)\,d\alpha}{1+4(\alpha-\Lambda)^2}-\int_{-B}^{B}\frac{2\sigma(\Lambda')\,d\Lambda'}{1+(\Lambda-\Lambda')^2} \tag{13b}$$

with

$$\frac{E}{L}=\int_{-Q}^{Q}\frac{16\alpha^2\rho(\alpha)\,d\alpha}{1+4\alpha^2},$$
$$\frac{N}{L}=\int_{-Q}^{Q}\rho(\alpha)\,d\alpha,\quad \frac{M}{L}=\int_{-B}^{B}\sigma(\Lambda)\,d\Lambda. \tag{14}$$

For $M/L\ll 1$, one could easily obtain ($N/L=r$)

$$\frac{E}{L}=2r-\frac{2}{\pi}\sin\pi r+\frac{M}{L}\left(\frac{2}{\pi}\sin\pi r-2r\cos\pi r\right)+\cdots. \tag{15}$$

The first two terms are the energy per unit length for a system with $M=0$, i.e., fermions with all parallel spins. The result is expected. On the other hand, the p's in (11a) can be complex numbers when $L\gg 1$. That is,

$$\alpha_i=\Lambda_i\pm i/2+O[\exp(-KL)], \tag{16}$$

$$p_i=2\tan^{-1}2\alpha_i, \tag{17}$$

where we take real part of p_i in the quadrant $(\pi/2,\pi)$ and $(-\pi/2,-\pi)$. Then similar to the δ-function interaction model in the attractive case,[4] one obtains the integral equations

$$\frac{2}{1+\Lambda^2}=2\pi\sigma(\Lambda)+\int_{-B}^{B}\frac{2\sigma(\Lambda')}{1+(\Lambda-\Lambda')^2}\,d\Lambda' +\int_{-Q}^{Q}\frac{4\tau(\gamma)}{1+4(\Lambda-\gamma)^2}\,dr, \tag{18a}$$

$$\frac{4}{1+4\gamma^2}=2\pi\tau(\gamma)+\int_{-B}^{B}\frac{4\sigma(\Lambda)}{1+4(\gamma-\Lambda)^2}\,d\Lambda \tag{18b}$$

with

$$\frac{N}{L}=2\int_{-B}^{B}\sigma\,d\Lambda+\int_{-Q}^{Q}\tau\,d\gamma,\quad \frac{M}{L}=\int_{-B}^{-B}\sigma\,d\Lambda, \tag{19}$$

and

$$\frac{E}{L}=2\int_{-B}^{B}\frac{3+4\Lambda^2}{1+\Lambda^2}\,\sigma(\Lambda)\,d\Lambda+\int_{-Q}^{Q}\frac{16\gamma^2}{1+4\gamma^2}\,\tau\,d\gamma.$$

For $M/L\ll 1$, one easily obtains

$$E/L=2r-(2/\pi)\sin\pi r+(M/L)[(2/r)\sin\pi r -2(r-2)\cos\pi r+2]+\cdots \tag{20}$$

(20) indicates that, at high density, pairing of spin up and spin down particles gives lower erergy than non-pairing as in (15). Therefore, one expects that, at certain density, two phases coexist: one phase defined as collections of fermions with no pairing and the other defined as with pairing. Details will be published elsewhere.

Case $\Delta=+1$

If we make the transformation

$$\alpha=\tfrac{1}{2}\cot(p/2) \tag{21}$$

so that

$$x_{ji}=i(\alpha_j-\alpha_i)^{-1}, \tag{22}$$

then again we can apply the same procedure as case $\Delta=-1$. But now complex solution of the forms (16) will give the lowest energy. The integral equations will be

$$\frac{2}{1+\Lambda^2}=2\pi\sigma(\Lambda)+\int_{[B]}\frac{2\sigma(\Lambda')}{1+(\Lambda-\Lambda')^2}\,d\Lambda' +\int_{[Q]}\frac{4\tau(\gamma)}{1+4(\Lambda-\gamma)^2}\,d\gamma, \tag{23a}$$

$$\frac{4}{1+4\gamma^2}=2\pi\tau(\gamma)+\int_{[B]}\frac{4\sigma(\Lambda)}{1+4(\gamma-\Lambda)^2}\,d\Lambda \tag{23b}$$

with

$$\frac{N}{L}=\int_{[Q]}\tau\,d\gamma+2\int_{[B]}\sigma\,d\Lambda,\quad \frac{M}{L}=\int_{[B]}\sigma\,d\Lambda, \tag{24}$$

and

$$\frac{E}{L}=\int_{[Q]}\frac{4}{1+4\gamma^2}\,\tau\,d\gamma+\int_{[B]}\frac{2}{1+\Lambda^2}\,\sigma\,d\Lambda, \tag{25}$$

where $[Q]$ indicates the integration range is $[-\infty,-Q]$ and $[Q,\infty]$.

REMARKS

One can also obtain the scattering S matrix and the thermodynamics of particles with higher spins. The problem of mixture of fermions and bosons can also be exactly solved. Details will be published elsewhere.

ACKNOWLEDGMENTS

The author is grateful to Dr. C.N. Yang who suggested the problem, and to Dr. B. Sutherland for his interest in this work.

[1] T. Matsubara and H. Matsuda, Progr. Theoret. Phys. (Kyoto) **16**, 569 (1956).
[2] C.N. Yang and C.P. Yang, Phys. Rev. **151**, 261 (1966).
[3] C.N. Yang, Phys. Rev. Lett. **19**, 1312 (1967). We will refer to the equations in this paper as Y1, Y2, etc.
[4] M. Gaudin, thesis, Faculté des Sciences d'Orsay, Université de Paris (1967).

Must quantum theory *assume* unrestricted superposition?

Herbert J. Bernstein

Hampshire College, Amherst, Massachusetts 01002
(Received 21 January 1974)

A conjecture that the $(n-1)^2$ independent moduli and $(2n-1)$ unphysical phases completely specify all n-dimensional unitary matrices is shown to be true in two and three dimensions, but false in four or more. The implications for quantum theory are discussed.

INTRODUCTION

Finite-dimensional unitary matrices arise in both the nonrelativistic and relativistic quantum theories as arrays of transformation coefficients between distinct orthonormal sets of state-vectors (bases). But (since the phase of a basis vector is of no physical significance) not all of the phases of a unitary transformation matrix may have physical significance. In particular, arbitrary diagonal unitary matrices D_1 and D_2 may be used as pre- and post-factors, respectively, in

$$U' = D_1 U D_2 \tag{1}$$

to produce a unitary matrix U' with precisely the same physical meaning as the original matrix, U.

Observe that a constant unimodular matrix may be extracted from both D_1 and D_2, and that it commutes with all matrices. Thus Eq. (1) demonstrates that within the n^2-dimensional continuum of $n\times n$ unitary matrices there are, in general, $(2n-1)$-dimensional subsets of physically equivalent unitary matrices. Furthermore, interpreting the matrix U alternately as an array of normalized column and row vectors, it is clear that there are only $(n-1)^2$ independent moduli of the matrix elements. But, since $(n-1)^2+(2n-1)=n^2$, the independent moduli would seem to complement the physically redundant diagonal factors in providing a parametrization of the entire group $U(n)$. There is, in fact, a recent conjecture[1] that the set of independent moduli of matrix elements completely specifies the physically distinct subset $\{U'\}$ generated in Eq. (1) when D_1 and D_2 run the entire gamut of diagonal unitary matrices.[2]

The theoretical significance of the conjecture is connected with this fact: The absolute value of any matrix element may be measured by ascertaining the transition probability from a single basis state in the initial basis to one in the final basis. In general, this may easily be made operational, since the experimenter *must* be capable of preparing and detecting systems in their basis states in order to define those states. The determination of the phase of a matrix element, however, requires the formation of a linear superposition of at least two basis states. Moreover, the disproof of the conjecture allows the situation that *all* the moduli of a transformation matrix may be measured, without removing all ambiguity in the matrix. This might introduce a new class of ambiguities in the experimental determination of scattering matrices, for example.[3]

After some technical preliminary discussion, we will proceed to prove the conjecture in two and three dimensions, and to demonstrate its failure in four (and hence in all higher) dimensions. The disproof proceeds by displaying a large class of counterexamples.

TECHNICAL PRELIMINARIES

Let us agree to use a "standard representative" of the double coset U' defined by Eq. (1). The previously mentioned constant unimodular matrix may be chosen to render the first entry of D_2 equal to 1. The phases of D_1 and D_2 may then be chosen so that the first row *and* first column of U' are made to be real and positive. For each U, there will be a corresponding standard representative U',

$$U' = \begin{pmatrix} \exp(-i\phi_1) & 0 & \cdots & 0 \\ 0 & \exp(-i\phi_2) & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \exp(-i\phi_n) \end{pmatrix} \times U \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \exp(-i\theta_2) & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \exp(-i\theta_n) \end{pmatrix}, \tag{2}$$

where $\phi_1, \phi_2 \ldots \phi_n$ are the phases of the first column of U and $\theta_2, \theta_3 \ldots \theta_n$ are the phases of the 2nd through nth elements of the first row of U. We will call such a standard U' a "real-bordered" unitary matrix.[4]

In terms of the real-bordered matrix just introduced, the conjecture reduces to the following: that the n^2 moduli of the real-bordered matrix specifies the matrix completely. Thus, a counterexample is provided by finding any two real-bordered unitary matrices whose moduli are equal (position for position), but whose phases differ. Now it is clear that any real-bordered matrix with complex entries will provide a counterexample to the conjecture, for the operation of complex conjugation would convert such a matrix to a different one, which nevertheless had equal moduli. But these two matrices must represent physically equivalent situations for it is precisely the complex conjugation which is necessary in quantum theory to implement the symmetry of time reversal.[5] Again, if the real-bordered matrix has two rows (or columns) whose corresponding entries have the same absolute value, then permutation of the rows (or columns) will provide another form of *discrete* counterexample. But these two matrices would differ merely by a labeling of the states; moreover, we would like to search for a wider class of counterexamples—one which demonstrates the impossibility of the parametrization hypothesized in the conjecture by exhibiting a continuum of standard real-bordered unitary matrices, all having the same moduli. We therefore exclude from our search all isolated discrete counterexamples.[6]

 Copyright © 1974 American Institute of Physics

PROOF OF THE CONJECTURE IN TWO AND THREE DIMENSIONS

For two-by-two matrices, the proof is trivial, since

$$U=\exp(i\alpha I)\exp\left(-i\frac{\psi}{2}\delta_z\right)\exp\left(-i\frac{\theta}{2}\sigma_y\right)\exp\left(-i\frac{\phi}{2}\sigma_z\right) \quad (3)$$

is a unique decomposition, and the y rotation is a real matrix. (Here the σ_i are Pauli spin matrices, and I is the identity matrix.) Clearly this leads to a real-bordered matrix determined by the factor $\exp[-i(\theta/2)\sigma_y]$, whose phases are completely specified by the moduli.

For three-by-three matrices, the argument can be made on geometric grounds, as follows:

$$U=\begin{pmatrix} X_1 & X_2 & X_3 \\ Y_2 & Ae^{i\alpha} & Ce^{i\gamma} \\ Y_3 & Be^{i\beta} & De^{i\delta}\end{pmatrix} \text{ is unitary only if}$$

$$\left.\begin{aligned} X_1X_2+Y_2Ae^{i\alpha}+Y_3Be^{i\beta}&=0\\ X_1X_3+Y_2Ce^{i\gamma}+Y_3De^{i\delta}&=0\\ X_1Y_2+X_2Ae^{i\alpha}+X_3Ce^{i\gamma}&=0\end{aligned}\right\}. \quad (4)$$

Each of these equations is a *triangle* on the complex plane with specified lengths, and whose orientation is fixed by having one real leg. Each therefore has only two solutions. For example, the first of Eqs. (4) can be solved for α and β in two ways, as shown in Fig. 1.

Thus $\alpha=\alpha_0$ and $\beta=\pi+\beta_0$ or $\alpha=-\alpha_0$ and $\beta=\pi-\beta_0$. These two solutions are related by complex conjugation. If the only degeneracy of solution were simultaneous complex conjugation of all phases, we would have the physically equivalent case discussed above.

This is indeed true of the three Eqs. (4); furthermore, the presence of zeroes in the matrix does not alter this conclusion, so the conjecture is proven in three dimensions.

THE COUNTEREXAMPLES

In four dimensions, the following matrix provides a wide class of counterexamples to the conjectured theorem. For every positive definite $a<1$, the matrix

$$U_4(\theta)=\begin{pmatrix} A & B & C\sqrt{1-a} & C\sqrt{a} \\ D & E & F\sqrt{1-a} & F\sqrt{a} \\ G\sqrt{1-a} & H\sqrt{1-a} & ae^{i\theta} & -\sqrt{a(1-a)}\,e^{i\theta} \\ G\sqrt{a} & H\sqrt{a} & -\sqrt{a(1-a)}\,e^{i\theta} & (1-a)\,e^{i\theta}\end{pmatrix} \quad (5)$$

will be a real-bordered unitary matrix as long as

$$U_3=\begin{pmatrix} A & B & C \\ D & E & F \\ G & H & 0\end{pmatrix} \quad (6)$$

is a real-bordered unitary matrix (which implies all entries are real, F and H are negative). Notice that the family of $U_4(\theta)$, as θ varies from 0 to 2π, is a continuous infinity of matrices whose corresponding moduli are always equal, but whose phases are quite different. Note that the counterexamples arise when U_3 is projected into four dimensions in such a way that an orthogonal dyadic may be added to it *without* altering any of the moduli. Clearly this procedure provides counterexamples in all higher dimensions as well.

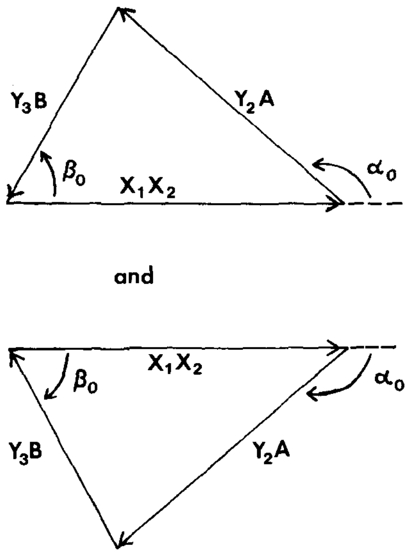

FIG. 1. The geometric solutions to Eq. (4).

It is not known if this is the most general counterexample possible in four dimensions. Since the presence of a zero in U_3 leaves it with only two independent moduli, the real-bordered U_4 has only four continuous parameters (i.e., the two of U_3 plus a and θ) instead of nine. Thus it would seem that a much larger class of counterexamples probably exists. It is likely, therefore, that the practical impossibility of operationalizing unrestricted superposition will lead to real-life examples of the phase indeterminacy adduced above.[7] Finally, the answer to the eponymous question is yes, at least for transformation matrices.

The author would like to thank Dr. J. Pietenpol for a helpful discussion which led to a particular counterexample, and Mr. J. Dell for stimulating and helping a return to this work. Special thanks are due to Professor R. Dashen for originally suggesting the problem.

[1]R. F. Dashen, private communication. The conjecture is due to Dashen and Y. Aharonov.

[2]A mathematician might more precisely state the conjecture: "that the double coset decomposition of the unitary group with respect to its diagonal subgroup is separated by the moduli of the matrix elements."

[3]See N.W. Dean and Ping Lee, Phys. Rev. D 5, 2741 (1972) and references therein. The current investigation may also bear upon the work of Moravcsik, Phys. Rev. D 5, 836 (1972) and its references.

[4]The special case where zeroes occur in the first row or column allows further specification of the "standard representative," but this is an inessential complication which in no way affects the general conclusions of this paper.

[5]See, for example, E. P. Wigner, *Group Theory* (Academic, New York, 1959), Chap. 26.

[6]We would restate the quotation in footnote 2 above: "that the double coset decomposition of the unitary group with respect to its diagonal subgroup is separated (up to discrete isomorphisms induced by either complex conjugation or by permutations) by the moduli of the matrix elements."

[7]An interesting example of the practical limitations on superposition occurs in the case of isotopic spin. It is clear that the physicist can prepare a state of, say one K° and one proton. Through approximate conservation of isotopic spin he might produce, in the same system, the state of total isospin 1, with charge 1 ($I=1$, $I_z=0$). But it is difficult to conceive of a controllable method of producing or detecting most of the linear combinations of states in a 1Kaon—1Nucleon system. Since the transformation matrices (in isospin space) are indeed four-dimensional in this case, the discussion above may be directly applicable. A similar remark may be made with respect to the πN system, where the role of isotopic spin states might be played by the Δ quartet.

Spectrality, cluster decomposition and small distance properties in Wightman field theory

A. Truman

Mathematics Department, Heriot-Watt University, Edinburgh, Scotland
(Received 12 April 1974)

We apply the results of a previous paper by Screaton and Truman to the truncated vacuum expectation values in Wightman field theory and, using spectrality, translational invariance, and Lorentz invariance, we derive the best bounds for the truncated vacuum expectation values at the real Jost points. In a local field theory these bounds include as a special case Araki's result on the exponential decrease of the truncated vacuum expectation value for large spacelike separations and the cluster decomposition property. The bounds also establish a connection between the small distance and high energy behaviors of the theory. In addition we evaluate the bounds in a nonlocal field theory and discuss some of their ramifications.

1. INTRODUCTION

In a previous paper[1] Screaton and Truman derived general holomorphy properties and bounds for the Fourier transforms of distributions with restricted support. In the context of a Wightman field theory, using spectrality, Lorentz invariance, and translation invariance, we here apply these bounds to the truncated vacuum expectation values of the products of a scalar field ϕ.

After Haag[2] the truncated vacuum expectation values, W^T, are inductively defined by

$$\begin{aligned} W(x_0, \ldots, x_n) &= \sum W^T(x_{l_1,1}, x_{l_1,2}, \ldots, x_{l_1}, r_1)\, W^T(x_{l_2,1}, \ldots, x_{l_2}, r_2) \cdots \\ &\quad \times W^T(x_{l_s,1}, \ldots, x_{l_s}, r_s), \end{aligned} \tag{1}$$

where $W(x_0, \ldots, x_n)$ is the vacuum expectation value of the product of field operators (Wightman[3] function):

$$W(x_0, \ldots, x_n) = \langle 0 | \phi(x_0) \cdots \phi(x_n) | 0 \rangle, \qquad W \in \mathcal{S}', \tag{2}$$

$|0\rangle$ is the vacuum state and the sum on the right-hand side of (1) runs over all partitions of $0, 1, \ldots, n$ and in each subset $l_k, 1, \ldots, l_k, r_k$ indices are taken in natural order.

In a translation invariant theory there is a unitary representation $T(a)$ of the translation group $\{a : a \in R^4\}$.

$$\begin{aligned} T(a) &= \exp(iP\cdot a) = \int \exp(ip\cdot a)\, d\epsilon(p), \quad T(a)\phi(x)T(-a) \\ &= \phi(x+a), \\ T(a)|0\rangle &= |0\rangle, \end{aligned} \tag{3}$$

where P is the self-adjoint energy—momentum operator and $\epsilon(p)$ is the corresponding spectral measure. It follows that the truncated vacuum expectation values are translation invariant and are distributions $W^T(\xi_1, \xi_2, \ldots, \xi_n) \in \mathcal{S}'$ in the difference variables $\xi_k = x_k - x_{k-1}$.

If also there is a lowest positive mass, m, particle so that the spectrum of P is $\{0\} \cup V_+^m$, where $V_+^m = \{p : p^2 \geq m^2,\ p_0 > 0\}$, then the inverse Fourier transform of W^T, $\tilde{W}^T(q_1, \ldots, q_n) \in \mathcal{S}'$, has support contained in $V_+^m \otimes^n$.

The results of Screaton and Truman can easily be generalized to yield Theorem 1.

Theorem 1: Let K be a closed subset of R^4 and let Fourier transform be defined as in Footnote 4. $\hat{K} = \{n : q\cdot n \geq \alpha(n) > -\infty,\ \forall\, q \in K\}$, $\hat{K} \subset R^4$, where the dot denotes Minkowski scalar product, $|n| = 1$ and $|\ |$ denotes Euclidean norm. If $\tilde{F}(q_1, \ldots, q_n) \in \mathcal{S}'$ has support $K\otimes^n$, then the Fourier transform of $\tilde{F}$, $F(\xi_1, \ldots, \xi_n)$ is holomorphic in the tube $R^{4n} + i\hat{K}\otimes^n$. If further

$$\xi_j = \mathrm{Re}\,\xi_j + i|\mathrm{Im}\,\xi_j|n_j, \quad j = 1, 2, \ldots, n,\ n_j \in \dot{\hat{K}},$$

$$\begin{aligned} |\tilde{F}(\xi_1, \ldots, \xi_n)| &< C \prod_{j=1}^{n} (|\mathrm{Im}\,\xi_j|^{-s_j} + 1)(|\xi_j|^{r_j} + 1) \\ &\quad \times \exp[-|\mathrm{Im}\,\xi_j|\,\alpha(n_j)], \end{aligned} \tag{4}$$

for some nonnegative integers r_j and s_j, $j = 1, 2, \ldots, n$, and some constant C.

If in the above theorem we identify K with V_+^m, $\hat{K}$ with V_+, and F with W^T, we arrive at bounds for $W^T(\xi_1, \ldots, \xi_n)$ in the forward tube

$$T, T = \{(\xi_1, \ldots, \xi_n) : \mathrm{Im}\,\xi_i \in V_+, i = 1, 2, \ldots, n\}.$$

If in addition the theory is invariant under the proper orthochronous Lorentz group, $L_+^\uparrow$, so that there is a unitary representation $U(\Lambda)$, $\Lambda \in L_+^\uparrow$, with

$$U(\Lambda)\phi(x)U^{-1}(\Lambda) = \phi(\Lambda x), \quad U(\Lambda)|0\rangle = |0\rangle, \tag{5}$$

then, by using the Bargmann—Hall—Wightman[5] theorem, the bounds in the forward tube can be extended to the real Jost[6] points

$$J = \{(\xi_1, \ldots, \xi_n) : (\textstyle\sum \lambda_i \xi_i)^2 < 0,\ \sum \lambda_i = 1,\ \lambda_i \geq 0\}.$$

By considering a certain Hermitian Lorentz transformation we simplify the bounds above for $W^T(\xi_1, \ldots, \xi_n)$, $(\xi_1, \ldots, \xi_n) \in J$. At an equal time Jost point $(\xi_1, \ldots, \xi_n)$, $\xi_i^0 = 0$, $i = 1, 2, \ldots, n$, the simplified formula enables us to evaluate the bounds explicitly. In a local field theory, where $\phi(x)$ and $\phi(y)$ commute for spacelike $(x-y)$, these bounds include, as a special case, Araki's[7] result on the exponential decrease of the truncated vacuum expectation values for large spacelike separations. This leads to the usual cluster decomposition property[8] in a local field theory, equivalent to the statistical independence of widely separated experiments. In a local field theory our results also give a relationship between the small distance and high energy behaviors of the theory—the small distance behavior of the vacuum expectation value in $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ is related to the polynomial growth of the inverse Fourier transform in the momentum vari-

 Copyright © 1974 American Institute of Physics

able conjugate to $(x_i - x_j)$. In a nonlocal field theory we obtain new bounds for the truncated vacuum expectation values. The new bounds lead to a restricted cluster decomposition property and a similar small distance property. The new bounds may be a useful tool in extending Haag–Ruelle collision theory[9] to nonlocal fields.

2. THE SUPPORT PROPERTY OF THE TRUNCATED VACUUM EXPECTATION VALUE

The truncated vacuum expectation values are defined recursively by Eq. (1). From translation invariance

$$W^T(x_0, \ldots, x_m) = W^T(\xi_1, \ldots, \xi_m), \quad \xi_k = x_k - x_{k-1}. \tag{6}$$

In this section we establish the support property of $\widetilde{W}^T(q_1, \ldots, q_{l-1})$, the inverse Fourier transform of $W^T(\xi_1, \ldots, \xi_{l-1})$, $\operatorname{supp} \widetilde{W}^T(q_1, \ldots, q_{l-1}) \subset V_+{}^m \otimes {}^{l-1}$, where

$$W^T(\xi_1, \ldots, \xi_{l-1}) = \int \exp(i\xi_1 \cdot q_1 + \cdots + i\xi_{l-1} \cdot q_{l-1}) \widetilde{W}^T(q_1, \ldots, q_{l-1}) dq_1 \cdots dq_{l-1}, \tag{7}$$

$V_+{}^m = \{q : q^2 \geq m^2,\ q_0 > 0\}$. We prove this result to be a direct consequence of the spectrum S of the energy–momentum operator P being given by $S = \{0\} \cup V_+{}^m$.

Let $X(a)$ be a function such that $\operatorname{supp} \widetilde{X} \cap S = \{0\}$. We shall prove by induction on l that

$$\int X(a) W^T(x_0, \ldots, x_{k-1},\ x_k + a, \ldots,\ x_{l-1} + a)\, d^4a = 0, \tag{8}$$

or

$$\int \widetilde{X}(q_k) \exp(i\xi_1 \cdot q_1 + \cdots + i\xi_{l-1} \cdot q_{l-1}) \times \widetilde{W}^T(q_1, \ldots, q_{l-1}) dq_1 \cdots dq_{l-1} = 0. \tag{9}$$

Since the above equation holds for all ξ it is equivalent to the support of $\widetilde{W}^T(q_1, \ldots, q_{l-1})$ being contained in $V_+{}^m \otimes {}^{l-1}$.

First of all when $l = 2$ we have that

$$\int X(a) W^T(x_0, x_1 + a) d^4a = \int X(a) W(x_0, x_1 + a) d^4a - \int X(a) W(x_0) W(x_1 + a) d^4a, \tag{10}$$

$$\therefore \int X(a) W^T(x_0, x_1 + a) d^4a = \langle \phi(x_0) \int X(a) T(a) d^4a\, \phi(x_1) \rangle_0 - \langle \phi(x_0) \rangle_0 \langle \phi(x_1) \rangle_0 \widetilde{X}(0). \tag{11}$$

However, since $T(a) = \exp(iP \cdot a) = \int \exp(ip \cdot a) d\epsilon(p)$, we have

$$\int X(a) T(a) d^4a = \widetilde{X}(P). \tag{12}$$

Hence from the support property of $\widetilde{X}$ we arrive at

$$\int X(a) T(a) d^4a = \widetilde{X}(0) |0\rangle\langle 0|, \tag{13}$$

where $|0\rangle\langle 0|$ is the projection onto the vacuum. Hence,

$$\int X(a) W^T(x_0, x_1 + a) d^4a = 0. \tag{14}$$

To establish the result inductively, we observe that

$$\begin{aligned} &\int X(a) W(x_0, \ldots, x_{k-1}, x_k + a, \ldots, x_m + a) d^4a \\ &= \langle \phi(x_0)\phi(x_1) \cdots \phi(x_{k-1}) \int T(a) X(a) d^4a\, \phi(x_k) \cdots \phi(x_m) \rangle_0 \\ &= \widetilde{X}(0) W(x_0, \ldots, x_{k-1}) W(x_k, \ldots, x_m). \end{aligned} \tag{15}$$

We now assume Eq. (8) holds for $l < m + 1$. Substitute from (1) for the lhs of (15). By the inductive assumption, apart from $\int X(a) W^T(x_0, \ldots, x_k + a, \ldots, x_m + a) d^4a$, any term from the rhs of (1) which includes W^T evaluated at arguments involving x's from both $(x_0, \ldots, x_{k-1})$ and $(x_k, \ldots, x_m)$ is zero. For instance, if $\{x_{a_1}, \ldots, x_{a_r}\} \cup \{x_{c_1}, \ldots, x_{c_s}\} = \{x_0, \ldots, x_{k-1}\}$ and $\{x_{b_1}, \ldots, x_{b_t}\} \cup \{x_{d_1}, \ldots, x_{d_u}\} = \{x_k, \ldots, x_m\}$, $m + 1 = r + s + t + u$, $a_1 a_2 \ldots a_r$ etc. in natural order,

$$\int = \int X(a) W^T(x_{a_1}, \ldots, x_{a_r}, x_{b_1} + a, \ldots, x_{b_t} + a) \times W^T(x_{c_1}, \ldots, x_{c_s}, x_{d_1} + a, \ldots, x_{du} + a) d^4a, \tag{16}$$

$$\therefore \int = \int X(a) W^T(\xi_1', \ldots, \xi_r' + a, \ldots, \xi_{r+t-1}') \times W^T(\xi_{r+t}', \ldots, \xi_{r+t+s-1}' + a, \ldots, \xi_{m-1}') d^4a, \tag{17}$$

where $\xi_1' = x_{a_2} - x_{a_1}$ etc.,

$$\therefore \int = \int \widetilde{X}(p'_r + p'_{r+t+s-1}) \widetilde{W}^T(p_1', \ldots, p_r', \ldots, p'_{r+t-1}) \times \widetilde{W}^T(p'_{r+t}, \ldots, p'_{r+t+s-1}, \ldots, p'_{m-1}) \times \exp\left(i \sum_{j=1}^{m-1} \xi_j' \cdot p_j'\right) dp'. \tag{18}$$

Since $p_r' \in V_+{}^m$, $p'_{r+t+s-1} \in V_+{}^m \Rightarrow (p_r' + p'_{r+t+s-1}) \in V_+^{2m}$ by inductive hypothesis $\int = 0$. Hence the only terms giving a nonvanishing contribution to the lhs of (15) are

$$\int X(a) W^T(x_0, \ldots, x_k + a, \ldots, x_m + a) d^4a + \widetilde{X}(0) \sum_{\text{part}'} W_{r_1}^T(\ )W_{r_2}^T(\ ) \cdots W_{r_s}^T(\ ), \tag{19}$$

where part′ indicates a summation over all subpartitions of the partition $(0, 1, \ldots, k-1)(k, \ldots, m)$. However, according to Eq. (1), this last term can be written as $\widetilde{X}(0) W(x_0, \ldots, x_{k-1}) W(x_k, \ldots, x_m)$ and

$$\int X(a) W^T(x_0, \ldots, x_k + a, \ldots, x_m + a) d^4a = 0. \tag{20}$$

The result is true by induction.

3. BOUNDS FROM SPECTRALITY

We have seen from spectrality that $\widetilde{W}^T (\in \mathcal{S}')$, the inverse Fourier transform of the truncated vacuum expectation value, is such that $\operatorname{supp} \widetilde{W}^T(q_1, \ldots, q_n) \subset V_+^m \otimes^n$. From Theorem 1 we can deduce that $W^T(\xi_1', \ldots, \xi_n')$ is holomorphic in the forward tube T, $T = \{(\xi_1', \ldots, \xi_n') : \operatorname{Im} \xi_i' \in V_+, i = 1, 2, \ldots, n\}$. Moreover, for $(\xi_1', \ldots, \xi_n') \in \overline{T}$,

$$|W^T(\xi_1', \ldots, \xi_n')| < C \prod_{i=1}^{n} (1 + |\operatorname{Im} \xi_i'|^{-s_i})(1 + |\xi_i'|^{r_i}) \times \exp[-|\operatorname{Im} \xi_i'| \alpha(n_i)], \tag{21}$$

where $\xi_j' = \operatorname{Re}(\xi_j') + i \operatorname{Im}(\xi_j')$, $\operatorname{Im} \xi_j' = |\operatorname{Im} \xi_j'| n_j$, $|n_j| = 1, j = 1, 2, \ldots, n$, ($|\ |$ denotes Euclidean norm) and $\alpha(n_j)$ is such that $\operatorname{supp} \widetilde{W}^T(q_1, \ldots, q_n) \subset \{(q_1, \ldots, q_n) : q_j \cdot n_j \geq \alpha(n_j), j = 1, 2, \ldots, n\}$.

Since $\operatorname{supp} \widetilde{W}^T(q_1, \ldots, q_n) \subset \{(q_1, \ldots, q_n) : q_i^2 \geq m^2, q_i^0 > 0, i = 1, 2, \ldots, n\}$ and the equation of the tangent plane to $q^2 = m^2$ at Q ($Q^2 = m^2$) is $q \cdot Q = m^2$, we have $\alpha(Q/|Q|) = m^2/|Q|$. Putting $Q/|Q| = \operatorname{Im} \xi_i'/|\operatorname{Im} \xi_i'|$ and using $Q^2 = m^2$ gives $|\operatorname{Im} \xi_i'| \alpha(n_i) = m(\operatorname{Im} \xi_i' \cdot \operatorname{Im} \xi_i')^{1/2}$. Hence, for $(\xi_1', \ldots, \xi_n') \in \overline{T}$,

$$|W^T(\xi_1', \ldots, \xi_n')| < C \prod_{i=1}^{n} (1 + |\operatorname{Im} \xi_i'|^{-s_i})(1 + |\xi_i'|^{r_i}) \times \exp[-m(\operatorname{Im} \xi_i' \cdot \operatorname{Im} \xi_i')^{1/2}]. \tag{22}$$

Thus, for $m > 0$, W^T is exponentially decreasing in the forward tube $\mathcal{T}$.

The truncated vacuum expectation values are also Lorentz invariant:

$$W^T(\xi_1, \ldots, \xi_n) = W^T(\Lambda\xi_1, \ldots, \Lambda\xi_n), \qquad \Lambda \in L_+^\dagger. \qquad (23)$$

The Bargmann—Hall—Wightman theorem implies that $W^T(\xi_1, \ldots, \xi_n)$ is holomorphic in the extended tube $\mathcal{T}'$, $\mathcal{T}' = \cup_\Lambda(\Lambda\mathcal{T})$, where the union is taken over $\Lambda \in L_+(C)$, the identity component of the complex Lorentz group. The value of $W^T(\xi_1, \ldots, \xi_n)$ at points of $\mathcal{T}'$ is given by

$$W^T(\xi_1, \ldots, \xi_n) = W^T(\Lambda\xi_1, \ldots, \Lambda\xi_n), \qquad \Lambda \in L_+(C), \quad (24)$$

where $(\xi_1', \ldots, \xi_n') = (\Lambda\xi_1, \ldots, \Lambda\xi_n) \in \mathcal{T}$. In particular $W^T(\xi_1, \ldots, \xi_n)$ is holomorphic at the *real* Jost points $J = \{(\xi_1, \ldots, \xi_n) : (\sum \lambda_i \xi_i)^2 < 0, \ \sum\lambda_i = 1, \lambda_i \geq 0\}$. For $(\xi_1, \ldots, \xi_n) \in J$,

$$|W^T(\xi_1, \ldots, \xi_n)| < C \inf \prod_{i=1}^n (1 + |\mathrm{Im}\xi_i'|^{-s_i})(1 + |\xi_i'|^{r_i}) \times \exp[-m(\mathrm{Im}\xi_i' \cdot \mathrm{Im}\xi_i')^{1/2}], \qquad (25)$$

where the infimum is taken over $(\xi_1', \ldots, \xi_n') = (\Lambda\xi_1, \ldots, \Lambda\xi_n) \in \mathcal{T}$, $\Lambda \in L_+(C)$.

From Eq. (25) we see that, for large $\xi \in J$, $W^T(\xi_1, \ldots, \xi_n)$ is essentially bounded by $\exp[-md(\xi)]$, where $d(\xi)$ is given by

$$d(\xi) = \sup \sum_{i=1}^n (\mathrm{Im}\Lambda\xi_i \cdot \mathrm{Im}\Lambda\xi_i)^{1/2}, \qquad (26)$$

the supremum being taken over $\Lambda \in L_+(C)$, $\mathrm{Im}\Lambda\xi_i \in V_+$, $i = 1, 2, \ldots, n$. We call $d(\xi)$ the diameter of the Jost point. [Araki derived the same exponential bound independently when evaluating $W^T(\lambda\xi_1, \ldots, \lambda\xi_n)$ at the equal time dilated Jost point $(\lambda\xi_1, \ldots, \lambda\xi_n)$.]

From equation (25) we also see that, for small $\xi \in J$, $W^T(\xi_1, \ldots, \xi_n)$ is essentially bounded by

$$C(\xi) = \prod_{i=1}^n (1 + (\mathrm{Im}\Lambda\xi_i \cdot \mathrm{Im}\Lambda\xi_i)^{-s_i/2}), \quad \Lambda \in L_+(C), \quad (\Lambda\xi_1, \ldots, \Lambda\xi_n) \in \mathcal{T}. \qquad (27)$$

In the next section by considering a certain Hermitian Lorentz transformation we find a simplified formula for $d(\xi)$ and the above bound. In a local field theory this leads to bounds for the behavior of $W^T(x_0, \ldots, x_n)$ for both large and small separations. For large separations we rediscover Araki's result—the exponential decrease of the truncated vacuum expectation value, while for small separations we see the value of the truncated vacuum expectation value is determined by the high energy behavior of the theory. In nonlocal field theories our results lead to new bounds for the truncated vacuum expectation values.

4. EVALUATION OF THE BOUND

Lemma 1: When $\Lambda \in L_+(C)$, the identity component of the complex Lorentz group, the Hermitian matrix $H = \Lambda^\dagger G\Lambda$ is a Lorentz transformation with eigenvalues $+1$, -1, k and k^{-1}, $k < 0$. (G is the matrix of the Minkowski inner product.) To see that H is a Lorentz transformation, we write

$$H^TGH = \Lambda^TG\bar{\Lambda}G\Lambda^\dagger G\Lambda = \Lambda^TG\bar{\Lambda}\bar{\Lambda}^{-1}\Lambda = \Lambda^TG\Lambda = G, \qquad (28)$$

where we have used $\Lambda^TG\Lambda = G$, or $\Lambda^{-1} = G\Lambda^TG$, so that $\bar{\Lambda}^{-1} = G\Lambda^\dagger G$.

Since $\Lambda \in L_+(C)$, $\det H = \det(\Lambda^\dagger G\Lambda) = -1$. We now consider the eigenvalue equation $\det(H - \lambda) = 0$. Because H is Hermitian, this equation will have real roots λ. Two of these roots are $\lambda = \pm 1$. This follows from (28) and $\det H = -1$. For we have

$$\begin{aligned}\det(H+1) &= \det(H + G^2) = \det(H + GH^TGH) \\ &= \det(1 + GH^TG)\det H = -\det(1 + H^T) \\ &= -\det(1 + H). \end{aligned} \qquad (29)$$

$$\therefore \det(H+1) = 0. \qquad (30)$$

Similarly $\det(H - 1) = 0$.

As H is Hermitian, $\det H = -1$, the eigenvalues of H must be $+1$, -1, k and k^{-1}, where k is real, $k \neq 0$. In fact, $k < 0$, as we now show. Consider $\mathrm{tr}(H) = \mathrm{tr}(\Lambda^\dagger G\Lambda) = k + k^{-1}$. For real k, $k \neq 0$, $\mathrm{tr}(H) \geq 2$ or $\mathrm{tr}(H) \leq -2$, according as $k > 0$ or $k < 0$, respectively. Also, $\mathrm{tr}(H) = \mathrm{tr}(\Lambda^\dagger G\Lambda)$ is a continuous function of $\Lambda \in L_+(C)$. However, $L_+(C)$ is connected and, when $\Lambda = 1 \in L_+(C)$, $\mathrm{tr}(H) = -2$, therefore, $\mathrm{tr}(H) \leq -2$, $\Lambda \in L_+(C)$. Hence, $k < 0$.

Lemma 2: When $\Lambda \in L_+(C)$, $\exists$ lightlike vectors l_+ and l_- (depending only on Λ) with $l_\pm \in \partial V_\pm$, $l_+ \cdot l_- = -2$, such that for all real 4-vectors ξ with $\mathrm{Im}\Lambda\xi \in V_\pm$

$$\mathrm{Im}\Lambda\xi \cdot \mathrm{Im}\Lambda\xi \leq (\xi \cdot l_+)(\xi \cdot l_-). \qquad (31)$$

Elementary computation yields

$$\mathrm{Im}\Lambda\xi \cdot \mathrm{Im}\Lambda\xi = \tfrac{1}{2}(\xi^TH\xi - \xi^2). \qquad (32)$$

Two cases arise: k, k^{-1} distinct and $k = k^{-1} = -1$.

Case 1: k, k^{-1} distinct: In this case all the eigenvalues of H are distinct. The spectral resolution of H then gives

$$1 = \sum_0^3 x_m x_m^\dagger, \qquad (33)$$

$$H = \sum_0^3 \lambda_m x_m x_m^\dagger, \qquad (34)$$

where x_m, $m = 0, 1, 2, 3$ are the eigenvectors of H:

$$Hx_m = \lambda_m x_m, \qquad m = 0, 1, 2, 3, \qquad (35)$$

$\lambda_0 = 1$, $\lambda_1 = -1$, $\lambda_2 = k$, $\lambda_3 = k^{-1}$, with

$$x_m^\dagger x_n = \delta_{mn}, \qquad m, n = 0, 1, 2, 3. \qquad (36)$$

As H is a Hermitian Lorentz transformation,

$$\bar{H} = H^T = GH^{-1}G. \qquad (37)$$

It follows that we can choose x_m so that, in addition,

$$x_0 = G\bar{x}_0, \quad x_1 = -G\bar{x}_1, \quad x_2 = -G\bar{x}_3. \qquad (38)$$

For instance,

$$Hx_0 = x_0 \Rightarrow \overline{Hx}_0 = \bar{x}_0 \Rightarrow GH^{-1}G\bar{x}_0 = \bar{x}_0 \Rightarrow H(G\bar{x}_0) = G\bar{x}_0, \qquad (39)$$

and similarly $H(G\bar{x}_1) = -G\bar{x}_1$, $H(G\bar{x}_2) = k^{-1}G\bar{x}_2$. (38) follows by suitably adjusting the phases of x_m. When we take (36) in conjunction with (38), we arrive at

$$\begin{aligned} &x_0^2 = -x_1^2 = -x_2 \cdot x_3 = 1, \\ &x_2^2 = x_3^2 = x_0 \cdot x_1 = x_0 \cdot x_2 = x_0 \cdot x_3 = x_1 \cdot x_2 = x_1 \cdot x_3 = 0. \end{aligned} \qquad (40)$$

From (33) and (38)

$$\xi=(\xi\cdot x_0)x_0-(\xi\cdot x_1)x_1-(\xi\cdot x_3)x_2-(\xi\cdot x_2)x_3, \tag{41}$$

$$\therefore \xi^2=(\xi\cdot x_0)^2-(\xi\cdot x_1)^2-2(\xi\cdot x_2)(\xi\cdot x_3). \tag{42}$$

Similarly from (34) and (38) with $k=-a^2$ (real $a>0$)

$$\xi^T H\xi=|\xi\cdot x_0|^2-|\xi\cdot x_1|^2-a^2|\xi\cdot x_2|^2-a^{-2}|\xi\cdot x_3|^2. \tag{43}$$

We now write $x_m=r_m+ij_m$, where r_m and j_m are real, $m=0,1,2,3$. Using Eqs. (32), (42) and (43), we finally arrive at

$$\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi=(\xi\cdot j_0)^2-(\xi\cdot j_1)^2-\tfrac{1}{2}[a(\xi\cdot r_3)-a^{-1}(\xi\cdot r_2)]^2 -\tfrac{1}{2}[a(\xi\cdot j_3)+a^{-1}(\xi\cdot j_2)]^2. \tag{44}$$

The vectors (r_2+ij_2) and (r_3+ij_3) are lightlike, $r_2^2-j_2{}^2=0$, $r_2\cdot j_2=0$, etc. Hence, if $r_2, j_2\neq 0$, the pair of vectors r_2, j_2 must be spacelike. Choose the Lorentz frame in which $r_2=(0,0,r_2,0)$, $j_2=(0,0,0,j_2)$. As $(r_3+ij_3)=-G(r_2-ij_2)$, $r_3=(0,0,r_2,0)$, $j_3=(0,0,0,-j_2)$. Then, since $(r_0+ij_0)=G(r_0-ij_0)$, r_0 is pure timelike and of necessity $j_0\cdot r_2=j_0\cdot j_2=j_0\cdot r_0=0$. Hence, in this Lorentz frame $x_0=(r_0, ij_0, 0, 0)$. Similarly in this Lorentz frame $x_1=(ij_1, r_1, 0, 0)$. In addition $x_0\cdot x_1=0$, so that $r_0 j_1=j_0 r_1$. However, $x_0^2=-x_1^2=1$; therefore, $j_0=\pm j_1$, $j_0^2\leqslant 1$. Finally then, referring back to (44), for $\mathrm{Im}\Lambda\xi\in V_\pm$,

$$\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi\leqslant(\xi\cdot j_0)^2-(\xi\cdot j_1)^2=j_0^2(\xi_1^2-\xi_0^2)\leqslant\xi_1^2-\xi_0^2. \tag{45}$$

Hence,

$$\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi\leqslant(\xi\cdot l_+)(\xi\cdot l_-),$$

where in our Lorentz frame $l_+=(1,1,0,0)$, $l_-=(-1,1,0,0)$, $l_+\cdot l_-=-2$, $l_\pm\in\partial V_\pm$.

The above argument is only valid when r_2, $j_2\neq 0$. When $j_2=0$, say, we have $j_3=0$ and $r_2{}^2=r_3{}^2=0$. Since r_0 and j_1 are pure timelike and $r_0\cdot r_2=j_1\cdot r_2=0$, we must have $r_0=j_1=0$. Hence, choosing a Lorentz frame in which the space-like j_0 and r_1 are given by $j_0=(0,0,1,0)$ and $r_1=(0,0,0,1)$ and satisfying $r_2\cdot r_3=-1$, $r_2=-Gr_3$, one possibility is $r_2=2^{-1/2}(+1,\pm 1,0,0)$ and $r_3=2^{-1/2}(-1,\pm 1,0,0)$. Thus, we see that

$$l_\pm=j_0\pm 2^{-1/2}(ar_2-a^{-1}r_3)=(\pm(a+a^{-1})/2, \pm(a-a^{-1})/2, 1, 0)\in\partial V_\pm \tag{46}$$

and $l_+\cdot l_-=-2$. Moreover, from Eq. (44)

$$\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi=(\xi\cdot l_+)(\xi\cdot l_-)\leqslant(\xi\cdot l_+)(\xi\cdot l_-). \tag{47}$$

The only remaining possibility $r_2=2^{-1/2}(-1,\pm 1,0,0)$, $r_3=2^{-1/2}(1,\pm 1,0,0)$ can be dealt with similarly.

The case $r_2=0$ can also be handled as above.

Case 2: $k=k^{-1}=-1$: As $x_0=G\bar{x}_0$, we can choose a Lorentz frame in which $x_0=(r_0, ij_0, 0, 0)$. We now choose the three vectors x_1, x_2, and x_3 to span the orthogonal complement of $\{x_0\}$. We put $x_1=(ij_0, r_0, 0, 0)$, $x_2=(0,0,1,0)$, $x_3=(0,0,0,1)$. Then $x_m\cdot x_n=G_{mn}$, $r_0^2+j_0^2=1$, and $x_0=G\bar{x}_0$, $x_1=-G\bar{x}_1$, $x_2=G\bar{x}_3$. Arguing as in Case 1, we find

$$\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi=(\xi\cdot j_0)^2-(\xi\cdot j_1)^2=\xi_1^2-\xi_0^2. \tag{48}$$

The result follows.

We can now prove the following theorems.

Theorem 2: A necessary and sufficient condition for $(\xi_1,\xi_2,\ldots,\xi_n)$ to be a Jost point is that $\exists\, l_\pm\in\partial V_\pm, l_+\cdot l_-=-2$, such that $\xi_i\cdot l_+>0$, $\xi_i\cdot l_->0$, $i=1,2,\ldots,n$.

The condition is necessary for if $(\xi_1,\ldots,\xi_n)$ is a Jost point $\exists\,\Lambda\in L_+(C)$ with $\mathrm{Im}\Lambda\xi_i\in V_+$, $i=1,2,\ldots,n$. However, V_+ is convex; from Lemma 2 then $\exists\, l_\pm\in\partial V_\pm$, $l_+\cdot l_-=-2$, such that

$$0<\mathrm{Im}\Lambda\xi\cdot\mathrm{Im}\Lambda\xi\leqslant(\xi\cdot l_+)(\xi\cdot l_-), \tag{49}$$

$\forall\,\xi\in H(\xi_1,\ldots,\xi_n)$, where $H(\xi_1,\ldots,\xi_n)$ is the convex hull of $(\xi_1,\ldots,\xi_n)$. Hence, $(\xi\cdot l_+)$ and $(\xi\cdot l_-)$ are nonzero and have the same sign $\forall\,\xi\in H(\xi_1,\ldots,\xi_n)$. Thus, either $\xi_i\cdot l_+>0$, $\xi_i\cdot l_->0$, $i=1,2,\ldots,n$, or $\xi_i\cdot l_+<0$, $\xi_i\cdot l_-<0$, $i=1,2,\ldots,n$. In the first case there is nothing further to prove and in the second case $-l_-$ and $-l_+$ can be taken as l_+ and l_-, respectively.

To see the condition is sufficient, choose the Lorentz frame in which $l_+=(1,1,0,0)$, $l_-=(-1,1,0,0)$. Then $\xi_0-\xi_1>0$, $\xi_0+\xi_1<0$. Consider the complex Lorentz transformation Λ: $(\xi_0,\xi_1,\xi_2,\xi_3)\to(-i\xi_1, i\xi_0,\xi_2,\xi_3)$. Then $\Lambda\in L_+(C)$ and $\mathrm{Im}\Lambda\xi_i=(-\xi_i^1,\xi_i^0,0,0)\in V_+$, $i=1,2,\ldots,n$. Hence, $(\xi_1,\ldots,\xi_n)$ is a Jost point, proving sufficiency.

[Note that $\mathrm{Im}\Lambda\xi_i\cdot\mathrm{Im}\Lambda\xi_i=(\xi_i\cdot l_+)(\xi_i\cdot l_-)$.]

It is not difficult to prove that the above condition for $(\xi_1,\ldots,\xi_n)$ to be a Jost point is equivalent to the usual condition:

$$(\textstyle\sum\lambda_i\xi_i)^2<0,\quad \lambda_i\geqslant 0,\quad \sum\lambda_i=1.$$

Theorem 3: Here we show that if

$$d'(\xi)=\sup\sum_{i=1}^{n}[(\xi_i\cdot l_+)(\xi_i\cdot l_-)]^{1/2}, \tag{50}$$

where the supremum is taken over $l_\pm\in\partial V_\pm$, $l_+\cdot l_-=-2$, with $\xi_i\cdot l_\pm>0$, $i=1,2,\ldots,n$, then

$$d'(\xi)=d(\xi). \tag{51}$$

Theorem 1 ensures $d'(\xi)>0$. From Lemma 2 evidently $d(\xi)\leqslant d'(\xi)$.

Also, from the second part of Theorem 2, for all $l_\pm\in\partial V_\pm$, $l_+\cdot l_-=-2$, $\xi_i\cdot l_\pm>0$, $i=1,2,\ldots,n$, $\exists\,\Lambda\in L_+(C)$ with $\mathrm{Im}\Lambda\xi_i\in V_+$, $\mathrm{Im}\Lambda\xi_i\cdot\mathrm{Im}\Lambda\xi_i=(\xi_i\cdot l_+)(\xi_i\cdot l_-)$, $i=1,2,\ldots,n$. Therefore, $d'(\xi)\leqslant d(\xi)$. Hence, as asserted, $d'(\xi)=d(\xi)$.

Evidently it is important to identify which points $l_\pm\in\partial V_\pm$ give rise to the supremum. A partial answer to this question is provided by the next theorem.

Theorem 4: If the Jost point $(\xi_1,\ldots,\xi_n)$ lies in the hyperplane $\xi^0=0$ so that $\xi_i=(0,\boldsymbol{\xi}_i)$, $i=1,2,\ldots,n$, and $\Sigma=\sum_{i=1}^n\boldsymbol{\xi}_i$, then

$$d(\xi)=\sup(\mathbf{m}\cdot\Sigma), \tag{52}$$

where the supremum is taken over $m\in M$, $M=\{m:\mathbf{m}\cdot\boldsymbol{\xi}_i>0,\ i=1,2,\ldots,n;\ \mathbf{m}^2=1\}$.

First of all we remove the artificial restriction $l_+\cdot l_-=-2$ in the definition of $d'(\xi)$. We write $l_+=\rho m_+$, $l_-=\sigma m_-$, $\rho,\sigma>0$. Then $\rho\sigma=-2/(m_+\cdot m_-)$. Hence,

$$d(\xi)=\sup\sum_{i=1}^{n}\left(-\frac{2(\xi_i\cdot m_+)(\xi_i\cdot m_-)}{m_+\cdot m_-}\right)^{1/2}, \tag{53}$$

where the supremum is taken over $m_\pm \in \partial V_\pm$, $m_+ \cdot m_- < 0$, with $m_\pm \cdot \xi_i > 0$, $i = 1, 2, \ldots, n$. We can now normalize m_+ and m_- so that $m_+ = (1, \mathbf{m}_+)$, $m_- = (-1, \mathbf{m}_-)$. Writing $\xi_i = (0, \boldsymbol{\xi}_i)$, then

$$d(\xi) = \sup \sum_{i=1}^{n} \left(\frac{2(\boldsymbol{\xi}_i \cdot \mathbf{m}_+)(\boldsymbol{\xi}_i \cdot \mathbf{m}_-)}{1 + \mathbf{m}_+ \cdot \mathbf{m}_-} \right)^{1/2}. \tag{54}$$

Introducing the unit vector $\mathbf{m} = -2^{-1/2}(1 + \mathbf{m}_+ \cdot \mathbf{m}_-)^{-1/2}(\mathbf{m}_+ + \mathbf{m}_-)$ and the vector $\mathbf{n} = 2^{-1/2}(1 + \mathbf{m}_+ \cdot \mathbf{m}_-)^{-1/2}(\mathbf{m}_+ - \mathbf{m}_-)$, we have

$$d(\xi) = \sup \sum_{i=1}^{n} [(\boldsymbol{\xi}_i \cdot \mathbf{m})^2 - (\boldsymbol{\xi}_i \cdot \mathbf{n})^2]^{1/2}, \tag{55}$$

$\mathbf{m} \cdot \boldsymbol{\xi}_i > 0$, $i = 1, \ldots, n$, $\mathbf{m}^2 = 1$. Hence, putting $\mathbf{n} = 0$, we have

$$d(\xi) = \sup (\mathbf{m} \cdot \Sigma), \tag{56}$$

where the supremum is taken over $m \in M$, $M = \{m : \mathbf{m} \cdot \boldsymbol{\xi}_i > 0,\ i = 1, \ldots, n;\ \mathbf{m}^2 = 1\}$.

We call the direction of $\mathbf{m}$ giving rise to the above supremum the Jost point diameter.

5. SOME RESULTS IN A LOCAL FIELD THEORY

We first show how Araki's very elegant result can be derived as a special case of Theorem 4. The reasoning is similar in the first part to that used by Araki.

Araki's result

If the point $(P^{-1}x_0, P^{-1}x_1, \ldots, P^{-1}x_n)$ satisfies $(P^{-1}x_i)^0 = (P^{-1}x_{i+1})^0$, $i = 0, 1, \ldots, n-1$, the $P^{-1}x_i$ can be permuted so that $(\xi_1, \xi_2, \ldots, \xi_n)$, $\xi_i = x_i - x_{i-1}$ is a Jost point and

$$d(\xi) = \sup_{i,j} |\mathbf{x}_i - \mathbf{x}_j| = R. \tag{57}$$

R is called the diameter of the set of points $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$.

Let $R = |P^{-1}\mathbf{x}_k - P^{-1}\mathbf{x}_l|$. Put $x_0 = P^{-1}x_k$ and $x_n = P^{-1}x_l$. Choose the 1 axis parallel to $(\mathbf{x}_n - \mathbf{x}_0)$ and the 2 and 3 axes any two perpendicular axes at right angles to $(\mathbf{x}_n - \mathbf{x}_0)$. We now permute the $P^{-1}x_i$ so that

$$\begin{aligned} &(x_i)^1 > (x_{i-1})^1, \\ \text{or } &(x_i)^1 = (x_{i-1})^1, \quad (x_i)^2 > (x_{i-1})^2, \\ \text{or } &(x_i)^1 = (x_{i-1})^1, \quad (x_i)^2 = (x_{i-1})^2, \quad (x_i)^3 > (x_{i-1})^3. \end{aligned} \tag{58}$$

Clearly the vector $\mathbf{l} = (1, 0, 0)$ is such that $\mathbf{l} \cdot \boldsymbol{\xi}_i \geq 0$, $i = 1, 2, \ldots, n$. As the function $\boldsymbol{\xi}_i \cdot \mathbf{m}$ is a continuous function of $\mathbf{m}$, for sufficiently small $\epsilon, \epsilon' > 0$, the vector $\mathbf{m} = ((1 - \epsilon^2 - \epsilon'^2)^{1/2}, \epsilon, \epsilon')$ satisfies $\boldsymbol{\xi}_i \cdot \mathbf{m} > 0$, $\boldsymbol{\xi}_i = \mathbf{x}_i - \mathbf{x}_{i-1}$, and $(\xi_1, \ldots, \xi_n)$ is, therefore, a Jost point. Moreover,

$$\Sigma \cdot \mathbf{m} = (\mathbf{x}_n - \mathbf{x}_0) \cdot \mathbf{m} = (1 - \epsilon^2 - \epsilon'^2)^{1/2} R. \tag{59}$$

From Theorem 4 then $d(\xi) = R$, establishing the result.

Cluster decomposition property

In a local field theory we can now deduce a general bound for $W^T(x_0, \ldots, x_n)$, $x_0^0 = x_1^0 = \cdots = x_n^0$. Referring back to Eq. (25) and permuting the x's by locality, we see that if $x_0^0 = x_1^0 = \cdots = x_n^0$ and $\mathbf{x}_i \neq \mathbf{x}_j$, $i \neq j$, then

$$|W^T(x_0, \ldots, x_n)| < C(\theta) P(R) \exp(-mR), \tag{60}$$

where R is the diameter of the set of points $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$, P is a fixed polynomial and $C(\theta)$ can be regarded as a constant depending upon the angles between $(\mathbf{x}_j - \mathbf{x}_k)$ and the diameter of $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$ for large separations.

We then have, $x_i \neq x_j$,

$$\exp(m'R) W^T(x_0, \ldots, x_n) \to 0, \tag{61}$$

as $R \to \infty$, for all $m' < m$. It is now a simple matter to use Eq. (1) to derive the cluster decomposition property, for all $m' < m$:

$$\begin{aligned} \exp(m'|\mathbf{a}|) \{ &W(x_0, \ldots, x_{k-1}, x_k + a, \ldots, x_n + a) \\ &- W(x_0, \ldots, x_{k-1}) W(x_k, \ldots, x_n)\} \to 0, \end{aligned} \tag{62}$$

as $|\mathbf{a}| \to \infty$, $a = (0, \mathbf{a})$. This establishes the statistical independence of widely separated experiments.

Small distance behavior

First of all, if the point $(x_0, x_1, \ldots, x_n)$ satisfies $(x_i)^0 = (x_{i+1})^0$, $i = 0, 1, \ldots, n-1$, and $\inf_{i,j} |\mathbf{x}_i - \mathbf{x}_j| = |\mathbf{x}_k - \mathbf{x}_l|$, then we can permute the x_i so that $(Px_0, Px_1, \ldots, Px_n)$ is a Jost point $(P\xi_1, \ldots, P\xi_n)$, $P\xi_i = Px_i - Px_{i-1}$, and, for some integer t, $P\xi_t = x_k - x_l$. To see this we merely choose the 3 axis parallel to $(\mathbf{x}_k - \mathbf{x}_l)$, two perpendicular axes at right angles to $(\mathbf{x}_k - \mathbf{x}_l)$ for the 1 and 2 axes and permute the x's as explained above.

Consider $W^T(x_0, \ldots, x_n)$ as $|\mathbf{x}_k - \mathbf{x}_l| \to 0$. From above, for sufficiently small $|\mathbf{x}_k - \mathbf{x}_l|$, $W^T(x_0, \ldots, x_n) = W^T(P\xi_1, \ldots, P\xi_n)$. The behavior of $W^T(P\xi_1, \ldots, P\xi_n)$ as $P\xi_t \to 0$, is determined by the function $C(\theta) = C(\xi)$, where

$$C(\xi) = \prod_{i=1}^{n} [1 + (\boldsymbol{\xi}_i \cdot \mathbf{m})^{-s_i}], \tag{63}$$

$\mathbf{m}$ being the direction of the Jost point diameter. Hence, for sufficiently small $r = |\mathbf{x}_k - \mathbf{x}_l|$,

$$|W^T(x_0, \ldots, x_n)| < M(1 + r^{-s_t}), \tag{64}$$

where M is a constant and s_t a positive integer.

It is not difficult to check that the integer s_t is essentially the degree of the polynomial growth of $\tilde{W}^T(q_1, \ldots, q_n)$ in the momentum variable q_t conjugate to $P\xi_t$.[10] The high energy behavior of the theory thus determines the small distance properties of the theory. This small distance property still obtains as $m \to 0$. The result is, therefore, also true for the vacuum expectation value of the product of field operators.

6. SOME RESULTS IN A NONLOCAL FIELD THEORY

The bound at an equal-time Jost point

When the Jost point $(\xi_1, \ldots, \xi_n)$ lies in the hyperplane $\xi^0 = 0$, so that $\xi_i = (0, \boldsymbol{\xi}_i)$, $i = 1, \ldots, n$, $\Sigma = \sum_{i=1}^n \boldsymbol{\xi}_i$, then

$$d(\xi) = \sup_{k,l,m} \{d_0, d_k, d_{lm}\}, \tag{65}$$

where $d_0 = |\Sigma|$, $\boldsymbol{\xi}_i \cdot \Sigma \geq 0$, $i = 1, \ldots, n$; $d_k = |\Sigma \wedge \boldsymbol{\xi}_k| / |\boldsymbol{\xi}_k|$, $\boldsymbol{\xi}_k \cdot \Sigma < 0$, $(\boldsymbol{\xi}_k \wedge \boldsymbol{\xi}_i) \cdot (\boldsymbol{\xi}_k \wedge \Sigma) \geq 0$, $i = 1, \ldots, n$; and $d_{lm} = \Sigma \cdot (\boldsymbol{\xi}_l \wedge \boldsymbol{\xi}_m) / |\boldsymbol{\xi}_l \wedge \boldsymbol{\xi}_m|$, where $(\boldsymbol{\xi}_l \wedge \boldsymbol{\xi}_m) \cdot (\Sigma \wedge \boldsymbol{\xi}_m) < 0, \wedge$ $(\boldsymbol{\xi}_m \wedge \boldsymbol{\xi}_l) \cdot (\Sigma \wedge \boldsymbol{\xi}_l) < 0$ (at least one of $\boldsymbol{\xi}_l \cdot \Sigma$, $\boldsymbol{\xi}_m \cdot \Sigma$ is negative) and $\boldsymbol{\xi}_i \cdot (\boldsymbol{\xi}_l \wedge \boldsymbol{\xi}_m) \geq 0$, $i = 1, \ldots, n$.

To establish the above result, referring to Theorem 4, one simply uses Lagrange multipliers to find the extrema of $f(m) = \mathbf{m} \cdot \Sigma$, $m \in M$, where $M = \{m : \mathbf{m}^2 = 1,\ \mathbf{m} \cdot \boldsymbol{\xi}_i > 0,\ i = 1, \ldots, n\}$. The values d_0, d_k and d_{lm} refer to the possibilities that the extrema occur at an interior point of M, on the boundary of M and at a corner of the boundary of M. The requirement that these extrema are suprema give rise to the subsidiary conditions.

The functions d_0, d_k, and d_{lm} are the following Lorentz invariants:

$$d_0^2 + \Sigma^2 = 0, \quad \xi_k^2 d_k^2 + \begin{vmatrix} \Sigma^2 & \Sigma\cdot\xi_k \\ \Sigma\cdot\xi_k & \xi_k^2 \end{vmatrix} = 0,$$

$$\begin{vmatrix} \xi_l^2 & \xi_l\cdot\xi_m \\ \xi_l\cdot\xi_m & \xi_m^2 \end{vmatrix} d_{lm}^2 + \begin{vmatrix} \Sigma^2 & \Sigma\cdot\xi_l & \Sigma\cdot\xi_m \\ \Sigma\cdot\xi_l & \xi_l^2 & \xi_l\cdot\xi_m \\ \Sigma\cdot\xi_m & \xi_l\cdot\xi_m & \xi_m^2 \end{vmatrix} = 0. \tag{66}$$

In a nonlocal field theory with a lowest positive mass m particle, for a given Jost point configuration $(\xi_1, \ldots, \xi_n)$, $\xi_i^0 = 0$, $i = 1, \ldots, n$, using the above, we can find the corresponding $d(\xi)$ explicitly. We then have

$$|W^T(\xi_1, \ldots, \xi_n)| < C(\xi)P(\xi)\exp[-m\, d(\xi)], \tag{67}$$

where $P(\xi)$ is a polynomial in the Lorentz invariants $(\xi_i\cdot\xi_j)$ and $C(\xi)$ is given by (63). It is not difficult to show that this bound implies a small distance property similar to that in a local field theory and a restricted cluster decomposition property.

The above bound for the truncated vacuum expectation value at an equal-time Jost point can also be derived for nonzero spin fields. It is an attractive proposition that it may be possible to extend this bound in some way to enable us to define a Haag–Ruelle collision theory for nonlocal fields. This is desirable as nonlocal field theories seem likely for higher spin particles.

The case $n < 3$

When $n < 3$, there always exists a Lorentz frame in which $(\xi_1, \ldots, \xi_n)$ lies in the hyperplane $\xi^0 = 0$.[11] Hence in a nonlocal field theory with a lowest positive mass particle the above gives us a bound for $W^T(x_0, \ldots, x_n)$, $n < 3$, at the Jost point configurations $(x_0, \ldots, x_n)$.

As an application of this result consider the value of the vacuum expectation value of the commutator $[\phi(x_0), \phi(x_1)]$ in a nonlocal field theory. Classically the mass spectrum condition implies that no signal can propagate with a speed greater than the speed of light. In quantum field theory this last condition is thought to correspond to the vanishing of the commutator $[\phi(x_0), \phi(x_1)]$ for spacelike $(x_0 - x_1)$. Since, when $n = 1$, $(x_1 - x_0)$ is a Jost point whenever $(x_1 - x_0)^2 = -\rho^2 < 0$ and from above $d(\xi) = \rho$, it is easy to see that the mass spectrum condition in a nonlocal field theory implies, for $(x_1 - x_0)^2 = -\rho^2 < 0$,

$$|\langle[\phi(x_0), \phi(x_1)]\rangle_0| < P(\rho)\exp(-m\rho), \tag{68}$$

where P is a fixed polynomial. This appears to be the analog of the classical result.

It is worthwhile observing that this last bound can be used instead of locality in defining asymptotic states in two nonlocal fields.

ACKNOWLEDGMENT

It is a pleasure to thank Professor N. H. Kuiper and his staff for the hospitality extended to me during a month's stay at IHES, Bures Sur Yvette, where most of this work was done.

[1] G. R. Screaton and A. Truman, J. Math. Phys. **14**, 982 (1973).
[2] R. Haag, Phys. Rev. **112**, 669 (1958).
[3] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), Chap. 2.
[4] We are using the convention $\mathcal{J}f(\xi) = \int f(q)\exp(i\xi\cdot q)dq$; $f \in \mathcal{S}$, where the dot denotes Minkowski scalar product. A discussion of Fourier and Laplace transforms of distributions can be found in R. F. Streater and A. S. Wightman, Ref. 3.
[5] D. Hall and A. S. Wightman, Mat.-Fys. Medd. Danske. Vid. Selsk. **31**, No. 5 (1957).
[6] R. Jost, Helv. Phys. Acta **31**, 263 (1958).
[7] H. Araki, Ann. Phys. (N.Y.) **11**, 260 (1960).
[8] R. F. Streater and A. S. Wightman, Ref. 3, Theorem 3.4 and R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, R.I., 1965), Sec. 3.6.
[9] R. Jost, Ref. 8, Chap. 6.
[10] G. R. Screaton and A. Truman, Ref. 1, see proof of Theorem 1.
[11] R. F. Streater and A. S. Wightman, Ref. 3, see for instance Theorem 4.17.

Orbits of the rotation group on spin states

H. Bacry

Centre de Physique Théorique, C.N.R.S., 31, Chemin Joseph Aiguier, 13274 Marseille Cedex 2, France
(Received 12 February 1973)

A simple theorem on projective spaces generalizes the concept of the Riemann sphere. This leads us to a generalized interpretation of the ray space associated with a finite-dimensional Hilbert space. An application is given about the way the rotation group acts on states of given spin j.

The set of pure states associated with the Hilbert space $\mathcal{H}_2$ of dimension two is known to be isomorphic to the ordinary real sphere S^2. That is the case of the sphere of spin $\frac{1}{2}$ states or the Poincaré–Stokes sphere of polarization states of a photon of given energy–momentum. From the mathematical point of view, such a sphere is known as the *Riemann sphere*, isomorphic to $\hat{\mathbb{C}}^2$, the projective space in $\mathbb{C}^2$, i.e., a compactified form of the complex line $\mathbb{C}$.

We first intend to generalize the concept of the Riemann sphere. For this purpose, let us denote by $\hat{\mathbb{C}}^n$ the projective space associated with $\mathbb{C}^n$. If $(a_1, a_2, \ldots, a_n)$ denotes an element of $\mathbb{C}^n$, the projective space $\hat{\mathbb{C}}^n$ is the set of classes of equivalent elements, the equivalence relation being defined by

$$(a_1, a_2, \ldots, a_n) \sim (b_1, b_2, \ldots, b_n) \Leftrightarrow \lambda \in \mathbb{C} - \{0\}, \quad a_i = \lambda b_i.$$

Now, let K_n be the set of all nonzero homogeneous polynomials of degree n in two complex variables. Let us define on K_n the following equivalence relation

$$p, p' \in K_n, \quad p \sim p' \Leftrightarrow \lambda \in \mathbb{C} - \{0\}, \quad p = \lambda p',$$

and let us denote by $\hat{K}_n$ the set of equivalence classes. It can be easily proved that there exists a bijection of $\hat{K}_{n-1}$ onto $\hat{\mathbb{C}}^n$. Indeed, we can associate with each element $(a_1, a_2, \ldots, a_n)$ of $\mathbb{C}^n$ the polynomial $a_1 x^{n-1} + a_2 x^{n-2} y + \cdots + a_{n-1} x y^{n-2} + a_n y^{n-1}$ of K_{n-1}. It is also clear that the equivalence relation preserves the one-to-one correspondence.

Due to the fundamental theorem on roots of a polynomial, any element of K_n can be factorized in a product of n elements of K_1. This decomposition is not unique since (i) we can multiply simultaneously two factors by α and $1/\alpha$, respectively, (ii) the order of factors is arbitrary. Nevertheless, the factorization of an element of $\hat{K}_n$ into elements of $\hat{K}_1$ is unique *up to a permutation.* Therefore, $\hat{K}_n$ is isomorphic to the symmetrized product $(\hat{K}_1 \times \hat{K}_1 \times \cdots \times \hat{K}_1)/S(n)$, where $S(n)$ denotes the permutation group of n elements. Remembering that $\hat{K}_1$ is in one-to-one correspondence with S^2, we then have the following theorem:

Theorem: The projective space $\hat{\mathbb{C}}^n$ can be identified with the set of all sets of $n-1$ points[1] on the real sphere S^2.

From the physical point of view, it readily follows that a state (a ray) in the Hilbert space $\mathcal{H}_n$ of dimension n can be considered as a set of $(n-1)$ points on S^2. As an example, *a state of spin j is represented by $2j$ points on S^2*. In the special case of spin states, one of the advantages of such a geometrical description is to make evident the action of the rotation group on states: We only have to embed S^2 in the ordinary three-dimensional Euclidean space.

Let us first examine the case of spin $\frac{1}{2}$. Let $|\hat{r}\rangle$ the state corresponding to the unit vector $\hat{r}$ (it is the eigenstate of the component $\mathbf{J} \cdot \hat{r}$ with eigenvalue $\hbar/2$). The scalar product $\langle \hat{r} | \hat{r}' \rangle$ is, up to a phase factor, equal to $\frac{1}{2}(1 + \hat{r} \cdot \hat{r}')$. In particular $|\hat{r}\rangle$ and $|-\hat{r}\rangle$ are two orthogonal states (opposite on the sphere S^2). The fact that any state is an eigenstate of some component of $\mathbf{J}$ is a consequence of the transitive action of $SO(3)$ on S^2. (This property is no longer true for higher spin states.) Obviously $SO(2)$ is the stabilizer on a given state which proves that S^2 is isomorphic to the coset space $SO(3)/SO(2)$.

Let us now look at spin 1 states. In the $|jm\rangle$ notation, such a state is a linear combination of eigenstates $|1\,1\rangle$, $|1\,0\rangle$, $|1\,-1\rangle$ of J_z. The state $|1\,1\rangle$ is known to be the symmetric tensor product of $|\frac{1}{2}\,\frac{1}{2}\rangle$ by itself. Therefore, the state $|1\,1\rangle$ will be represented in our geometry by two points at, say, North pole. The state $|1\,0\rangle$ is obtained either as the symmetric tensor product of $|\frac{1}{2}\,\frac{1}{2}\rangle$ by $|\frac{1}{2}\,-\frac{1}{2}\rangle$ (orthogonal states) or by applying the lowering operator J^-. Therefore, we have the description of the three eigenstates in Fig. 1.

This can be easily generalized: The lowering operator J_- associated with the z direction (South–North direction) when applied to a state with points at North and South poles positions, puts down one point.[2]

Let us now make the rotation group acting on a spin 1 state. Let us first consider the case of Fig. 1a. The stabilizer of the state $|1\,1\rangle$ is $SO(2)$. The corresponding orbit is $SO(3)/SO(2)$. The same result is valid for the state $|1\,-1\rangle$. (Obviously $|1\,1\rangle$ and $|1\,-1\rangle$ are on the same orbit.[3]) In the case of Fig. 1b, the stabilizer is the two-sheeted group containing $SO(2)$ as a subgroup and a rotation of angle π around an equatorial axis. This group will be denoted by $O(2)$ to which it is isomorphic. Therefore, the orbit of $|1\,0\rangle$ is $SO(3)/O(2)$. Now the most general orbit will be three-dimensional. This can be shown in the following way. If we put two points on S^2 in arbitrary positions, i.e., not on the same diameter, there always exists a rotation of angle π which maps

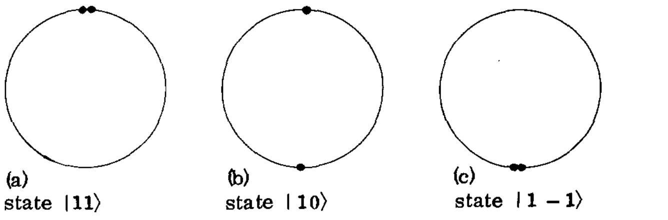

FIG. 1.

Copyright © 1974 American Institute of Physics

the points one on the other. This rotation together with the identity transformation form the stability group $C(2)$ of the state. Therefore, the most general orbit is isomorphic to $SO(3)/C(2)$. Such orbits are parametrized by an angle θ such that $0 < \theta < \pi$. A union of orbits with the same stabilizer is called a stratum. We have then proved that we have three strata on spin 1 states. They are

$\theta = 0$, a single orbit of dimension 2, isomorphic to $SO(3)/SO(2)$,

$0 < \theta < \pi$, a continuous set of orbits of dimension 3, isomorphic to $SO(3)/C(2)$,

$\theta = \pi$, a single orbit of dimension 2, isomorphic to $SO(3)/O(2)$.

It is now interesting to look for a generalization of the above results for all spin j states. For this purpose, we found more convenient to look for those space representations which involve a given type of orbit. In other words, given a (closed) subgroup H of $SO(3)$, how many points can be put on the sphere in such a way that this set of points has H as a stabilizer?

Let us first consider the case of a stabilizer of type $C(n)$, the cyclic subgroup of order n. Let Δ be the axis of rotations of angle $2\pi/n$ which form the group $C(n)$. We will call it the vertical axis. It is clear that the points representing a state which is invariant under $C(n)$ are necessarily either on the axis Δ itself or at the vertices of some polygon having Δ as an axis and the order of which is a multiple of n. We want $C(n)$ to be the stabilizer of the state, that is the *maximal* subgroup which leaves the state invariant. This implies that the state contains at least one polygon of order n since without any polygon the stabilizer would contain $SO(2)$ as a subgroup. Because the number of points on Δ is unlimited a necessary and sufficient condition for $C(n)$ to be the stabilizer of some state of spin j is $2j \geq n$.[4]

Let us now examine the case of the dihedral groups $D(n)$ as stabilizers. We denote by Δ the vertical axis of symmetry and by δ the corresponding diameter. The $2j$ points must be either on δ in even number and/or the vertices of a polygon the order of which is a multiple of n ($n \geq 2$). We must distinguish between the two following cases:

(i) There are nonequatorial polygons. The number of them is necessarily even (at least two) due to the symmetry properties of $D(n)$. The corresponding number of points is a multiple of $2n$. Since the number of points in the equatorial plane is a multiple of n and the number of axial points is even, we get the condition

$$2j = 2na + nb + 2c,$$

where $a \geq 1$, $b \geq 0$, $c \geq 0$.

(ii) There is no polygon except in the equatorial plane. In such a case, we get

$$2j = nb + 2c,$$

with $b \geq 1$, $c \geq 0$. In fact, this result is not valid when n equals 2 because this is a situation where the symmetry is larger if $c = 0$, $b = 1$ (no point on Δ) or if $c = b = 1$ (symmetry of the square) and when $n = 4$, $b = c = 1$ because the symmetry is the one of the octahedron.

The results are the following:

$SO(3)/D(2)$ is present for integral values of j (except $j = 1$),
$SO(3)/D(4)$ is present for integral values of j (except $j = 1$ and 3),
$SO(3)/D(n)$ with $n \neq 2$, 4 is present for $2j = n + nb + 2c$ (b and c nonnegative integers).

The situation is much simpler with the polyhedron groups. It is for instance quite obvious that for the tetrahedron group T, $2j$ must be a multiple of four. This is due to the fact that the tetrahedron is the only polyhedron with T symmetry which can be inscribed in a sphere.

For the octahedron group O, the $2j$ points must be at the vertices of an octahedron and/or a cube. Therefore, $SO(3)/O$ occurs for all values of $2j$ satisfying $2j = 8a + 6b > 0$, where a and b are nonnegative integers. In the same way, the icosahedron group Y will provide orbits for $2j$, $2j = 20a + 12b > 0$.

We are now left with the trivial cases $SO(2)$ and $O(2)$. It is quite obvious that $SO(2)$ occurs in all cases (consider states $|jj\rangle$) and $O(2)$ occurs in integral representations (orbits of states $|j0\rangle$). Table I gives a résumé of the above results.

We have thus classified all orbits associated with *closed* subgroups of $SO(3)$. The last line in Table I corresponds to the trivial case where points are put on a sphere without symmetry property.

The same geometric properties could be used to find out the orbits of the group $O(3)$. Other applications will be derived elsewhere.

ACKNOWLEDGMENTS

The author is grateful to A. Grossmann, B. Morin, and R. Stora for their helpful critical remarks.

APPENDIX

It is interesting to make explicit the relationship between the generalized Riemann sphere and spinor theory. This can be made in the following way.

TABLE I.

Stability group	Type of representation $D_j (j > 0)$
$O(2)$	j integer
$SO(2)$	all
$C(n)$	$2j \geq n$
$D(2)$	j integer (except 1)
$D(4)$	j integer (except 1 and 3)
$D(n)$ for $n > 2$ (except 4)	$2j = n + na + 2b$ [a]
T	j even
O	$j = 4a + 3b$ [a]
Y	$j = 10a + 6b$ [a]
Unit element	$j > 1$

[a] a and b are nonnegative integral.

A. spin $\frac{1}{2}$ states

A stereographic projection from North pole maps S^2 on $\mathbb{C}$. The point with spherical coordinates (θ, φ) is mapped on $z = \operatorname{cotg}(\theta/2)\exp(i\varphi) \in \mathbb{C}$. Corresponding spinors are

$$\hat{\psi} = \begin{pmatrix} \cos(\theta/2)\exp(i\varphi/2) \\ \sin(\theta/2)\exp(-i\varphi/2) \end{pmatrix} \sim \begin{pmatrix} z \\ 1 \end{pmatrix}$$

(this notation obviously includes the possibility $z = \infty$ corresponding to $\theta = 0$).

Two states $\hat{\psi}$ and $\hat{\psi}'$ are orthogonal if

$$\bar{z}'z + 1 = 0,$$

which means that the corresponding points (θ, φ) and (θ', φ') are opposite on S^2.

The $SO(3)$ action on S^2 is described by the $SU(2)$ action on $\mathbb{C}$:

$$z \xrightarrow{U} \frac{az+b}{-\bar{b}z+\bar{a}}, \quad \text{where } U = \begin{vmatrix} a & b \\ -\bar{b} & \bar{a} \end{vmatrix} \in SU(2).$$

B. Spin j states

States of spin j are known to be obtained as symmetric tensors on spinor space. Let us consider the symmetric tensor built on the spinors

$$\begin{pmatrix} z_1 \\ 1 \end{pmatrix}, \begin{pmatrix} z_2 \\ 1 \end{pmatrix}, \cdots \begin{pmatrix} z_n \\ 1 \end{pmatrix},$$

with $n = 2j$. Its components are

$$T_{11\cdots 11} = z_1 z_2 \cdots z_n,$$

$$T_{11\cdots 12} = \frac{1}{\sqrt{n}}(z_1 z_2 \cdots z_{n-1} + z_1 z_2 \cdots z_{n-2} z_n + \cdots + z_2 z_3 \cdots z_n),$$

$$\vdots$$

$$T_{22\cdots 2} = 1.$$

More generally the component $T_{11\cdots 12\cdots 2}$ with p indices 1 is given by $[(n-p)!p!/n!]^{1/2}\, S(z_{i_1} z_{i_2} \cdots z_{i_p})$, where S is the symmetrizer. These equations provide us with the *exact* relationship between the *Euclidean* sphere S^2 and homogeneous polynomials.

[1]Not necessarily distinct.
[2]It readily follows that the set of all lowering operators is identical to the set of all rising operators and is in one-to-one correspondence with S^2.
[3]All states with magnetic quantum number equal to one (resp. zero) in some direction are on a unique orbit.
[4]Note that the case $2j = n$ corresponds to a nonequatorial n-gon.

Free-particle-like formulation of Newtonian instantaneous action-at-a-distance electrodynamics

D. Hirondel

Physics Department, University of Delaware, Newark, Delaware
(Received 13 August 1973)

From the infinite order equations of motion of conventional electrodynamics, one can extract by order depression a subclass of second order equations of motion parametrized only by initial positions and velocities. This article presents, with a view toward possible later quantization, a canonical formulation of this electrodynamics. It happens to have the same aspect as for free particles: $H = (m_1^2+\mathbf{p}_1^2)^{1/2} + (m_2^2+\mathbf{p}_2^2)^{1/2}$. The $\mathbf{p}$'s are constant, and the canonical variables q's describe straight lines (particle positions cannot be canonical). The extension to the many-body problem is given.

1. INTRODUCTION

In conventional electrodynamics, the scheme of interaction is the following:

particle 1 ⟷ field ⟷ particle 2.

The field produced by a particle is obtained as solution of Maxwell's equations; the force a particle undergoes is obtained from the Lorentz force. The advanced and retarded solutions lead to difference differential equations of motion of interacting charges, that is to say, after proceeding to a Taylor expansion about the present time t, to equations of motion of infinite order.[1] This infinite number of degrees of freedom left the two-body problem unsolved, and not even well formulated, and led to important difficulties in field theory, such as infinite self-energy, owing to the necessity of having a charge interact with itself on the same basis as interaction with other charges.

The framework of the present article, "Newtonian instantaneous action-at-a-distance electrodynamics," is free of these difficulties. There is action-at-a-distance because the mediating fields are eliminated, and one concentrates directly on the particle orbits. The scheme of the interaction is: particle 1 ⟷ particle 2. The interaction is instantaneous and Newtonian in the sense that the acceleration of one particle is a function of only velocities and relative positions of all other particles (and no derivatives of order higher than second) evaluated at the present time t, not a retarded or advanced time. The Newtonian order of the equations of motion does not here signify Galilean covariance: Lorentz covariance is in fact maintained, meaning that the equations of motion in one Lorentz frame, expressing accelerations as functions of positions and velocities, will look, in another Lorentz frame, just the same, with the same functions of positions and velocities.

Kerner[2] showed formally that it was possible to obtain such an electrodynamics, by starting from the infinite order equations of motion of conventional electrodynamics, and depressing the order from infinite to second.

The main topic covered here is a canonical formulation of this Newtonian instantaneous electrodynamics. The reason for seeking such a formulation is to prepare the ground for quantization. There is an important theorem whose conclusion we must bear in mind before undertaking any canonical formulation: it is the zero-interaction theorem.[3] It states that, in a Hamiltonian dynamics giving invariant world lines in which (a) the inhomogeneous Lorentz group is canonically represented and (b) physical particle positions are taken to be canonical coordinates, then only free particle motions are possible.

However, this is not an obstacle to Hamiltonizing an interaction situation; rather it is a guide. It only means that we can try to keep particle positions as canonical variables: $\mathbf{q}_i = \mathbf{r}_i$. Then, Lorentz transformations cannot be canonically represented. But it is easily seen that the two straight line approximations of the electromagnetic forces cannot stem from a common Lagrangian, expanded up to order ϵ, because each of them contains a different kind of square root: $[\mathbf{r}^2 - (\mathbf{r}\times\mathbf{v}_2)^2]^{-3/2}$ for particle 1, $[\mathbf{r}^2 - (\mathbf{r}\times\mathbf{v}_1)^2]^{-3/2}$ for particle 2.

Thus, this leads to consideration of the second alternative: We do not require that $\mathbf{q}_i = \mathbf{r}_i$, and we look for some $\mathbf{q}_i(\mathbf{r}_i, \mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$. Lorentz transformations may or may not be canonically represented.

It happens that the particular set of coordinates we end up with presently is such that $\mathbf{p}_1$ and $\mathbf{p}_2$ are conserved, $\mathbf{q}_1$ and $\mathbf{q}_2$ describe two straight lines, and $H = H_1 + H_2 = (m_1^2 + \mathbf{p}_1^2)^{1/2} + (m_2^2 + \mathbf{p}_2^2)^{1/2}$. In other words, the problem is given a free particle aspect.

We will sketch briefly how, working first at the level of the straight line approximation, we discovered such a set, finding first a common action principle using 3-vectors and a single time t, and then realizing it is more powerful to look for some private action principles, one for each particle, using 4-vectors and two independent proper times. This helped to establish the method.

Then, we will give the general result for the complete order-reduced electrodynamics: We construct some conserved energy—momentum $\mathfrak{P}_i$ for each particle, and some 4-vectorial canonical coordinate $\mathfrak{Q}_i$ running in straight line motion; then the $\mathbf{q}$'s and p's are easily obtained from them: upon extracting the timelike part $\mathbf{K}_i$ from the 4-tensor $\mathfrak{Q}_i \times \mathfrak{P}_i$, the 3-vector $\mathbf{q}_i$ is given by $\mathbf{q}_i = (\mathbf{K}_i + \mathbf{p}_i t)/H_i$, which guarantees the satisfaction of the canonical equations of motion.

As the canonical variables are found particle by particle, the scheme does not depend on the number of charged particles, and can be applied to more than two charges, provided that we know how to compute the accelerations in the case of more than two charges; the procedure to find these accelerations is devised.

 Copyright © 1974 American Institute of Physics

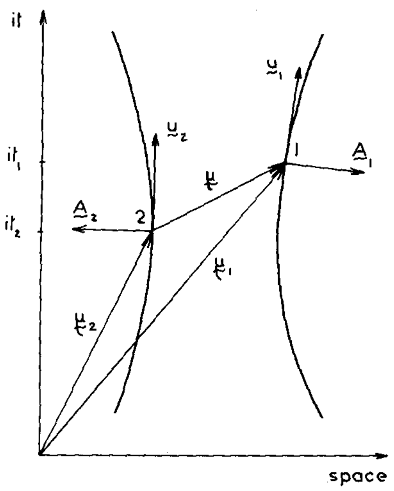

FIG. 1. Notations.

Notations and conventions

The velocity of light c is taken to be 1, unless otherwise specified. $\epsilon \equiv e_1 e_2$ is the product of the two charges e_1 and e_2. A lower index to a quantity Q refers to the label of the particle. An upper index refers to an expansion in powers of ϵ: $Q_i = \sum_{n=0}^{\infty} \epsilon^n Q_i^n$.

A and B being two 4-vectors, $\mathbf{A}\times\mathbf{B}$ will be the 4-tensor whose components are $(\mathbf{A}\times\mathbf{B})_{ij} = (\mathbf{A})_i(\mathbf{B})_j - (\mathbf{A})_j(\mathbf{B})_i$. The 3-vector angular momentum will be denoted by $\mathbf{L}$, and $\mathfrak{L}$ will be the 4-tensor formed by $\mathbf{L}$ and $\mathbf{K}$, the barycentric momentum. $(\mathbf{A})_S$ will be the 3-vector made of the first three components of $\mathbf{A}$, and $(\mathbf{A})_4$ the fourth component: $\mathbf{A} = [(\mathbf{A})_S, (\mathbf{A})_4]$.

We will use the following 4-vectors:

$$\mu_j \equiv (\mathbf{r}_j, it_j), \quad \mu_{ij} \equiv \mu_i - \mu_j, \quad \mu \equiv \mu_{12}, \quad d\tau_j = (1 - \mathbf{v}_j^2)^{1/2} dt_j,$$

$$\frac{d\mu_j}{d\tau_j} = \mathbf{u}_j = \frac{(\mathbf{v}_j, i)}{(1 - \mathbf{v}_j^2)^{1/2}}, \quad \frac{d\mathbf{u}_j}{d\tau_j} = \mathbf{A}_j,$$

and ∂_μ, for example, will mean $\partial/\partial\mu$. (see Fig. 1.)

$R_i(\xi)$ is a displacement operator which shifts μ_i into $\mu_i + \xi \mathbf{u}_i$ in everything that follows it. If the operand itself contains an operation such as $\partial_{\mathbf{u}_i}$, the shift should be done last.

Review of several basic results

From conventional electrodynamics, and in the case of retarded interaction for example, we have

$$\mathbf{a}_1(t) = \dot{\mathbf{v}}_1(t) = \mathbf{f}_1[\mathbf{r}_1(t) - \mathbf{r}_2^{\text{ret}}, \mathbf{v}_1(t), \mathbf{v}_2^{\text{ret}}],$$

or, after a Taylor expansion about the present time

$$\mathbf{a}_1(t) = \mathbf{g}_1\left[\mathbf{r}(t), \mathbf{v}_1(t), \left(\frac{d}{dt}\right)^i \mathbf{v}_2(t)\right], \quad \text{all } i = 0, 1, 2 \cdots.$$

The essence of Kerner's order reduction process[2] is to compute the time derivatives of order higher than second from the equations of motion, using power series in $\epsilon \equiv e_1 e_2$. If $\mathbf{a}_i = \sum_{n=1}^{\infty} \epsilon^n \mathbf{a}_i^n(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$, then

$$\dot{\mathbf{a}}_i = \Delta \mathbf{a}_i, \quad \ddot{\mathbf{a}}_i = \Delta \dot{\mathbf{a}}_i \cdots,$$

$$\Delta \equiv (\mathbf{v}_1 - \mathbf{v}_2)\cdot\partial_{\mathbf{r}} + \left(\sum_{n=1}^{\infty} \epsilon^n \mathbf{a}_1^n\right)\cdot\partial_{\mathbf{v}_1} + \left(\sum_{n=1}^{\infty} \epsilon^n \mathbf{a}_2^n\right)\cdot\partial_{\mathbf{v}_2}.$$

$\mathbf{a}_i^1$ is the so-called straight line approximation; it is enough to know the two $\mathbf{a}_i^1$ to start the whole procedure (the question of convergence of the series is unanswered at present).

This electrodynamics satisfies the Currie–Hill conditions.[4] These are obtained by asking quite generally that a dynamics $\ddot{\mathbf{r}}_i = \mathbf{a}_i(\mathbf{r}, \dot{\mathbf{r}}_1, \dot{\mathbf{r}}_2)$ keeps the same form after a Lorentz transformation, namely that $\mathbf{a}_1$, $\mathbf{a}_2$ are the same functions as before transformation. This functional invariance puts all frames on the same footing; none is privileged.

The electromagnetic accelerations satisfy these conditions because, starting from the advanced or retarded point which is a Lorentz invariant point on the trajectory, one can make a Taylor expansion about any time considered as the present time. Thus the dynamics has the same form in any frame.

For a class of solutions to Currie–Hill conditions, which, as will be shown later, includes electromagnetism, it is possible to make use of 4-vectors. Then, the 4-accelerations $\mathbf{A}_i$ satisfy Wray's equations,[5] the only ones used here.

To obtain them, consider one point on each world trajectory, 1 and 2. There is a time axis for which these two points are simultaneous. (See Fig. 2.) Making a Lorentz transformation amounts to taking another time axis, which means, for example, keeping particle 1 fixed and moving 2 to 2′. The arguments of $\mathbf{A}(\mu, \mathbf{u}_1, \mathbf{u}_2)$ are shifted, but $\mathbf{A}_1$ should not vary as it is $(d\mathbf{u}_1/d\tau_1)$-related to the shape of trajectory of particle 1:

$$(\mu_2 \cdot \partial_{\mu_2} + \mathbf{A}_2 \cdot \partial_{\mu_2})\mathbf{A}_1(\mu, \mathbf{u}_1, \mathbf{u}_2) = 0.$$

We will also write this as $\partial\mathbf{A}_1/\partial\tau_2 = 0$, or $\partial_2\mathbf{A}_1 = 0$. Similarly, $\partial_1\mathbf{A}_2 = 0$. We see that the two particles are shifted independently, in other words, that their proper times are considered independent.

Hill gave integro-differential equations[4] for the Currie–Hill conditions. Instead of these, we will integrate the manifestly covariant equations as follows:

$$\mathbf{A}_1(\mu, \mathbf{u}_1, \mathbf{u}_2) = \lambda_1 \mathbf{A}_1^{\text{ret}}(\mu, \mathbf{u}_1, \mathbf{u}_2) + (1 - \lambda_1)\mathbf{A}_1^{\text{adv}}(\mu, \mathbf{u}_1, \mathbf{u}_2),$$

$$\mathbf{A}_1^{\text{ret}}(\mu, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{A}_1^*[\mu + \zeta_2\mathbf{u}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{A}_2(\mu + \zeta_2\mathbf{u}_2, \mathbf{u}_1, \mathbf{u}_2)]$$

$$- \int_0^{\zeta_2} d\zeta R_2(-\zeta)\mathbf{A}_2(\mu, \mathbf{u}_1, \mathbf{u}_2)\cdot\partial_{\mathbf{u}_2}\mathbf{A}_1^{\text{ret}}(\mu, \mathbf{u}_1, \mathbf{u}_2)$$

where

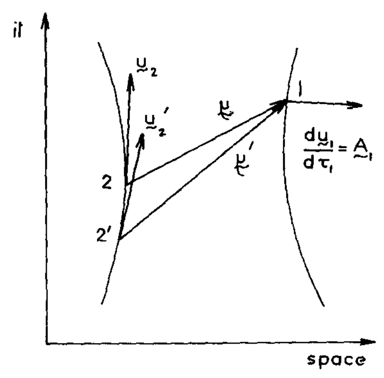

FIG. 2. Lorentz shift in Wray's equations.

$$\zeta_2 = \mu \cdot u_2 + [\mu^2 + (\mu \cdot u_2)^2]^{1/2}$$

and where the functional A_1^* is the same as in the Liénard—Wiechert formulas of conventional electrodynamics:

$$\frac{du_1^{\rm ret}}{d\tau_1} = A_1^*[\mu_1 - \mu_2^{\rm ret}, u_1, u_2^{\rm ret}, A_2(\text{evaluated at } \mu_2^{\rm ret})].$$

A similar relation holds for $A_1^{*\rm adv}$ (within changing the sign of the square root in ζ_2) but, in each case, it is the *full* A_2 [namely $\lambda_2 A_2^{\rm ret} + (1-\lambda_2)A_2^{\rm adv}$] which comes into play in the integrand and the last argument of A_1^*. It is enough to apply $u_2 \cdot \partial_\mu$ to check that A_1 satisfies the manifestly covariant equations.

Note that when μ is already in the retarded position (then $\mu^2 = 0$ and $\mu \cdot u_2 = -|\mu \cdot u_2|$) the upper limit of integration ζ_2 in $A_1^{\rm ret}$ becomes zero, and what remains is A_1^* evaluated for the retarded value of its arguments. Thus, the boundary conditions of the Liénard—Wiechert formulas are satisfied (I thank L. Bel for helping me precise this point). Ordering by powers of ϵ: $A_i^1 = \epsilon A_i^1 + \epsilon^2 A_i^2 + \cdots$ allows to find everything knowing only the two straight line approximations.

While $du_1/d\tau_1 = A_1$ is attached to point μ_1, a conserved quantity S is attached to no point in particular, and is characteristic of the world lines in their totality:

$$\partial_1 S = 0, \quad \partial_2 S = 0,$$

where S can be a 4-scalar, a 4-vector, a 4-tensor, etc.

2. THE CANONICAL VARIABLES UP TO ORDER ϵ

Computations with 3-vectors and a single time t

The method employed as a starting point involves the Lie—Koenigs theorem,[6] followed by a solution to a Pffaf's problem. Let us present that theorem for two particles labelled 1 and 2, whose positions are $\mathbf{r}_1, \mathbf{r}_2$, velocities $\mathbf{v}_1, \mathbf{v}_2$ and accelerations $\dot{\mathbf{v}}_1 = \mathbf{a}_1(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$. The Lie—Koenigs formulation considers a vector with twelve components: $\mathbf{r}_1, \mathbf{r}_2, \mathbf{v}_1, \mathbf{v}_2$; it is not known yet that $\dot{\mathbf{r}}_i = \mathbf{v}_i$, $i = 1, 2$. It also supposes that the integrand of the variational principle is linear in the time derivative of this vector; the linearity ensures a family of first order equations of motion:

$$\delta \int Z\, dt = \delta \int (\mathbf{U}_1 \cdot \dot{\mathbf{r}}_1 + \mathbf{U}_2 \cdot \dot{\mathbf{r}}_2 + \mathbf{V}_1 \cdot \dot{\mathbf{v}}_1 + \mathbf{V}_2 \cdot \dot{\mathbf{v}}_2 - H)\, dt = 0$$

where $\mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2$ and H are functions of $\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2$. The equations of motion

$$\dot{\mathbf{U}}_i = \partial_{\mathbf{r}_i} Z(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2, \dot{\mathbf{r}}_1, \dot{\mathbf{r}}_2, \dot{\mathbf{v}}_1, \dot{\mathbf{v}}_2),$$

$$\dot{\mathbf{V}}_i = \partial_{\mathbf{v}_i} Z \quad (\text{we have } \partial_{\mathbf{v}_1} \dot{\mathbf{r}}_1 \equiv 0)$$

should imply $\dot{\mathbf{r}}_i = \mathbf{v}_i$, $\dot{\mathbf{v}}_i = \mathbf{a}_i(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$, with prescribed $\mathbf{a}_i$.

The search for Z is greatly facilitated if we already know the constants of the motion: $\mathbf{P} = \mathbf{U}_1 + \mathbf{U}_2$, $\mathbf{L} = \mathbf{r}_1 \times \mathbf{U}_1 + \mathbf{r}_2 \times \mathbf{U}_2 + \mathbf{v}_1 \times \mathbf{V}_1 + \mathbf{v}_2 \times \mathbf{V}_2$, and H (linear momentum, angular momentum, and energy).

The canonical variables are obtained by solving Pfaff's problem, that is to say by seeking functions $\mathbf{p}_i(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ and $\mathbf{q}_i(\mathbf{r}_i, \mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$, $i = 1, 2$, such that

$$\mathbf{U}_1 \cdot d\mathbf{r}_1 + \mathbf{U}_2 \cdot d\mathbf{r}_2 + \mathbf{V}_1 \cdot d\mathbf{v}_1 + \mathbf{V}_2 \cdot d\mathbf{v}_2 = \mathbf{p}_1 \cdot d\mathbf{q}_1 + \mathbf{p}_2 \cdot d\mathbf{q}_2,$$

within, possibly, an exact differential.

Concerning our case, we wish to have a Lie—Koenigs action principle whose equations of motion will yield the straight line approximation of the accelerations: $\dot{\mathbf{v}}_i = \epsilon \mathbf{a}_i^1(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$.

For free particles, we would have

$$\mathbf{U}_i = \mathbf{U}_i^0 = \frac{m_i \mathbf{v}_i}{(1-\mathbf{v}_i^2)^{1/2}}, \quad \mathbf{V}_i = \mathbf{V}_i^0 = 0,$$

$$H = H^0 = \frac{m_1}{(1-\mathbf{v}_1^2)^{1/2}} + \frac{m_2}{(1-\mathbf{v}_2^2)^{1/2}}.$$

Let us then expand all quantities and the time differentiation operator in powers of ϵ:

$$\mathbf{U}_i = \mathbf{U}_i^0 + \epsilon \mathbf{U}_i^1, \quad \mathbf{V}_i = \epsilon \mathbf{V}_i^1, \quad H = H^0 + \epsilon H^1,$$

$$\frac{d}{dt} = D^0 + \epsilon D^1 = (\mathbf{v}_1 - \mathbf{v}_2) \cdot \partial_{\mathbf{r}} + \epsilon(\mathbf{a}_1^1 \cdot \partial_{\mathbf{v}_1} + \mathbf{a}_2^1 \cdot \partial_{\mathbf{v}_2}).$$

The ϵ term of the equation of motion is

$$D^1 \mathbf{U}_i^0 + D^0 \mathbf{U}_i^1 = \partial_{\mathbf{r}_i}(\mathbf{U}_1^1 \cdot \dot{\mathbf{r}}_1 + \mathbf{U}_2^1 \cdot \dot{\mathbf{r}}_2 - H^1),$$

$$D^0 \mathbf{V}_i^1 = \partial_{\mathbf{v}_i}(\mathbf{U}_1^1 \cdot \dot{\mathbf{r}}_1 + \mathbf{U}_2^1 \cdot \dot{\mathbf{r}}_2 - H^1).$$

Kennedy[7] had already worked out the constants of the motion up to order ϵ: $\mathbf{P} = \mathbf{P}^0 + \epsilon \mathbf{P}^1 \cdots$. In fact, each of Kennedy's conserved quantities, for example $\mathbf{P}$, can be written as $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2 = (\mathbf{P}_1^0 + \epsilon \mathbf{P}_1^1) + (\mathbf{P}_2^0 + \epsilon \mathbf{P}_2^1)$ where $\mathbf{P}_1$ and $\mathbf{P}_2$ are separately conserved up to order ϵ: $D^1 \mathbf{P}_1^0 + D^0 \mathbf{P}_1^1 = 0$. The term $D^1 \mathbf{P}_1^0 = (\mathbf{a}_1^1 \cdot \partial_{\mathbf{v}_1})[m_1 \mathbf{v}_1/(1-\mathbf{v}_1^2)^{1/2}]$ contains only $[\mathbf{r}^2 - (\mathbf{r} \times \mathbf{v}_2)^2]^{-1/2}$ {not $[\mathbf{r}^2 - (\mathbf{r} \times \mathbf{v}_1)^2]^{-1/2}$}, which helps in choosing $\mathbf{P}_1^1$ as the part of Kennedy's $\mathbf{P}^1$ containing that square root, $\mathbf{P}_2^1$ containing the other square root for the same reason. A similar dissection occurs for the other constants: $H, \mathbf{L}, \mathbf{K}$.

Upon noticing that $\mathbf{P}_1^1 \cdot \mathbf{v}_1 - H_1^1 = 0$, $\mathbf{P}_2^1 \cdot \mathbf{v}_2 - H_2^1 = 0$, one sees immediately that $D^1 \mathbf{U}_i^0 + D^0 \mathbf{U}_i^1 = \partial_{\mathbf{r}_i}(\mathbf{U}_1^1 \cdot \dot{\mathbf{r}}_1 + \mathbf{U}_2^1 \cdot \dot{\mathbf{r}}_2 - H^1)$ is satisfied with $\mathbf{U}_i^1 = \mathbf{P}_i^1$.

The remainder of the equations of motion becomes $D^0 \mathbf{V}_i^1 = \partial_{\mathbf{v}_i}(\mathbf{P}_1^1 \cdot \dot{\mathbf{r}}_1 + \mathbf{P}_2^1 \cdot \dot{\mathbf{r}}_2 - H^1) = -\mathbf{P}_i^1$. But we also have

$$\mathbf{v}_1 \times \mathbf{V}_1^1 + \mathbf{v}_2 \times \mathbf{V}_2^1 = \mathbf{L} - \mathbf{r}_1 \times \mathbf{U}_1^1 - \mathbf{r}_2 \times \mathbf{U}_2^1$$
$$= (\mathbf{L}_1^1 - \mathbf{r}_1 \times \mathbf{P}_1^1) + (\mathbf{L}_2^1 - \mathbf{r}_2 \times \mathbf{P}_2^1)$$

which, because of the square roots, we dissect into $\mathbf{v}_i \times \mathbf{V}_i^1 = \mathbf{L}_i^1 - \mathbf{r}_i \times \mathbf{P}_i^1$, $i = 1, 2$. This determines $\mathbf{V}_i^1$ up to a vector colinear to $\mathbf{v}_i$. For example, $\mathbf{V}_1^1 = A_1 \mathbf{r} + B \mathbf{v}_2 + \lambda_1 \mathbf{v}_1$, where A_1 and B_1 are known, and λ_1 is unknown. Choosing λ_1 such that

$$\mathbf{V}_1^1 = A_1\left(\mathbf{r} + \frac{(\mathbf{r} \cdot \mathbf{v}_1)}{1 - \mathbf{v}_1^2}\mathbf{v}_1\right) + B\left(\mathbf{v}_2 + \frac{\mathbf{v}_1 \cdot \mathbf{v}_2 - 1}{1 - \mathbf{v}_1^2}\mathbf{v}_1\right)$$

allows us to satisfy $D^0 \mathbf{V}_1^1 = -\mathbf{P}_1^1$. The recipe to obtain the undetermined term in $\mathbf{V}_i^1$ will be justified later.

Finally, we are now in possession of the following variational principle:

$$\delta \int (\mathbf{P}_1 \cdot \dot{\mathbf{r}}_1 + \mathbf{P}_2 \cdot \dot{\mathbf{r}}_2 + \epsilon \mathbf{V}_1^1 \cdot \dot{\mathbf{v}}_1 + \epsilon \mathbf{V}_2^1 \cdot \dot{\mathbf{v}}_2 - H)\, dt = 0.$$

To solve Pffaf's problem, we use the identity

$$\mathbf{V} \cdot d\mathbf{v} = d[(1 - \mathbf{v}^2)\mathbf{v} \cdot \mathbf{V}] + \frac{m\mathbf{v}}{(1-\mathbf{v}^2)^{1/2}} \cdot d\left(-\frac{(1-\mathbf{v}^2)^{1/2}}{m}(\mathbf{V} - \mathbf{v}\mathbf{v} \cdot \mathbf{V})\right)$$

and throw away the exact differential. This gives[8]

$$\mathbf{p}_i = \mathbf{P}_i = \frac{m_i \mathbf{v}_i}{(1-\mathbf{v}_i^2)^{1/2}} + \epsilon \mathbf{P}_i^1,$$

$$\mathbf{q}_i = \mathbf{r}_i - \frac{\epsilon}{m_i}(1-\mathbf{v}_i^2)^{1/2}(1-\mathbf{v}_i\mathbf{v}_i)\cdot \mathbf{V}_i.$$

Because of $\mathbf{P}_i^1 \cdot \mathbf{v}_i = H_i^1$, $i = 1, 2$, the Hamiltonian is $H = H_1 + H_2 = (m_1^2 + \mathbf{p}_1^2)^{1/2} + (m_2 + \mathbf{p}_2^2)^{1/2}$ up to order ϵ.

Computations with 4-vectors and two independent proper times

Actually, one sees that the common action principle breaks up into two private action principles, one for each particle:

$$\delta_1 \int (\mathbf{P}_1 \cdot \dot{\mathbf{r}}_1 + \epsilon \mathbf{V}_1^1 \cdot \dot{\mathbf{v}}_1 - H_1)\,dt = 0,$$

$$\delta_2 \int (\mathbf{P}_2 \cdot \dot{\mathbf{r}}_2 + \epsilon \mathbf{V}_2^1 \cdot \dot{\mathbf{v}}_2 - H_2)\,dt = 0,$$

where δ_i means that only the trajectory of the ith particle is varied. This is true because of the relations

$$\dot{\mathbf{P}}_i = \partial_{\mathbf{r}_i}(\mathbf{P}_i \cdot \mathbf{v}_i - H_i) = 0, \quad D^0 \mathbf{V}_i^1 = -\mathbf{P}_i \quad (i = 1 \text{ or } 2).$$

We also see that we solved Pfaff's problem separately for each

$$\mathbf{P}_i \cdot d\mathbf{r}_i + \epsilon \mathbf{V}_i^1 \cdot d\mathbf{v}_i = \mathbf{p}_i \cdot d\mathbf{q}_i \quad (i = 1 \text{ or } 2),$$

thus putting them into the form

$$\delta_i \int \mathbf{p}_i \cdot d\mathbf{q}_i - (m_i^2 + \mathbf{p}_i^2)^{1/2}\,dt = 0.$$

This suggests accentuating the separation by using two independent proper times instead of one single time, and 4-vectors instead of 3-vectors. We thus reach the idea of two private 4-vectorial Lie–Koenigs action principles:

$$\delta_i \int \mathbf{X}_i(\mu, \mathbf{u}_1, \mathbf{u}_2) \cdot d\mu_i + \mathbf{Y}_i(\mu, \mathbf{u}_1, \mathbf{u}_2) \cdot d\mathbf{u}_i = 0 \quad (i = 1 \text{ or } 2)$$

where $\mathbf{X}_i$ and $\mathbf{Y}_i$ are some 4-vectors. Their equations of motion are

$$\partial_i \mathbf{X}_i = \partial_{\mu_i} S_i, \quad \partial_i \mathbf{Y}_i = \partial_{\mathbf{u}_i} S_i$$

with

$$S_i\left(\mu, \mathbf{u}_1, \mathbf{u}_2, \frac{d\mu_i}{d\tau_i}, \frac{d\mathbf{u}_i}{d\tau_i}\right) = \mathbf{X}_i \cdot \frac{d\mu_i}{d\tau_i} + \mathbf{Y}_i \cdot \frac{d\mathbf{u}_i}{d\tau_i}, \quad i = 1 \text{ or } 2.$$

A close connection is suspected between $\mathbf{P}_1 \cdot d\mathbf{r}_1 + \epsilon \mathbf{V}_1^1 \cdot d\mathbf{v}_1 - H_1\,dt$ and $\mathbf{X}_1 \cdot d\mu_1 + \mathbf{Y}_1 \cdot d\mathbf{u}_1$. This is confirmed by the fact that there exists a 4-vector $\mathbf{E}_1(\mu, \mathbf{u}_1, \mathbf{u}_2)$ such that, when making $t_1 = t_2$ in it $[\mu \to (\mathbf{r}, 0); \mathbf{u}_j \to (\mathbf{v}_j, i)(1 - \mathbf{v}_j^2)^{-1/2}]$, it becomes equal to $(\mathbf{P}_1, iH_1)$. It has to satisfy $(\mathbf{u}_1 \cdot \partial_\mu + \epsilon \mathbf{A}_1^1 \cdot \partial_{\mathbf{u}_1})\mathbf{E}_1 = 0$ which can be integrated into

$$\mathbf{E}_1^1 = -\int^0 m_1 \mathbf{A}_1^1(\mu + \xi \mathbf{u}_1, \mathbf{u}_1, \mathbf{u}_2)\,d\xi, \quad (\mathbf{E}_1 = m_1 \mathbf{u}_1 + \epsilon \mathbf{E}_1^1)$$

where the primitive is simply evaluated at $\xi = 0$. With

$$\mathbf{A}_1^1 = \frac{\mathbf{u}_2(\mu \cdot \mathbf{u}_1) - \mu(\mathbf{u}_1 \cdot \mathbf{u}_2)}{[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{3/2}},$$

this yields

$$\mathbf{E}_1^1 = \frac{\mathbf{u}_2}{[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2}} + \frac{\mathbf{u}_1 \cdot \mathbf{u}_2}{-N^2}\left(\mathbf{u}_2 \frac{\mu \cdot \mathbf{u}_1 \mu \cdot \mathbf{u}_2 - \mu^2 \mathbf{u}_1 \cdot \mathbf{u}_2}{[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2}} - \mathbf{u}_1[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2} + \mu \frac{\mu \cdot \mathbf{u}_1 + \mu \cdot \mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2}{[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2}}\right)$$

with

$$-N^2 = \mu^2[(\mathbf{u}_1 \cdot \mathbf{u}_2)^2 - 1] - [(\mu \cdot \mathbf{u}_1)^2 + 2\mu \cdot \mathbf{u}_1 \mu \cdot \mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2 + (\mu \cdot \mathbf{u}_2)^2].$$

We have $\mathbf{E}_1^1 \cdot \mathbf{u}_1 = 0$ which means that $\mathbf{E}_1 \cdot \mathbf{E}_1 = -m_1^2$ up to order ϵ.

Similarly, there exists a 4-tensor $\mathfrak{L}_1$ equal to $\mu_1 \times \mathbf{E}_1 + \epsilon \mathbf{u}_1 \times \mathbf{Y}_1^1$ with

$$\mathbf{Y}_1^1 = -\int^0 m_1 \xi \mathbf{A}_1^1(\mu + \xi \mathbf{u}_1, \mathbf{u}_1, \mathbf{u}_2)\,d\xi = \frac{\mathbf{u}_1 \cdot \mathbf{u}_2}{N^2}\left(\mu - \frac{\mathbf{u}_1(\mu \cdot \mathbf{u}_1 + \mu \cdot \mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2) + \mathbf{u}_2(\mu \cdot \mathbf{u}_2 + \mu \cdot \mathbf{u}_1 \mathbf{u}_1 \cdot \mathbf{u}_2)}{(\mathbf{u}_1 \cdot \mathbf{u}_2)^2 - 1}\right)[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2}$$
$$- \frac{\mathbf{u}_2 + \mathbf{u}_1(\mathbf{u}_1 \cdot \mathbf{u}_2)}{[(\mathbf{u}_1 \cdot \mathbf{u}_2)^2 - 1)]^{3/2}} \ln\left|\frac{[\mu^2 + (\mu \cdot \mathbf{u}_2)^2]^{1/2}[(\mathbf{u}_1 \cdot \mathbf{u}_2)^2 - 1]^{1/2} - (\mu \cdot \mathbf{u}_1 + \mu \cdot \mathbf{u}_2 \mathbf{u}_1 \cdot \mathbf{u}_2)}{(-N^2)^{1/2}}\right|,$$

whose components, once one has made $t_1 = t_2$, are equal to Kennedy's $\mathbf{L}_1$ and $\mathbf{K}_1$.

The choice $\mathbf{X}_1^1 = \mathbf{E}_1^1$ satisfies $\partial_1 \mathbf{X}_1 = \partial_{\mu_1} S_1$ as $\partial_1 \mathbf{E}_1 = 0$ and $S_1 = (m_1 \mathbf{u}_1 + \epsilon \mathbf{E}_1^1) \cdot \mathbf{u}_1 = -m_1$. For $\mathbf{Y}_1$ let us take $\epsilon \mathbf{Y}_1^1$, $\mathbf{Y}_1^1$ being the previously computed quantity. The second half of the equations of motion $\partial_1 \mathbf{Y}_1 = \partial_{\mathbf{u}_1} S_1$ becomes, to order ϵ,

$$(\mathbf{u}_1 \cdot \partial_\mu)\mathbf{Y}_1^1 = \partial_{\mathbf{u}_1}\left(\mathbf{E}_1^1 \cdot \frac{d\mu_1}{d\tau_1}\right) = \partial_{\mathbf{u}_1}(\mathbf{E}_1^1 \cdot \mathbf{u}_1) - \mathbf{E}_1^1 = -\mathbf{E}_1^1.$$

It is easy to check that this is true with the above values for $\mathbf{E}_1^1$ and $\mathbf{Y}_1^1$

It is also easy to check that $\delta_1 \int \mathbf{P}_1 \cdot d\mathbf{r}_1 + \epsilon \mathbf{V}_1^1 \cdot d\mathbf{v}_1 - H\,dt = 0$ is obtained from: $\delta_1 \int \mathbf{E}_1 \cdot d\mu_1 + \epsilon \mathbf{Y}_1^1 \cdot d\mathbf{u}_1 = 0$ by making $t_1 = t_2 = t$ in the integrand. $\mathbf{Y}_1^1 \cdot d\mathbf{u}_1$ can be written

$$\left.\frac{\mathbf{Y}_1^1 + \mathbf{u}_1 \mathbf{u}_1 \cdot \mathbf{Y}_1^1}{(1 - \mathbf{v}_1^2)^{1/2}}\right|_s \cdot d\mathbf{v}_1$$

thus yielding

$$\mathbf{V}_1^1 = \left.\frac{\mathbf{Y}_1^1 + \mathbf{u}_1 \mathbf{u}_1 \cdot \mathbf{Y}_1^1}{(1 - \mathbf{v}_1^2)^{1/2}}\right|_s$$

which explains the recipe to obtain the undetermined part of $\mathbf{V}_1^1$ in the preceding section. The above relation says that $\mathbf{V}_1^1$ is of the form

$$\mathbf{V}_1^1 = A_1\left(\mathbf{r} + \frac{\mathbf{r} \cdot \mathbf{v}_1}{1 - \mathbf{v}_1^2}\mathbf{v}_1\right) + B_1\left(\mathbf{v}_2 + \frac{\mathbf{v}_1 \cdot \mathbf{v}_2 - 1}{1 - \mathbf{v}_1^2}\mathbf{v}_1\right).$$

It is also interesting to solve Pfaff's problem in terms of 4-vectors to yield some 4-vectorical canonical variables $\mathfrak{Q}_1(\mu_1, \mu, \mathbf{u}_1, \mathbf{u}_2)$ and $\mathfrak{P}_1(\mu, \mathbf{u}_1, \mathbf{u}_2)$. Using $\mathbf{u}_1 \cdot \mathbf{Y}_1^1 = 0$, we write $\mathbf{E}_1 \cdot d\mu_1 + \epsilon \mathbf{Y}_1^1 \cdot d\mathbf{u}_1 = (m_1 \mathbf{u}_1 + \epsilon \mathbf{E}_1^1) \cdot d\mu_1 - m_1 \mathbf{u}_1 \cdot d[(\epsilon/m_1)\mathbf{Y}_1^1] = \mathfrak{P}_1 \cdot d\mathfrak{Q}_1$ with $\mathfrak{P}_1 = \mathbf{E}_1$, $\mathfrak{Q}_1 = \mu_1 - (\epsilon/m_1)\mathbf{Y}_1^1$. These 4-vectorial canonical variables satisfy

$$\partial_1 \mathfrak{P}_1 = 0, \quad \partial_2 \mathfrak{P}_1 = 0, \quad \partial_1(m_1 \mathfrak{Q}_1) = \mathfrak{P}_1, \quad \partial_2 \mathfrak{Q}_1 = 0, \quad \mathfrak{L}_1 = \mathfrak{Q}_1 \times \mathbf{P}_1,$$

which parallels closely the relations for a free particle:

$\partial_1(m_1\mathbf{u}_1)=0,\quad \partial_2(m_1\mathbf{u}_1)=0,\quad \partial_1(m_1\mu_1)=m_1\mathbf{u}_1,\quad \partial_2\mu_1=0,$

$\mathfrak{L}_1=\mu_1\times m_1\mathbf{u}_1.$

It is easy to check that

$$\mathrm{P}_1=m_1\mathbf{u}_1+\frac{\epsilon\mathbf{u}_2}{[\mu^2+(\mu\cdot\mathbf{u}_2)^2]^{1/2}}+\epsilon\partial_\mu\Sigma_1,$$

$$\mathfrak{Q}_1=\mu_1-\frac{\epsilon}{m_1}\partial_{\mathbf{u}1}\Sigma_1,$$

$$\Sigma_1=-\frac{\mathbf{u}_1\cdot\mathbf{u}_2}{[(\mathbf{u}_1\cdot\mathbf{u}_2)^2-1]^{1/2}}\times\ln\left|\frac{[\mu^2+(\mu\cdot\mathbf{u}_2)^2]^{1/2}[(\mathbf{u}_1\cdot\mathbf{u}_2)^2-1]^{1/2}+\mu\cdot\mathbf{u}_1+\mu\cdot\mathbf{u}_2\mathbf{u}_1\cdot\mathbf{u}_2}{(-N^2)^{1/2}}\right|,$$

which shows that $\mathfrak{Q}_1$ and $\mathfrak{P}_1$ are related to μ_1 and $m_1\mathbf{u}_1+\epsilon\mathbf{u}_2/[\mu^2+(\mu\cdot\mathbf{u}_2)^2]^{1/2}$ by a canonical transformation, but we may insist on saying that this is only a private canonical transformation; a different generating function Σ_2 would be used for particle 2, so that, as shown previously, it is not possible to take particle positions as canonical variables for the whole problem.

At last, it is easy to see that $\mathbf{K}_1=\mathbf{q}_1H_1-\mathbf{p}_1t$ is true up to order ϵ. Actually, this constitutes the shortest method to obtain the canonical variables: (a) compute $\mathbf{E}_1$. Then write $\mathbf{E}_{1,\,t_1=t_2}=(\mathbf{p}_1,iH_1)$; $\mathbf{E}_1\cdot\mathbf{E}_1=-m_1^2$ implies $H_1=(m_1^2+\mathbf{p}_1^2)^{1/2}$; (b) compute $\mathfrak{L}_1$. Extract $\mathbf{K}_1$ by $(\mathbf{K}_1)_j=i(\mathfrak{L}_1)_{4j,\,t_1=t_2}$. Then $\mathbf{q}_1=(\mathbf{K}_1+\mathbf{p}_1t)/H_1$. Note that in this improved method, we no longer have recourse to any Lie—Koenigs formulation.

3. THE CANONICAL VARIABLES TO ALL ORDERS IN ϵ

Extending what we have discovered at the order ϵ, we are going to construct one conserved energy-momentum 4-vector per particle; for example

$$\partial_1\mathbf{E}_1=(\mathbf{u}_1\cdot\partial_\mu)\mathbf{E}_1+(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1=0,$$

$$\partial_2\mathbf{E}_1=-(\mathbf{u}_2\cdot\partial_\mu)\mathbf{E}_1+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1=0.$$

The first condition of conservation can be integrated as follows:

$$\mathbf{E}_1=m_1\mathbf{u}_1-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1\,,$$

provided that $(\mathbf{u}_1\cdot\partial_\mu)\xi_0(\mu,\mathbf{u}_1,\mathbf{u}_2)=-1$ as can be checked by applying $\mathbf{u}_1\cdot\partial_\mu$. Writing $\mathbf{E}_1=m_1\mathbf{u}_1+\sum_{n=1}^\infty\epsilon^n\mathbf{E}_1^n$ allows one to compute $\mathbf{E}_1$ term by term. A direct consequence of this expression is that

$$\mathbf{E}_1(\mu,\mathbf{u}_1,\mathbf{u}_2)=m_1\mathbf{u}_1+\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{u}_1\cdot\partial_\mu)\mathbf{E}_1$$
$$=m_1\mathbf{u}_1+\mathbf{E}_1(\mu,\mathbf{u}_1,\mathbf{u}_2)-\mathbf{E}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2),$$

or $\mathbf{E}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=m_1\mathbf{u}_1$. This result is used while proving that the second condition of conservation $\partial_2\mathbf{E}_1=0$ is satisfied. We will also use

$$\partial_1\mathbf{E}_1=(\mathbf{u}_1\cdot\partial_\mu)\mathbf{E}_1+(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1=0,$$

$$\partial_2\mathbf{A}_1=-(\mathbf{u}_2\cdot\partial_\mu)\mathbf{A}_1+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{A}_1=0,$$

$$\partial_1\mathbf{A}_2=(\mathbf{u}_1\cdot\partial_\mu)\mathbf{A}_2+(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{A}_2=0,$$

and impose $(\mathbf{u}_2\cdot\partial_\mu)\xi_0=0$. Apply $-\mathbf{u}_2\cdot\partial_\mu$:

$$-(\mathbf{u}_2\cdot\partial_\mu)\mathbf{E}_1=\int_{\xi_0}^0 d\xi R_1(\xi)\{[(\mathbf{u}_2\cdot\partial_\mu)\mathbf{A}_1]\cdot\partial_{\mathbf{u}_1}\mathbf{E}_1+(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})[(\mathbf{u}_2\cdot\partial_\mu)\mathbf{E}_1]\}.$$

But

$$\begin{aligned}[(\mathbf{u}_2\cdot\partial_\mu)\mathbf{A}_1]\cdot\partial_{\mathbf{u}_1}\mathbf{E}_1
&=[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{A}_1]\cdot\partial_{\mathbf{u}_1}\mathbf{E}_1\\
&=(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1]-\mathbf{A}_2\cdot(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\partial_{\mathbf{u}_2}\mathbf{E}_1\\
&=-(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{u}_1\cdot\partial_\mu)\mathbf{E}_1]-(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1]\\
&\quad+[(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{E}_1\\
&=-(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{u}_1\cdot\partial_\mu)\mathbf{E}_1]-(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1]\\
&\quad-[(\mathbf{u}_1\cdot\partial_\mu)\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{E}_1\\
&=-(\mathbf{u}_1\cdot\partial_\mu)[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1]-(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1],\end{aligned}$$

giving

$$\begin{aligned}-(\mathbf{u}_2\cdot\partial_\mu)\mathbf{E}_1
&=-[R_1(\xi)(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1]_{\xi=\xi_0}^{\xi=0}\\
&\quad-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})[(-\mathbf{u}_2\cdot\partial_\mu+\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{E}_1].\end{aligned}$$

Also $\mathbf{E}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=m_1\mathbf{u}_1$, thus leaving

$$\partial_2\mathbf{E}_1=-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})(\partial_2\mathbf{E}_1).$$

$\partial_2\mathbf{E}_1$ is zero to lowest order in ϵ. As $\mathbf{A}_1$ contains at least one ϵ, ordering this equation by powers of ϵ yields zero for all successive terms of $\partial_2\mathbf{E}_1$; thus $\partial_2\mathbf{E}_1=0$ to all orders. More generally, each time we will have an integro-differential equation with a straight line approximation term which is zero or constant, the quantity satisfying it will be zero or constant.

Next, we prove that $(\mathbf{E}_1)^2=-m_1^2$:

$$(\mathbf{E}_1)^2=-m_1^2-2m_1\mathbf{u}_1\cdot\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1+\Big[\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1\Big]^2.$$

Define

$$\mathbf{W}(\mu,\mathbf{u}_1,\mathbf{u}_2)\equiv\int_{\xi_0}^0 d\xi' R_1(\xi')(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1.$$

We need to change μ into $\mu+\xi\mu_1$ everywhere in this relation, including ξ_0. The solution of $(\mathbf{u}_1\cdot\partial_\mu)\xi_0=-1$, $(\mathbf{u}_2\cdot\partial_\mu)\xi_0=0$ is

$$\xi_0=-\frac{\mu\cdot\mathbf{u}_1+\mu\cdot\mathbf{u}_2\mathbf{u}_1\cdot\mathbf{u}_2}{(\mathbf{u}_1\cdot\mathbf{u}_2)^2-1}+g[-N^2,\mathbf{u}_1\cdot\mathbf{u}_2];$$

thus

$$\xi_0(\mu+\xi\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=\xi_0(\mu,\mathbf{u}_1,\mathbf{u}_2)-\xi,$$

and

$$\mathbf{W}(\xi)\equiv\mathbf{W}(\mu+\xi\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=\int_{\xi_0-\xi}^0 d\xi' R_1(\xi+\xi')(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1.$$

Note that $\mathbf{W}(\xi)=m_1\mathbf{u}_1-\mathbf{E}_1(\mu+\xi\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)$ and, consequently, $\mathbf{W}(\xi_0)=0$. Differentiating

$$\begin{aligned}\frac{d\mathbf{W}(\xi)}{d\xi}&=R_1(\xi+\xi')(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}\Big|_{\xi'=\xi_0-\xi}\\
&\quad+\int_{\xi_0-\xi}^0 d\xi'\frac{d}{d\xi}[R_1(\xi+\xi')(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1]\\
&=R_1(\xi_0)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1\\
&\quad+\int_{\xi_0-\xi}^0 d\xi'\frac{d}{d\xi'}[R_1(\xi+\xi')(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1]\\
&=R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1.\end{aligned}$$

Then

$$\mathbf{W}^2(0)-\mathbf{W}^2(\xi_0)=2\int_{\xi=\xi_0}^{\xi=0} d\xi\,\mathbf{W}(\xi)\cdot\frac{d\mathbf{W}(\xi)}{d\xi},$$

or

$$[\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1]^2$$
$$=2\int_{\xi_0}^0 d\xi[m_1\mathbf{u}_1-\mathbf{E}_1(\mu+\xi\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)]\cdot R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1$$

which shows that $(\mathbf{E}_1)^2=-m_1^2-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})(\mathbf{E}_1)^2$, and thus that $(\mathbf{E}_1)^2=-m_1^2$ to all orders.

We choose to take $\mathfrak{P}_1=\mathbf{E}_1$ and $\mathbf{E}_{1,\,t_1=t_2}=(\mathbf{P}_1,iH_1)=(\mathbf{p}_1,iH_1)$. The above relation gives the problem a free particle aspect:

$$H_1=(m_1^2+\mathbf{p}_1^2)^{1/2},\quad H=H_1+H_2.$$

Now, let us turn to $\mathfrak{Q}_1$. Write it as $\mu_1-\mathbf{Y}_1/m_1$, $\mathbf{Y}_1$ being solution of the following integro-differential equation

$$\mathbf{Y}_1=\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})(\xi\mathbf{E}_1+\mathbf{Y}_1),$$

with the same ξ_0. Applying $\mathbf{u}_1\cdot\partial_\mu$ gives $\partial_1\mathbf{Y}_1=\int_{\xi_0}^0 d\xi R_1(\xi)\times(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathbf{E}_1=m_1\mathbf{u}_1-\mathbf{E}_1$ and, consequently $\partial_1(m_1\mathfrak{Q}_1)=\mathbf{E}_1=\mathfrak{P}_1$. Thus $\mathfrak{Q}_1$ describes a straight line.

It is easy to show that $\partial_2\mathbf{Y}_1=0$ the same way as we showed that $\partial_2\mathbf{E}_1=0$. This implies $\partial_2\mathfrak{Q}_1=0$. This relation guarantees manifest world line invariance for the $\mathfrak{Q}$'s, though there is no such thing for the $\mathbf{q}$'s; this is one of the advantages of the $\mathfrak{Q}$'s over the $\mathbf{q}$'s.

By the same method as for $\mathbf{E}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=m_1\mathbf{u}_1$, one can show that $\mathbf{Y}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=0$; this is also obvious from

$$\xi_0(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)=\xi_0-\xi_0=0$$

which makes the lower limit of integration equal to the upper limit in $\mathbf{Y}_1(\mu+\xi_0\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2)$. Thus $\mathfrak{Q}_1(\mu_1+\xi_0\mathbf{u}_1,\mu_2,\mathbf{u}_1,\mathbf{u}_2)=\mu_1$. This allows the following proof:

$$\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})(\mathfrak{Q}_1\times\mathfrak{P}_1)$$
$$=\int_{\xi_0}^0 d\xi R_1(\xi)\left[\left(\frac{\mathfrak{P}_1}{m_1}-(\mathbf{u}_1\cdot\partial_\mu)\mathfrak{Q}_1\right)\times\mathfrak{P}_1+\mathfrak{Q}_1\times[-(\mathbf{u}_1\cdot\partial_\mu)\mathfrak{P}_1]\right]$$
$$=-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{u}_1\cdot\partial_\mu)(\mathfrak{Q}_1\times\mathfrak{P}_1)$$
$$=-R_1(\xi)(\mathfrak{Q}_1\times\mathfrak{P}_1)\Big|_{\xi=\xi_0}^{\xi=0}$$
$$=-\mathfrak{Q}_1\times\mathfrak{P}_1+\mu_1\times m_1\mathbf{u}_1.$$

Thus $\mathfrak{Q}_1\times\mathfrak{P}_1$ satisfies an integro-differential equation which is the same in form as the one for the quantity $\mathfrak{L}_1$ that one would compute knowing its straight line approximation value

$$\mathfrak{L}_1=\mu_1\times m_1\mathbf{u}_1-\int_{\xi_0}^0 d\xi R_1(\xi)(\mathbf{A}_1\cdot\partial_{\mathbf{u}_1})\mathfrak{L}_1,\quad(\text{same } \xi_0).$$

Thus $\mathfrak{L}_1=\mathfrak{Q}_1\times\mathfrak{P}_1$.

This is useful to show that, of the ten constants $\mathbf{E}_1$, $\mathfrak{L}_1$ we can build for particle 1, only six are independent. We expect to find four relations. The first one is $\mathbf{E}_1\cdot\mathbf{E}_1=-m_1^2$. The remainder are found by considering the 4-vector $\mathbf{M}_1$ such that $(\mathbf{M}_1)_i=\epsilon_{ijkl}(\mathfrak{L}_1)_{jk}(\mathbf{E}_1)_l$. As $\mathfrak{L}_1$ is $\mathfrak{Q}_1\times\mathfrak{P}_1$, $\mathbf{M}_1$ is zero. But this is only three relation as, once the first three components are zero, the fourth is necessarily zero:

$$\mathbf{M}_1=2[i(\mathbf{P}_1\times\mathbf{K}_1+\mathbf{L}_1H_1),-\mathbf{L}_1\cdot\mathbf{P}_1].$$

One can take $\mathbf{P}_1$ and $\mathbf{K}_1$ as the six independent quantities; then

$$H_1=(m_1^2+\mathbf{P}_1^2)^{1/2},\quad \mathbf{L}_1=(\mathbf{K}_1\times\mathbf{P}_1)/H_1\quad(\mathbf{P}_1=\mathbf{p}_1).$$

Extracting $\mathbf{K}_1$ from $\mathfrak{L}_1$, we write $\mathbf{q}_1=(\mathbf{K}_1+\mathbf{p}_1t)/H_1$ (whereupon $\mathbf{L}_1=\mathbf{q}_1\times\mathbf{p}_1$). This formula guarantees that all canonical equations of motion are satisfied. $\dot{\mathbf{p}}_1=-\partial_{\mathbf{q}_1}H$ is satisfied as $\mathbf{p}_1$ is conserved and H does not depend on the $\mathbf{q}$'s.

$\dot{\mathbf{q}}_1=\partial_{\mathbf{p}_1}H$ is satisfied because

$$\dot{\mathbf{q}}_1=\frac{d}{dt}\frac{\mathbf{K}_1+\mathbf{p}_1t}{H_1}=\frac{\mathbf{p}_1}{H_1},$$

$$\partial_{\mathbf{p}_1}H=\partial_{\mathbf{p}_1}(m_1^2+\mathbf{p}_1^2)^{1/2}=\mathbf{p}_1/H_1.$$

The private Lie–Koenigs action principle is $\delta_1\times\int\mathfrak{P}_1\cdot d\mathfrak{Q}_1$. It is straightforward to show that $\mathfrak{P}_1\cdot d\mathfrak{Q}_1$ and $\mathbf{p}_1\cdot d\mathbf{q}_1-H_1dt_1$ are equal within an exact differential:

$$(\mathbf{K}_i)_j=i(\mathfrak{L}_1)_{4j}=i[(\mathfrak{Q}_1)_4(\mathfrak{P}_1)_j-(\mathfrak{Q}_1)_j(\mathfrak{P}_1)_4]$$
$$=i\left[\left(it-\frac{(\mathbf{Y}_1)_4}{m_1}\right)\mathbf{p}_1-(\mathfrak{Q}_1)_S iH_1\right]_j$$

$$\mathbf{q}_1=\frac{\mathbf{K}_1+\mathbf{p}_1t}{H_1}(\mathfrak{Q}_1)_S-\frac{\mathbf{p}_1}{H_1}\frac{i(\mathbf{Y}_1)_4}{m_1},$$

$$\mathbf{p}_1\cdot d\mathbf{q}_1=\mathbf{p}_1\cdot(\mathfrak{Q}_1)_S-\mathbf{p}_1\cdot d\left(\frac{\mathbf{p}_1}{H_1}\frac{i(\mathbf{Y}_1)_4}{m_1}\right).$$

Then, using the constraint $H_1^2-\mathbf{P}_1^2=m_1^2$ and $H_1dH_1=\mathbf{P}_1\cdot d\mathbf{P}_1$, we find

$$\mathbf{p}_1\cdot d\left(\frac{\mathbf{p}_1}{H_1}(\mathbf{Y}_1)_4\right)=\mathbf{p}_1\cdot\left(\frac{d\mathbf{p}_1}{H_1}-\frac{\mathbf{p}_1}{(H_1)^2}dH_1\right)(\mathbf{Y}_1)_4$$
$$+\frac{H_1^2-m_1^2}{H_1}d(\mathbf{Y}_1)_4$$
$$=H_1d(\mathbf{Y}_1)_4-m_1^2d\frac{(\mathbf{Y}_1)_4}{H_1}.$$

Thus

$$\mathbf{p}_1\cdot d\mathbf{q}_1-H_1dt_1=\mathbf{p}_1\cdot d(\mathfrak{Q}_1)_S+iH_1d\left(-\frac{(\mathbf{Y}_1)_4}{m_1}\right)$$
$$-H_1dt_1+d\left(m_1\frac{i(\mathbf{Y}_1)_4}{H_1}\right)$$
$$=(\mathfrak{P}_1)_S\cdot d(\mathfrak{Q}_1)_S+iH_1d\left(it_1-\frac{i(\mathbf{Y}_1)_4}{m_1}\right)$$
$$+d\left(m_1\frac{i(\mathbf{Y}_1)_4}{H_1}\right)$$
$$=\mathfrak{P}_1\cdot d\mathfrak{Q}_1+d\left(m_1\frac{i(\mathbf{Y}_1)_4}{H_1}\right).$$

Obviously

$$(\mathbf{q}_1,it_1)=\left((\mathfrak{Q}_1)_S-\frac{\mathbf{p}_1}{H_1}\frac{i(\mathbf{Y}_1)_4}{m_1},it_1\right)$$

is not a 4-vector, as it is obtained from the 4-vector

$$\mathfrak{Q}_1=\left((\mathfrak{Q}_1)_S,it_1-\frac{(\mathbf{Y}_1)_4}{m_1}\right).$$

by transferring part of its fourth component on its spatial part to obtain a fourth component containing no ϵ contribution. But (q_1,it_1) belongs to the same straight world line as $\mathfrak{Q}_1$.

With the help of $\mathbf{K}_1 = \mathbf{q}_1 H_1 - \mathbf{p}_1 t$ it is easy to verify that the q's transform like physical coordinates, namely satisfy the following Poisson brackets:

$$[\mathbf{q}_i, \mathbf{K}] = [\mathbf{q}_i, H]\mathbf{q}_i - 1t, \quad [\mathbf{q}_i, \mathbf{p}_j] = 1,$$

$$[\mathbf{q}_i, \mathbf{L}] = -1 \times \mathbf{q}_i, \quad \frac{\partial \mathbf{q}_i}{\partial t} = 0.$$

Thus the q's behave as physical coordinates and are also canonical variables: $[\mathbf{q}_i, \mathbf{q}_j] = 0$, $i = 1, 2$, $j = 1, 2$. Hill[9] has proved that in such a case, $\dot{\mathbf{q}}_i$ had to be independent of $\mathbf{q}_j$ and $\dot{\mathbf{q}}_j$ $(j \neq i)$. This is the result of the zero-interaction theorem, illustrated in q-space. There is complete agreement with what we have found in our present canonical formulation, as, precisely $\ddot{\mathbf{q}}_i = 0$.

Naturally, the charge e_i is located at the particle position $\mathbf{r}_i$, not at the canonical position $\mathbf{q}_i$. Hill[9] made some remarks related to measurement theory when the formalism is quantized in terms of canonical positions.

Note that the separation of internal and external motions is made immediate by the fact that our formulation is the same as for free particles:

$$H = (m_1^2 + \mathbf{p}_1^2)^{1/2} + (m_2^2 + \mathbf{p}_2^2)^{1/2}, \quad \mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2,$$

$$\mathbf{L} = \mathbf{q}_1 \times \mathbf{p}_1 + \mathbf{q}_2 \times \mathbf{p}_2, \quad \mathbf{K} = \mathbf{q}_1 H_1 - \mathbf{p}_1 t + \mathbf{q}_2 H_2 - \mathbf{p}_2 t,$$

and that Bakamjian and Thomas[10] worked out this separation for real free particles. They give the external variables $\mathbf{Q}$ and $\mathbf{P}$, and the internal variables $\mathbf{q}$, $\mathbf{p}$ in terms of $\mathbf{q}_i = \mathbf{r}_i$, $\mathbf{p}_i = m_i \mathbf{v}_i (1 - \mathbf{v}_i^2)^{-1/2}$. Then their result can be used directly for our problem, where now $\mathbf{q}_i$ and $\mathbf{p}_i$ are $\mathbf{q}_i(\mathbf{r}_i, \mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ and $\mathbf{p}_i(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$. Application to first order in ϵ, and at the nonrelativistic level gives

$$\mathbf{p} = m\mathbf{v} + \epsilon \frac{(\mathbf{r} \times \mathbf{v}) \times \mathbf{r}}{r(\mathbf{r} \times \mathbf{v})^2},$$

$$\mathbf{q} = \mathbf{r} + \frac{\epsilon}{m}\left[\frac{(\mathbf{r} \times \mathbf{v}) \times \mathbf{v} r}{v^2 (\mathbf{r} \times \mathbf{v})^2} + \frac{\mathbf{v}}{v^3} \ln\left(\frac{rv + \mathbf{r} \cdot \mathbf{v}}{rv - \mathbf{r} \cdot \mathbf{v}}\right)^{1/2}\right]$$

with $m^{-1} = m_1^{-1} + m_2^{-1}$, $\mathbf{v} \equiv \mathbf{v}_1 - \mathbf{v}_2$, $v \equiv |\mathbf{v}|$, $r \equiv |\mathbf{r}|$. $\mathbf{p}$ happens to be exactly conserved for the Kepler problem ($m\ddot{\mathbf{r}} = \epsilon \mathbf{r} r^{-3}$); $\mathbf{p}$ is along the minor axis of the conic. Also $\mathbf{q} \times \mathbf{p} = \mathbf{r} \times m\mathbf{v}$, and $m^{-1}\mathbf{p} \times (\mathbf{q} \times \mathbf{p}) = \mathbf{v} \times (\mathbf{r} \times m\mathbf{v}) + \epsilon \hat{r}$ which is the Runge Lenz vector. This suggests proposing the conserved $\mathbf{p} \times (\mathbf{q} \times \mathbf{p})$ as a relativistic generalization of the classical Runge Lenz vector.

To all orders, the internal Hamiltonian $h = (H^2 - \mathbf{P}^2)^{1/2}$ can be expressed as $(m_1^2 + \mathbf{p}^2)^{1/2} + (m_2^2 + \mathbf{p}^2)^{1/2}$. Thus real $\mathbf{p}$ corresponds to $h \geq m_1 + m_2$.

4. THE MANY BODY PROBLEM

We already know how to compute the accelerations for the two body problem by Hill's integro-differential equations. We are going to find how to do so for more than two charged particles in two steps: first, carrying out the process of order reduction in one dimension for three charges, then finding integro-differential equations for the accelerations in 4-vectorial form.

We start from

$$\frac{m_1 a_1}{(1 - v_1^2)^{3/2}} = e_1[(\text{electric field at 1 due to 2}) + (\text{electric field at 1 due to 3})]$$

$$= e_1\left\{e_2\left[\lambda_2\left(\frac{1}{x_{12}^2}\frac{1+v_2}{1-v_2}\right)_{\text{ret}} + (1-\lambda_2) \times \left(\frac{1}{x_{12}^2}\frac{1-v_2}{1+v_2}\right)_{\text{adv}}\right] + e_3\left[\lambda_3\left(\frac{1}{x_{13}^2}\frac{1+v_3}{1-v_3}\right)_{\text{ret}} + (1-\lambda_3)\left(\frac{1}{x_{13}^2}\frac{1-v_3}{1+v_3}\right)_{\text{adv}}\right]\right\}$$

which expresses the superposition principle in conventional electrodynamics (λ_2 and λ_3 are any numbers). x_{ij} is $x_i - x_j$, and we will suppose $x_3 < x_2 < x_1$.

The fields from particle 2 and 3 will be separately order reduced, so that, at the order reduced level, a_1 will be the sum of two contributions, one from particle 2, one from particle 3: $a_1 = a_1(2) + a_1(3)$, but each contribution will not be the one corresponding to each two body problem.

We will compute only $(1/x_{12}^2)[(1+\theta v_2)/(1-\theta v_2)]$ where $\theta = 1$ for the retarded field of 2 at 1, and $\theta = -1$ for the advanced field.

To compute $x_2(t_2)$ and $v_2(t_2)$, we make a Taylor expansion about t_1, and evaluate time derivatives of position of order higher than second from the equations known at the next lower order of approximation. (see Fig. 3.)

$$x_2(t_2) = x_2(t_1) + (t_2 - t_1)v_2(t_1) + \sum_{n=2}^{\infty}\frac{(t_2 - t_1)^n}{n!}(\partial_{t_1})^{n-2}a_2(t_1),$$

$$v_2(t_2) = v_2(t_1) + \sum_{n=1}^{\infty}\frac{(t_2 - t_1)^n}{n!}(\partial_{t_1})^{n-1}a_2(t_1),$$

with

$$a_2 = \frac{(1 - v_2^2)^{3/2}}{m_2} e_2\left(-\frac{e_1(1 - v_1^2)}{x_{12}^2} + \frac{e_3(1 - v_3^2)}{x_{23}^2}\right)$$

and

$$\partial_{t_1} = (v_1 - v_2)\partial_{x_{12}} + (v_2 - v_3)\partial_{x_{23}}.$$

The exact light cone condition $t_2 - t_1 = \theta[x_2(t_2) - x_1(t_1)]$ is approximated by $t_2 - t_1 = x_{12}/(v_2 - \theta)$.

The series are easily summed to give

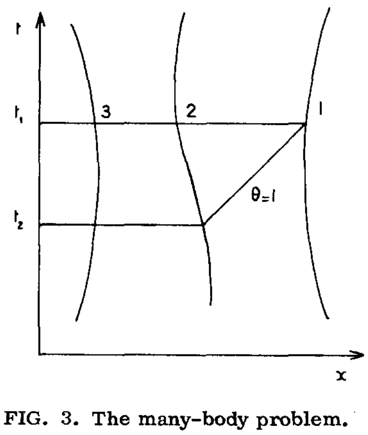

FIG. 3. The many-body problem.

$$\frac{1}{x_{12}^2}\frac{1+\theta v_2}{1-\theta v_2}=\frac{1-v_2^2}{x_{12}^2}+\frac{2e_1e_2(1-v_2^2)^{3/2}}{m_2(v_1-v_2)x_{12}^3}$$
$$\times\left((1-v_1v_2)(v_1+\theta)+\frac{(1-v_1^2)(1-v_2^2)}{v_1-v_2}\ln\frac{1-\theta v_1}{1-\theta v_2}\right)$$
$$-\frac{2e_2e_3}{m_2}\frac{(1-v_2^2)^{3/2}(1-v_3^2)}{(v_2-v_3)x_{12}^2}\left(\frac{\theta}{x_{23}}\frac{x_{23}+v_2(v_3x_{12}-v_2x_{13})}{x_{23}+\theta(v_3x_{12}-v_2x_{13})}\right.$$
$$\left.+\frac{1-v_2^2}{x_{12}(v_2-v_3)}\ln\left|\frac{x_{23}+\theta(v_3x_{12}-v_2x_{13})}{x_{23}(1-\theta v_2)}\right|\right).$$

The term in e_1e_2 is the one already found by Hill[4]; the one in e_2e_3 is new.

In four-dimensional space—time, we put, in a very natural way,

$$\mathbf{A}_1=\mathbf{A}_1(2)+\mathbf{A}_1(3),$$
$$\mathbf{A}_1(i)=\lambda_{1i}\mathbf{A}_1^{\mathrm{ret}}(i)+(1-\lambda_{1i})\mathbf{A}_1^{\mathrm{adv}}(i),$$
$$\mathbf{A}_1^{\mathrm{ret}}(i)=\mathbf{A}_1^{*\mathrm{ret}}(i)-\int_0^{\zeta_i}d\zeta R_i(-\zeta)\mathbf{A}_i\cdot\partial_{\mathbf{u}_i}\mathbf{A}_1^{\mathrm{ret}}(i),$$
$$\mathbf{A}_1^{*\mathrm{ret}}(i)=\mathbf{A}_1^*[\boldsymbol{\mu}_{1i}+\zeta_i\mathbf{u}_i,\mathbf{u}_1,\mathbf{u}_i,R_i(-\zeta_i)\mathbf{A}_i],$$
$$\zeta_i=\boldsymbol{\mu}_{1i}\cdot\mathbf{u}_i+[\boldsymbol{\mu}_{1i}^2+(\boldsymbol{\mu}_{1i}\cdot\mathbf{u}_i)^2]^{1/2}$$

[change the sign of the square root in ζ_i for $\mathbf{A}_1^{\mathrm{adv}}(i)$].

$\mathbf{A}_1(2)$ satisfies $\partial_2\mathbf{A}_1(2)=0$ by construction. To prove $\partial_3\mathbf{A}_1(2)=0$, apply $\mathbf{u}_3\cdot\partial_{\mu_3}$, and use $\partial_2\mathbf{A}_1(2)=0$, $\partial_3\mathbf{A}_2=0$, $\partial_2\mathbf{A}_3=0$; note that to prove $\partial_3\mathbf{A}_1(2)=0$ up to order ϵ^n, we need these three relations to be true only up to order ϵ^{n-1}. Let us show how this works on $\mathbf{A}_1^{\mathrm{ret}}(2)$, for example

$$(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{\mathrm{ret}}(2)=(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{*\mathrm{ret}}(2)-\int_0^{\zeta_2}d\zeta R_2(-\zeta)\{[(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{A}_1^{\mathrm{ret}}(2)+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{\mathrm{ret}}(2)\},$$
$$\begin{aligned}[(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{A}_1^{\mathrm{ret}}(2)&=-[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{A}_1^{\mathrm{ret}}(2)\\&=-(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{A}_1^{\mathrm{ret}}(2)]+\mathbf{A}_3\cdot(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\partial_{\mathbf{u}_3}\mathbf{A}_1^{\mathrm{ret}}(2)\\&=(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})[(\mathbf{u}_2\cdot\partial_{\mu_2})\mathbf{A}_1^{\mathrm{ret}}(2)]+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)]-[(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\mathbf{A}_3]\cdot\partial_{\mathbf{u}_3}\mathbf{A}_1^{\mathrm{ret}}(2)\\&=(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})[(\mathbf{u}_2\cdot\partial_{\mu_2})\mathbf{A}_1^{\mathrm{ret}}(2)]+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)]+[(\mathbf{u}_2\cdot\partial_{\mu_2})\mathbf{A}_3]\cdot\partial_{\mathbf{u}_3}\mathbf{A}_1^{\mathrm{ret}}(2)\\&=(\mathbf{u}_2\cdot\partial_{\mu_2})[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)]+(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)],\end{aligned}$$
$$R_2(-\zeta)[(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_2]\cdot\partial_{\mathbf{u}_2}\mathbf{A}_1^{\mathrm{ret}}(2)=-\frac{d}{d\zeta}[R_2(-\zeta)(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)]+R_2(-\zeta)(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})[(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)].$$

Finally,

$$(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{\mathrm{ret}}(2)=(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{*\mathrm{ret}}(2)+R_2(-\zeta)(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)\Big|_{\zeta=0}^{\zeta=\zeta_2}-\int_0^{\zeta_2}d\zeta R_2(-\zeta)(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})(\mathbf{u}_3\cdot\partial_{\mu_3}+\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_1^{\mathrm{ret}}(2)$$

or

$$\partial_3\mathbf{A}_1^{\mathrm{ret}}(2)=(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{*\mathrm{ret}}(2)+R_2(-\zeta_2)\mathbf{A}_3\cdot\partial_{\mathbf{u}_3}\mathbf{A}_1^{\mathrm{ret}}(2)-\int_0^{\zeta_2}d\zeta R_2(-\zeta)(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\partial_3\mathbf{A}_1^{\mathrm{ret}}(2).$$

Let us show that the first term of the right-hand side equals minus the second. Remembering that $\mathbf{A}_1^{*\mathrm{ret}}(2)$ is linear in $\mathbf{A}_2$,

$$\begin{aligned}\mathbf{A}_1^{*\mathrm{ret}}(2)&=\mathbf{A}_1^*[\boldsymbol{\mu}+\zeta_2\mathbf{u}_2,\mathbf{u}_1,\mathbf{u}_2,R_2(-\zeta_2)\mathbf{A}_2]\\&=R_2(-\zeta_2)\epsilon\left(\frac{\mathbf{A}_2(\boldsymbol{\mu}\cdot\mathbf{u}_1)-\boldsymbol{\mu}(\mathbf{A}_2\cdot\mathbf{u}_1)}{(\boldsymbol{\mu}\cdot\mathbf{u}_2)^2}\right.\\&\quad\left.-\frac{(1+\boldsymbol{\mu}\cdot\mathbf{A}_2)(\boldsymbol{\mu}\mathbf{u}_1\cdot\mathbf{u}_2-\mathbf{u}_2\boldsymbol{\mu}\cdot\mathbf{u}_1)}{|\boldsymbol{\mu}\cdot\mathbf{u}_2|^3}\right),\end{aligned}$$

we write, using $\partial_3\mathbf{A}_2=0$ at the next lower order,

$$\begin{aligned}(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_1^{*\mathrm{ret}}(2)&=\mathbf{A}_1^*[\cdots,R_2(-\zeta_2)(\mathbf{u}_3\cdot\partial_{\mu_3})\mathbf{A}_2]\\&=\mathbf{A}_1^*[\cdots,-R_2(-\zeta_2)(\mathbf{A}_3\cdot\partial_{\mathbf{u}_3})\mathbf{A}_2]\\&=\mathbf{A}_1^*[\cdots,-[R_2(-\zeta_2)\mathbf{A}_3]\cdot\partial_{\mathbf{u}_3}R_2(-\zeta_2)\mathbf{A}_2]\\&=-[R_2(-\zeta_2)\mathbf{A}_3]\cdot\partial_{\mathbf{u}_3}\mathbf{A}_1^{*\mathrm{ret}}(2)\\&=-[R_2(-\zeta_2)\mathbf{A}_3]\cdot\partial_{\mathbf{u}_3}[R_2(-\zeta_2)\mathbf{A}_1^{\mathrm{ret}}(2)]\\&=-R_2(-\zeta_2)[\mathbf{A}_3\cdot\partial_{\mathbf{u}_3}\mathbf{A}_1^{\mathrm{ret}}(2)].\end{aligned}$$

Thus, what remains is the following integro-differential equation:

$$\partial_3\mathbf{A}_1^{\mathrm{ret}}(2)=-\int_0^{\zeta_2}d\zeta R_2(-\zeta)(\mathbf{A}_2\cdot\partial_{\mathbf{u}_2})\partial_3\mathbf{A}_1^{\mathrm{ret}}(2),$$

whose solution is $\partial_3\mathbf{A}_1^{\mathrm{ret}}(2)=0$, since $\partial_3\mathbf{A}_1^{\mathrm{ret}}(2)=0$ at the lowest order.

Thus $\mathbf{A}_1(2)$ is such that $\partial_2\mathbf{A}_1(2)=0$, $\partial_3\mathbf{A}_1(2)=0$; $\mathbf{A}_1(3)$ satisfies the same relations, and, consequently, $\mathbf{A}_1=\mathbf{A}_1(2)+\mathbf{A}_1(3)$ satisfies them also.

To compare with the preceeding method, we start the computation of

$$-\int_0^{\zeta_2}d\zeta R_2(-\zeta)[\mathbf{A}_2^1(3)\cdot\partial_{\mathbf{u}_2}]\mathbf{A}_1^1(2)$$

in four-dimensional space—time; namely, we make the shift $\boldsymbol{\mu}_2\to\boldsymbol{\mu}_2-\zeta\mathbf{u}_2$, then consider the first three components, go to a one-dimensional space, and make $t_1=t_2$ [: $\boldsymbol{\mu}_{ij}\to(x_{ij},0)$] before doing the integration over ζ, and we find the same result as before. (In one dimension, $\mathbf{A}_1^*$ brings no contribution to second, nor to higher order.)

It is necessary to end this section on the many body problem by a remark on the domain of applicability of the manifestly covariant equation. In general, we cannot deduce them for more than four particles in our four-dimensional space—time: Write that the variation of $d\mathbf{u}_1/d\tau_1=\mathbf{A}_1(\boldsymbol{\mu}_{1i},\mathbf{u}_1,\mathbf{u}_i)$ (all $i\neq1$) is zero when one shifts the hyperplane containing all particles (they are thus instantaneous with respect to the time axis perpendicular to that hyperplane) to a slightly different hyperplane still going through particle 1, $\sum_{i\neq1}d\tau_i\partial_i\mathbf{A}_1=0$; only if $n\leqslant4$ can the $d\tau_i$ be chosen arbitrarily, implying $\partial_i\mathbf{A}_1=0$ $(i\neq1)$.

However, for n particles in electromagnetism, the straight line approximations satisfy $\partial_k\mathbf{A}_j=0$ at the lowest order:

$$(\mathbf{u}_k \cdot \partial_{\mu_k}) \sum_{i \neq j} e_i e_j \frac{\mathbf{u}_i(\mu_{ij} \cdot \mathbf{u}_j) - \mu_{ij}(\mathbf{u}_i \cdot \mathbf{u}_j)}{[\mu_{ij}^2 + (\mu_{ij} \cdot \mathbf{u}_i)^2]^{3/2}} = 0 \quad (k \neq j).$$

Then, from the preceding proof by recurrence, this implies $\partial_k \mathbf{A}_j = 0$ to all successive orders.

This allows us to apply our scheme of free-particle-like formulation of particles.

CONCLUSION

We are now in possession of a formalism giving a certain set of canonical variables such that each $\mathbf{p}_i$ is constant, and each $\mathbf{q}_i$ describes a straight line. Its value resides in several features: it is relatively simple, paralleling free-particle dynamics, is well defined and computable to any order in ϵ, and it applies to any number of particles.

On the other hand, the outstanding problem of quantizing with some $\mathbf{q}_i(\mathbf{r}_i, \mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ is totally unsolved. A new recipe has to be found. Will it be some rule applicable in any set of canonical variables?

About the last possibility, we note that, working only up to order $\epsilon, \mathbf{q}_i$ does not go to $\mathbf{r}_i$ in the nonrelativistic limit ($1/c = 0$), nor to first order in $1/c^2$ (Darwin—Breit theory); however, the new canonical coordinates obtained by making a canonical transformation with the generating function Σ_0, defined as the common part of Σ_1 and Σ_2, do go to particle positions in the same conditions (but not to order $1/c^4$); thus, this new set stays as close as possible to the set usually presumed to be within known quantization rules.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. Kerner for an interesting and fruitful interaction during the preparation of this work, and Dr. Hill for discussions. Financial support from the Physics Department of the University of Delaware is gratefully acknowledged.

*Abstract from Ph.D. thesis, "Canonical Formulation of Instantaneous Electrodynamics. Gravitation," University of Delaware, June 1973.

† Present address: 4 Rue Sédillot, Paris 7°, France.

[1]Nevertheless, under certain conditions for the electromagnetic equations, it is found that instantaneous values of positions and velocities do indeed determine the solutions uniquely [Driver, Phys. Rev. **178**, 2051 (1969)].

[2]E.H. Kerner, J. Math. Phys. **6**, 1218 (1965).

[3]D.G. Currie, T.F. Jordan, and E.C.G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); T.F. Jordan, Phys. Rev. **166**, 1308 (1968).

[4]R.N. Hill, J. Math. Phys. **8**, 201 (1967); D.G. Currie, Phys. Rev. **142**, 817 (1966).

[5]J.G. Wray, Phys. Rev. D **1**, 2212 (1970). Since this article was written, I learnt that Ph. Droz-Vincent was in fact the first one to write these equations [Lett. al Nuovo Cim., Ser. **1**, 839 (1969)], and that L. Bel proved it was always possible to use 4-vectors.

[6]E.T. Whittaker, *Analytical Dynamics* (Cambridge U.P., Cambridge, 1937).

[7]F.J. Kennedy, J. Math. Phys. **10**, 1349 (1969).

[8]Kennedy worked out this result independently.

[9]R.N. Hill, J. Math. Phys. **8**, 1756 (1967).

[10]Bakamjian and Thomas, Phys. Rev. **92**, 1302 (1953).

Radial charged particle trajectories in the extended Reissner–Nordstrom manifold

J. D. Finley III

The University of New Mexico, Albuquerque, New Mexico 87131
(Received 20 March 1974)

It is shown that the trajectory of a charged particle on the extended Reissner–Nordström manifold can be such as to carry it into regions of the manifold where the definition of energy at infinity is different from the one at its point of origin. The various types of radial trajectories are classified. In the event one considers the manifold as having been produced by a collapsed star, there exist trajectories which go through both horizons, reach a minimum value of r, and go through two more horizons to a copy of the space in which it originated (flat at $r=+\infty$) without colliding with the matter of the collapsed star.

In a recent paper[1] it was shown that there are two distinct types of radial geodesics in the complete Kerr manifold, which can be classified by their place of origin on the manifold. This manifold contains infinitely many copies of two distinct spaces, both flat at $r=\pm\infty$. It is also shown that geodesics cannot cross over from one space to the other. However, this is possible if there is properly applied acceleration. It is the purpose of this paper to show in detail how this crossing over occurs for a very similar manifold: the complete Reissner–Nordström manifold. There has been renewed interest lately in this manifold. Ruffini has suggested that a magnetized rotating object should have a nonzero net charge in order to achieve a minimum energy configuration, and also that a very rapidly rotating, sufficiently small star would be able to maintain this charge in interstellar space.[2]

We will start with the Reissner–Nordström metric in Schwarzschild-like coordinates[3]

$$ds^2 = H^{-1}\,dr^2 + r^2\,d\Omega^2 - H\,dt^2, \tag{1}$$

where

$$H = H(r) = 1 - 2m/r + e^2/r^2,$$

and $d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2$ is the usual spherical surface element. Only the case $m^2 > e^2$ will be considered since otherwise the manifold is already complete. The complete extension was first determined by Graves and Brill[4] and given in a more convenient form by Carter.[5] Carter's extension is created by the repeated use of two null metrics. We define one coordinate system (r, u, θ, ϕ) with metric

$$ds^2 = 2dr\,du - H\,du^2 + r^2\,d\Omega^2 \tag{2a}$$

and another similar coordinate system (r, w, θ, ϕ) with metric

$$ds^2 = 2dr\,dw - H\,dw^2 + r^2\,d\Omega^2, \tag{2b}$$

where

$$u = \tfrac{1}{2}F(r) + t, \quad w = \tfrac{1}{2}F(r) - t, \quad \text{and} \quad \frac{dF}{dr} = \frac{2}{H}. \tag{2c}$$

This implies that

$$F(r) = 2r + mK_+^{-1}\log|r/r_+ - 1| + mK_-^{-1}\log|r/r_- - 1|, \tag{3a}$$

where $r_\pm$ are the roots of H with

$$K_\pm = (r_\pm - r_\mp)/(2r_\pm^2) \quad \text{and} \quad r_\pm = m \pm (m^2 - e^2)^{1/2}. \tag{3b}$$

Note that the function $F(r)$ is separately monotonic in each of the three regions

$$\begin{aligned} &\text{I: } r_+ < r, \\ &\text{II: } r_- < r < r_+, \\ &\text{III: } 0 < r < r_-. \end{aligned} \tag{4}$$

Each of these coordinate systems is analytic and extensible to a manifold larger than the one upon which the original coordinates were defined. Where these two manifolds overlap, one may introduce full null coordinates (u, w, θ, ϕ) with the metric

$$ds^2 = H\,du\,dw + r^2\,d\Omega^2. \tag{5}$$

This overlap region will be one of the three regions in Eq. (4); therefore, given u, w, and a region, one may uniquely determine r. We may then introduce, following Carter, a new coordinate system $(\xi, \psi, \theta, \phi)$ by

$$\pm h(u) = \tan(\psi + \xi), \quad \pm h(w) = \tan(\psi - \xi), \tag{6}$$

where $h(z)$ must be a monotone increasing function such that $h(z) = O[\exp(-K_\pm z)]$ as $z \to \mp\infty$. The complete manifold will then consist of an infinite sequence of (r, u) patches labeled $(-, m)$, and superimposed on this, a similar sequence of (r, w) patches labeled $(n, -)$ running perpendicularly to the (r, u) sequence. By labeling each intersection by (n, m) the manifold consists of those intersections where $|n - m| \leqslant 1$. If $n = m$ is odd (even), then it is a II ($\overline{\text{II}}$) region; if n is even (odd) and $<$ ($>$) m, then it is a I (I′) region; if n is even (odd) and $>$ ($<$) m, then it is a III (III′) region. The choice in sign in the definition of ξ and ψ is determined by which of the regions I, I′, II, etc. is under consideration. Given an (n, m), the sign is $+h(u)$ $[-h(u)]$ for m odd [even], and equivalently for n with $\pm h(w)$.[6]

By denoting by E the constant of the motion associated with the timelike Killing vector, in the original coordinates of Eq. (1), and using a prime to denote the total derivative with respect to proper time τ, the equations of motion for a particle in radial motion with charge to mass ratio X are

$$(r')^2 = D^2 - H, \tag{7a}$$

$$t' = D/H, \tag{7b}$$

 Copyright © 1974 American Institute of Physics

where $D=D(r)=E-eX/r$. (7c)

Solving Eq. (7a) for the constant E (which has the interpretation of the energy per unit mass in unprimed regions and the negative of the energy per unit mass in primed regions[7]), we have

$$\begin{aligned} E &= eX/r \pm [H+(r')^2]^{1/2} \\ &= eX/r \pm [1+(r')^2 - 2m/r + e^2/r^2]^{1/2} \\ &\approx eX/r \pm [1+\tfrac{1}{2}(r')^2 - m/r + e^2/2r^2], \end{aligned} \tag{8}$$

since, for large r, both $(r')^2$ and $H-1$ are small compared to 1. Since eX/r is just the classical potential energy per unit mass of the electromagnetic interaction between the black hole and the test particle, this equation has a reasonable appearance for an energy equation, where the term $e^2/2r^2$ is an additional gravitational term due to the energy of the electric field associated with the charge e, and the ± sign is reminiscent of problems with the Klein—Gordon equation in particle physics.[8] Here, however, both signs are needed, since the sign of E must be negative at $r=\infty$ in primed regions.

The solutions to the equations can be written in the following form when $E^2<1$ (bound test particle), in terms of a parameter η which is adjusted to be 0 at maximum r,

$$r=m(\alpha+\beta\cos\eta), \tag{9a}$$

$$\tau-\tau_0=m(\alpha\eta+\beta\sin\eta)/(1-E^2)^{1/2}, \tag{9b}$$

$$\begin{aligned} t-t_0 &= (\tau-\tau_0)E + (2mE-eX)\eta/(1-E^2)^{1/2} \\ &+ \tfrac{1}{2}K_+^{-1}[\operatorname{sgn}D(r_+)]\log\left|\frac{\tan(\eta/2)+\tan(\eta_+/2)}{\tan(\eta/2)-\tan(\eta_+/2)}\right| \\ &+ \tfrac{1}{2}K_-^{-1}[\operatorname{sgn}D(r_-)]\log\left|\frac{\tan(\eta/2)+\tan(\eta_-/2)}{\tan(\eta/2)-\tan(\eta_-/2)}\right|, \end{aligned} \tag{9c}$$

where $\eta_\pm$ are the values of η at which $r=r_\pm$, while $\alpha\pm\beta$ are the roots of $r'(r)=0$:

$$\alpha=(m-EXe)/(1-E^2),\quad \beta=[m^2-2EXem+e^2-1)]^{1/2}/|1-E^2|. \tag{9d}$$

Solutions for $E^2>1$ are similar and may be obtained from Eqs. (9a)—(9c) by the following substitutions. Change everywhere $(1-E^2)^{1/2}$ to $(E^2-1)^{1/2}$. Then there are two cases: If β is real, replace $\cos\eta$ by $[\operatorname{sgn}(r-\alpha-\beta)]\cosh\psi$, $\sin\eta$ by $[\operatorname{sgn}(r-\alpha-\beta)]\sinh\psi$, and $\tan(\eta/2)$ by $\tanh(\psi/2)$, where ψ increases from $-\infty$ if $r>\alpha+\beta$ and from 0 if $r<\alpha-\beta$. If β is complex, define $\gamma^2=-\beta^2$ and replace $\beta\cos\eta$ by $\gamma\sinh\psi$, $\beta\sin\eta$ by $\gamma\cosh\eta$ and $\tan(\eta/2)$ by $\tanh(\psi/2)$, where ψ increases from $-\infty$. In particular instances $\psi_\pm$ may both be complex, which means the particular trajectory never crosses the horizons, $r=r_\pm$. From Eq. (9a) we see that these radial trajectories are oscillatory in the coordinate r, although we shall see that they do not actually come back to their starting point on the extended manifold (unless, of course, one identifies various different regions of the same type, which leads to serious causal problems); $\alpha\pm\beta$ are just the turning points of this r motion. It is, however, possible for $\alpha-\beta$ to be negative in which case the particle strikes $r=0$ first, which is a singularity. It is also clear that t becomes infinite at $r=r_\pm$, which merely indicates that it is no longer a good coordinate; however, either u or w is finite at $r=r_\pm$. From Eqs. (3) and (7) we find that

$$u'=[D+(\operatorname{sgn}r')(D^2-H)^{1/2}]/H, \tag{10a}$$

$$w'=[-D+(\operatorname{sgn}r')(D^2-H)^{1/2}]/H. \tag{10b}$$

It is then easily seen that at $r=r_\pm$ (roots of $H=0$) u' (w') is finite if $\operatorname{sgn}(r'D)=-1$ (+1).

We now proceed to discuss the possible trajectories in more detail. In particular we divide all trajectories originating in a given region I into classes, as a function of E, X, and e, which have a given future history. From Eq. (7a) one sees that D may vanish along a trajectory only when H is negative, which happens in regions II and $\overline{\text{II}}$. There may then exist trajectories for which D changes sign while the particle is passing through such a region. This would then change which of u or w is finite as the boundary of the region is crossed, and therefore change which boundary is crossed. For sufficiently large r, D and E must have the same sign [Eq. (7c)], so that D is positive in region I and negative in I′. We now restrict consideration to particles originating in region I, while in this region the energy is given by Eq. (8) with a plus sign and is, of course, a fixed number for a given trajectory thereafter. Defining

$$V_\pm=eX/r\pm[H(r)]^{1/2}, \tag{11}$$

we see that for $r>r_+$, $E\geq V_+$, but for $r<r_-$, we have either $E\geq V_+$ (if $D>0$) or $E\leq V_-$ (if $D<0$). There are then five possible types of trajectories. In Fig. 1 is exhibited an E,X plane, for a specific choice of $e=0.8\ m$, which is divided into regions according to the future history of a trajectory with those initial conditions. If $X\leq-1$ then the trajectory ends at the singularity $r=0$ in region III [type (a)]. If $-1<X\leq 0$, then the trajectory enters region $\overline{\text{III}}$, reaches a minimum value of r and rebounds through $\overline{\text{II}}$ back into another I region [type (b)]. However, when $X>0$, there are more possibilities since D now may change sign. For $0<X<1$, if $E>eX/r_-$ the minimum r lies in region III as above. But for $E<eX/r_-$ an infalling particle starting in region I enters region II and, at some point in region II, D becomes negative. The particle must then continue into region III′, reach a minimum value of r there and rebound back into $\overline{\text{II}}$, where D becomes positive again, allowing it to exit into

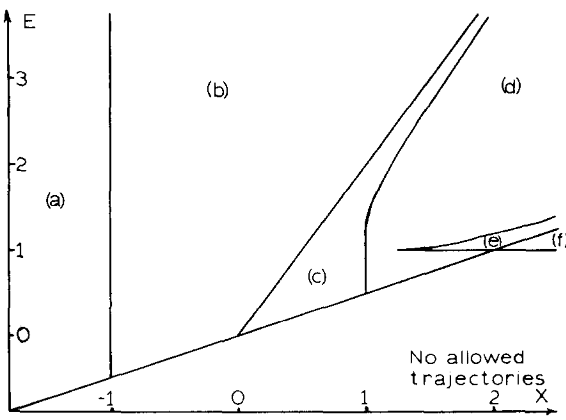

FIG. 1. Determination of the future history of a trajectory which originated in region I with given values of the energy per unit mass, E, and the charge per unit mass, X.

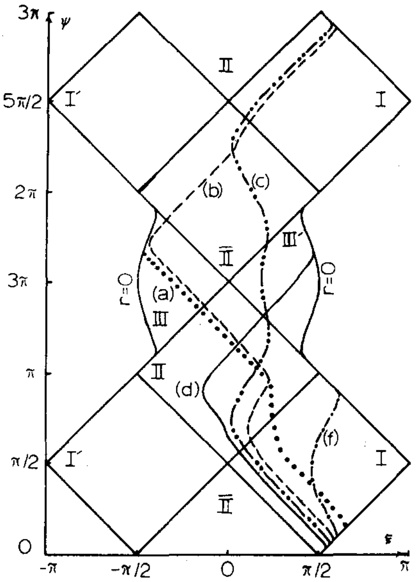

FIG. 2. Typical examples of different types of radial trajectories on the extended Reissner–Nordström manifold. The particular values chosen for these curves were: (a), $E=05$, $X=-2$, $t_0=-13.57$ m; (b), $E=0.3$, $X=-0.7$, $t_0=-4.88$ m; (c), $E=0.7$, $X=0.5$, $t_0=-8.74$ m; (d), $E=0.96$, $X=1.2$, $t_0=-23.09$ m; (e), $E=2.0$, $X=4.2$, $t_0=-2.71$ m.

another region I [type (c)].

If $1\leqslant X\leqslant m/e$, then there are three possibilities. If $E<\Delta_+$, the trajectory enters into region III′ ($D<0$) and hits the singularity at $r=0$ [type (d)], where $\Delta_\pm=[mX \pm (X^2-1)^{1/2}(m^2-e^2)^{1/2}]/e$. If $m/e<X$, there are four possibilities. If $E<1$, the trajectory is of type (c), ending at $r=0$ in III′. If $1\leqslant E<eX/r_+$ the trajectory will stay in region I, eventually going toward $r=+\infty$ [type (f)]. If $\Delta_-<E$, then one has trajectories of types (d), (c), and (b), as shown in Fig. 1. However, for $1<E<\Delta_-$, the situation is more complicated because $V_+(r)$ has a maximum at $r=s$,

$$s=e^2\{m-X|(m^2-e^2)/(X^2-1)|^{1/2}\}^{-1}\geqslant r_+. \tag{12}$$

So if the initial value of r is greater than s, the trajectory will stay always in region I, eventually going toward $r=+\infty$. If the initial value of r is less than s, the trajectory will end at $r=0$ in region III′ [type (e), a choice between motions of types (d) and (f).]. For larger values of E there are trajectories of types (d), (c), and (b), as is shown in Fig. 1. In Fig. 2 typical examples of these various possible trajectories are shown on the extended manifold for a fixed θ and ϕ.

We note that for $X<0$ there exist trajectories for which the energy is negative; i.e., states in region I for which $E<0$ even though $D>0$. These trajectories are an indication that the energy of electrical attraction can be so negative as to overwhelm the energy associated with the rest mass.[9] In the case $E<0$ the maximum value of r for the orbit, d, must satisfy

$$m+(m^2-e^2)^{1/2}=r_+\leqslant d\leqslant m+(m^2-e^2+e^2X^2)^{1/2}. \tag{13}$$

For any particular fixed value of E, with $D>0$, there is a maximum value of X for which that E can be realized by a particle on a radial orbit—$X_{\max}=Er_+/e$. For $X=X_{\max}$, $d=r_+$ and the gap between states with positive D is zero. Therefore, increasing X so that $X>X_{\max}$ causes D to become negative, and the value of d now increases with increasing X, but the starting point of the motion is in region I′. Also the energy is now positive since the energy has been seen to be $-E$ in region I′.

We consider in detail a sequence of particles all released at the same starting point, $d>r_+$, but such that the members of the sequence have increasing charge to mass ratio, X. Since all the particles are momentarily at rest at $r=d$, the energy depends on d, and is given by

$$E_d=eX/d+[H(d)]^{1/2}, \tag{14a}$$

with the turn around point at minimum r given as

$$d_-=e^2(1-X^2)/[d(1-A^2)]. \tag{14b}$$

Starting with $X=0$ and looking at particles with larger and larger values of X, one obtains trajectories of type (b), above, similar to geodesic trajectories. But as X approaches

$$X_0=\frac{r_-}{e}\left|\frac{d-r_+}{d-r_-}\right|^{1/2}<1,$$

d_- approaches r_- and u_-, the value of u at $r=r_-$, approaches $+\infty$. For $X>X_0$, $D(r_-)<0$ and w_- is finite rather than u_-, while d_- is again less than r_- but in region III′. So the trajectory now exits from region II into III′ [type (c)]. Increasing X further to $X=1$, we find that the particle hits the singularity at $r=0$ in region III′ [type (d)]. However, there is a point at which the charge to mass ratio gets so large that there is no longer an attractive force at $r=d$. For X greater than

$$X_1=[m-(e^2/d)]/[e^2H(d)]^{1/2}>1,$$

a particle released at $r=d$, momentarily at rest there, will be repelled toward $r=\infty$, all in region I [type (f)].

It is seen that a full set of (radial) trajectories on the extended manifold requires use of both the plus and the minus sign for the energy in Eq. (8). On those trajectories for which D changes sign, one must use both signs in Eq. (8) for a single trajectory. Also note that even in the case where the collapsing matter which caused the horizon is not ignored, the trajectories of types (c) and (d), as well as (f) are perfectly feasible since the matter lies only in unprimed regions[10] and no collision with it occurs for these orbits.

[1]R.H. St. John and J.D. Finley III, J. Math. Phys. 15, 147 (1974).
[2]R. Ruffini and A. Treves, Astrophys. Lett. 13, 109 (1973) and R. Ruffini, in *Black Holes*, edited by C. deWitt and B.S. deWitt (Gordon and Breach, New York, 1973), p. 525.

[3]We use units in which $c=1=G$. When observed from very far away, the central region has mass m and electric charge e,

which we assume is positive. For purposes of comparison, in Gaussian units, with $c=1=G$, the charge of a single proton is 1.381×10^{-39} km $=9.353\times10^{-40}$ solar masses.

[4]J.C. Graves and D.R. Brill, Phys. Rev. **120**, 1507 (1960).

[5]B. Carter, Phys. Lett. **21**, 423 (1966). Note, however, that his figure (Fig. 1b) relevant to the case $0<e^2<m^2$ does not correctly indicate the locations of the singularities $r=0$. [A similar figure is also in Misner, Thorne, and Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Fig. 34.4.] There is no allowable choice of the function $h(z)$ [Eq. (6)] which makes the singularity a vertical line in the ξ,ψ plane, since $h(z)$ may not be solely even or odd, except when $e^2=m^2$. It is always a curve with two symmetrical bulges toward the ψ axis, as indicated in Fig. 2, where the choice $h(z) = e^{-K_- z} - e^{-K_+ z}$ has been made.

[6]A more complete description of the manifold will be found in B. Carter, Phys. Rev. **141**, 1242 (1966).

[7]See Ref. 1 for a complete discussion of this.

[8]See also R. Ruffini and D. Chrsitodoulou, Phys. Rev. D **4**, 3552 (1971), for some reference to this problem.

[9]These trajectories are discussed in more detail, in region I, by R. Ruffini, in *Black Holes*, p. 503 (see Ref. 2).

[10]For example, see Ya. B. Zeldovich and I.D. Novikov, *Relativistic Astrophysics* (Univ. of Chicago Press, Chicago, 1971), p. 147.

A general method for obtaining Clebsch-Gordan coefficients of finite groups. I. Its application to point and space groups

Isao Sakata

Department of Physics, Osaka City University, Osaka, Japan
(Received 7 September 1972)

A general method is developed for obtaining Clebsch–Gordan coefficients of finite groups. With this method Clebsch–Gordan coefficients are obtained in a matrix form, whereas the so-called basis-function generating machine generates these coefficients one by one. The method is applied to double point group $\bar{\mathbf{D}}_3$, the point group $\mathbf{T}$, and the nonsymmorphic space group $\bar{\mathbf{D}}_{4h}^{14}$. It will be shown that the method can be simplified by the conservation law of the reduced wave vectors when applied to space groups.

1. INTRODUCTION

In case when Clebsch—Gordan (or CG for short) coefficients of a given finite group are to be obtained, one usually makes use of the so-called basis-function generating machine to obtain them.[1] In this method, by the successive applications of projection and step operators to the basis functions for a direct product representation, one can generate basis functions one by one for the irreducible representations which are to be obtained by reducing the direct product representation. Since this method is somewhat heuristic, one sometimes makes vain efforts. If one operates a projection operator on a product basis function and obtains a vanishing result, one must operate it on another function. And this procedure must be repeated until a nonvanishing result is achieved.

A prescription to be presented in this paper straightforwardly gives in a single matrix a whole set of CG coefficients for a direct product of two irreducible representations. Moreover, the prescription is found to be very useful when applied to space groups.

In Sec. 2 a theorem is presented which provides us with a similarity transformation matrix connecting two equivalent irreducible representations. Klauder and Gay's method[2] to induce the irreducible representations of solvable groups proves to be a special case of this theorem. In Sec. 3 the theorem is extended to reducible representations, leading to a general prescription for obtaining CG coefficients. In Sec. 4 the prescription is applied to two point groups $\bar{\mathbf{D}}_3$ and $\mathbf{T}$. In Sec. 5 the prescription is also applied to a nonsymmorphic space group $\bar{\mathbf{D}}_{4h}^{14}$ ($P4_2/mnm$), the symmetry group for the rutile structure in paramagnetic phase. Through this application, it will be shown that the prescription can be simplified by the conservation law of the reduced wave vectors when applied to space groups.

The discussion in this paper is limited to unitary groups. The extension of the method to antiunitary groups will be discussed in a later paper.

Since every representation of finite groups is equivalent to a unitary representation we assume, without loss of generality, that all the representations appearing in this paper are unitary.

In addition, Schoenflies' notation is employed to express point groups and space groups.

2. A MATRIX CONNECTING TWO EQUIVALENT IRREDUCIBLE REPRESENTATIONS

The starting point for this paper is the following theorem:

If D and D' are two equivalent irreducible representations of a finite group $\mathbf{G}$, being related by a unitary matrix U through

$$D'(r) = UD(r)U^\dagger \quad \text{for every element } r \text{ in } \mathbf{G}, \tag{1}$$

then a matrix given by

$$\sum_{r\in\mathbf{G}} D'(r)AD(r)^\dagger \tag{2}$$

is equal to the matrix U in Eq. (1) apart from a constant factor, where A is an arbitrary matrix.

Proof: Consider a matrix

$$\sum_{r\in\mathbf{G}} D(r)BD(r)^\dagger, \tag{3}$$

where B is an arbitary matrix. The matrix (3), which is well known as a matrix utilized to prove the orthogonality relation for the irreducible representations, is by Schur's lemma equal to a scalar multiple of unit matrix:

$$\sum_{r\in\mathbf{G}} D(r)BD(r)^\dagger = \lambda\mathbf{1}. \tag{4}$$

If the matrix B is replaced by a matrix A through $B = U^\dagger A$, Eq. (4) becomes

$$\sum_{r\in\mathbf{G}} D(r)U^\dagger AD(r)^\dagger = \lambda\mathbf{1}. \tag{5}$$

Multiplying this by U on the left, we get

$$\lambda U = \sum_{r\in\mathbf{G}} D'(r)AD(r)^\dagger, \tag{6}$$

where the relation (1) is used. Thus the theorem is proved.

If, in the above discussion, $\mathbf{G}$ is an invariant subgroup of prime index of some larger group and D is a self-conjugate irreducible representation, then the matrix (2) is equal to the matrix $C(X)$ in Klauder and Gay's paper,[2] where X is used for A.

According to the above theorem, when two irreducible representations D and D' are proved equivalent, i.e., when characters of D and D' are the same, one can find out the matrix U in Eq. (1) by calculating the matrix (2).

 Copyright © 1974 American Institute of Physics

3. A GENERAL METHOD FOR OBTAINING CLEBSCH-GORDAN COEFFICIENTS OF FINITE GROUPS

In this section we shall discuss a general method for obtaining CG coefficients of finite groups. This is done by extending the theorem of the last section to reducible representations.

Let D be a reducible representation of a finite group $\mathbf{G}$, and consider the matrix

$$\sum_{r\in\mathbf{G}} D(r)BD(r)^\dagger, \tag{7}$$

where B is an arbitrary matrix. Assume that D is completely reduced to a direct sum of two irreducible representations $D^{(1)}$ and $D^{(2)}$[3]:

$$D(r)=\begin{bmatrix} D^{(1)}(r) & 0 \\ 0 & D^{(2)}(r)\end{bmatrix} \text{ for every } r \text{ in } \mathbf{G}. \tag{8}$$

Corresponding to the block diagonal form of D, let us block off the matrix B in a similar way. Then the matrix (7) can be written as

$$\sum_{r\in\mathbf{G}} D(r)BD(r)^\dagger$$
$$=\sum_{r\in\mathbf{G}}\begin{bmatrix} D^{(1)}(r) & 0 \\ 0 & D^{(2)}(r)\end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{bmatrix}\begin{bmatrix} D^{(1)}(r)^\dagger & 0 \\ 0 & D^{(2)}(r)^\dagger\end{bmatrix},$$
$$=\begin{bmatrix} \sum_r D^{(1)}(r)B_{11}D^{(1)}(r)^\dagger & \sum_r D^{(1)}(r)B_{12}D^{(2)}(r)^\dagger \\ \sum_r D^{(2)}(r)B_{21}D^{(1)}(r)^\dagger & \sum_r D^{(2)}(r)B_{22}D^{(2)}(r)^\dagger\end{bmatrix}.$$

In the matrix of the right-hand side, if $D^{(1)}$ and $D^{(2)}$ are inequivalent, the diagonals are scalar matrices and the off-diagonals are null matrices, i.e.,

$$\sum_{r\in\mathbf{G}} D(r)BD(r)^\dagger=\begin{bmatrix} \lambda 1 & 0 \\ 0 & \mu 1'\end{bmatrix}. \tag{9}$$

The scalar constants λ and μ are related to the traces of B_{11} and B_{22}, respectively.

Now let us denote by D' a reducible representation to which the completely reduced representation D given by (8) is transformed by a unitary matrix M:

$$D'(r)=MD(r)M^\dagger \quad \text{for every } r \text{ in } \mathbf{G}. \tag{10}$$

Replacing B in (9) by $M^\dagger A$ and multiplying both sides of (9) on the left by M, we obtain the matrix equation

$$\sum_{r\in\mathbf{G}} D'(r)AD(r)^\dagger=\begin{bmatrix} \lambda M_{11} & \mu M_{12} \\ \lambda M_{21} & \mu M_{22}\end{bmatrix}, \tag{11}$$

where the relation (10) is used.

The representation D' in (11) can be general reducible one. If, in particular, D' is a direct product representation of two irreducible representations $D^{(\alpha)}$ and $D^{(\beta)}$, then the matrix on the left-hand side of (11) provides us with *unnormalized* CG coefficients. In other words, CG coefficients are obtained by normalizing the columns of the matrix

$$F^{\alpha\times\beta}(\mathbf{G})\equiv\sum_{r\in\mathbf{G}}[D^{(\alpha)}(r)\times D^{(\beta)}(r)]AD(r)^\dagger, \tag{12}$$

where the symbol $\times$ stands for the direct product of two irreducible representations, and D is a completely reduced representation for $D^{(\alpha)}\times D^{(\beta)}$.

Before applying above results to practical problems we shall mention two points which will prove useful later on.

If a group $\mathbf{G}$ has a subgroup $\mathbf{H}$, the group $\mathbf{G}$ can be expressed as

$$\mathbf{G}=a_1\mathbf{H}+a_2\mathbf{H}+\cdots+a_m\mathbf{H},$$

where $a_1, a_2, \ldots, a_m$ are coset representatives of $\mathbf{G}$ with respect to $\mathbf{H}$; we can take $a_1=e$ (the identity element). In this case, calculation of the matrix (12) is practically simplified in the following two steps. Let us first calculate the matrix

$$F^{\alpha\times\beta}(\mathbf{H})=\sum_{r\in\mathbf{H}}[D^{(\alpha)}(r)\times D^{(\beta)}(r)]AD(r)^\dagger \tag{13}$$

summed over all the elements of $\mathbf{H}$, then the matrix for $\mathbf{G}$

$$F^{\alpha\times\beta}(\mathbf{H})+[D^{(\alpha)}(a_2)\times D^{(\beta)}(a_2)]F^{\alpha\times\beta}(\mathbf{H})D(a_2)^\dagger.$$
$$+\cdots+[D^{(\alpha)}(a_m)\times D^{(\beta)}(a_m)]F^{\alpha\times\beta}(\mathbf{H})D(a_m)^\dagger. \tag{14}$$

Equation (14) is clearly identical with $F^{\alpha\times\beta}(\mathbf{G})$.

When $\mathbf{G}$ is a double rotation group, there exists a barred element $\bar{r}$ for any element r of $\mathbf{G}$. If r is a rotation through an angle ϕ about some axis, the $\bar{r}$ may be interpreted to be a rotation through an angle $\phi+2\pi$ about the same axis. Representations of a double group can be classified into two types according to whether $D(\bar{r})=D(r)$ or $D(\bar{r})=-D(r)$. Since in either case the relationship holds

$$[D^{(\alpha)}(\bar{r})\times D^{(\beta)}(\bar{r})]AD(\bar{r})^\dagger=[D^{(\alpha)}(r)\times D^{(\beta)}(r)]AD(r)^\dagger,$$

it is sufficient to take summation in (12) or (13) over only the unbarred elements of the double group.

4. TWO EXAMPLES: $\bar{\mathbf{D}}_3$ AND T

We are now in a position to apply the prescription (12) or (14) to practical problems. Let us first take the double point group $\bar{\mathbf{D}}_3$ as an example. The group $\bar{\mathbf{D}}_3$ has $\bar{\mathbf{C}}_3$ as a subgroup of index two: $\bar{\mathbf{D}}_3=\bar{\mathbf{C}}_3+C_{2x}\bar{\mathbf{C}}_3$, where C_{2x} is a rotation by π around the x axis. The matrices in six irreducible representations of $\bar{\mathbf{D}}_3$ are given in Table I for three elements of $\bar{\mathbf{C}}_3$ and for C_{2x}. Among these representations, D_1, D_2, and D_6 are the representations such that $D(\bar{r})=D(r)$, and $\bar{D}_3$, $\bar{D}_4$ and $\bar{D}_5$ are otherwise.

Let us consider a product representation $\bar{D}_5\times D_6$, which is reducible to $\bar{D}_3+\bar{D}_4+\bar{D}_5$. In Table I, the matrices in $\bar{D}_5\times D_6$ are also shown. Thus Eq. (13) is

TABLE I. Irreducible representations and a direct product representation $\bar{D}_5\times D_6$ of $\bar{\mathbf{D}}_3$.[a]

$\bar{\mathbf{D}}_3$	E	C_{3z}	C_{3z}^2	C_{2x}
D_1	(1)	(1)	(1)	(1)
D_2	(1)	(1)	(1)	(−1)
$\bar{D}_3$	(1)	(−1)	(1)	(i)
$\bar{D}_4$	(1)	(−1)	(1)	(−i)
$\bar{D}_5$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}\omega&0\\0&-\omega^2\end{bmatrix}$	$\begin{bmatrix}\omega^2&0\\0&-\omega\end{bmatrix}$	$\begin{bmatrix}0&\omega\\\omega^2&0\end{bmatrix}$
D_6	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}\omega^2&0\\0&-\omega\end{bmatrix}$	$\begin{bmatrix}-\omega&0\\0&\omega^2\end{bmatrix}$	$\begin{bmatrix}0&\omega^2\\-\omega&0\end{bmatrix}$
$\bar{D}_5\times D_6$	$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	$\begin{bmatrix}-1&0&0&0\\0&-\omega^2&0&0\\0&0&\omega&0\\0&0&0&-1\end{bmatrix}$	$\begin{bmatrix}1&0&0&0\\0&-\omega&0&0\\0&0&\omega^2&0\\0&0&0&1\end{bmatrix}$	$\begin{bmatrix}0&0&0&-1\\0&0&-\omega^2&0\\0&-\omega&0&0\\1&0&0&0\end{bmatrix}$

[a] $\omega=\exp(\pi i/3)$.

$$F^{\bar{5}\times 6}(\bar{\mathbf{C}}_3)=2\times\left\{\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\\a_{41}&a_{42}&a_{43}&a_{44}\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\right.$$

$$+\begin{bmatrix}-1&0&0&0\\0&-\omega^2&0&0\\0&0&\omega&0\\0&0&0&-1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\\a_{41}&a_{42}&a_{43}&a_{44}\end{bmatrix}\begin{bmatrix}-1&0&0&0\\0&-1&0&0\\0&0&-\omega^2&0\\0&0&0&\omega\end{bmatrix}$$

$$\left.+\begin{bmatrix}1&0&0&0\\0&-\omega&0&0\\0&0&\omega^2&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\\a_{41}&a_{42}&a_{43}&a_{44}\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&-\omega&0\\0&0&0&\omega^2\end{bmatrix}\right\},$$

$$=6\times\begin{bmatrix}a_{11}&a_{12}&0&0\\0&0&0&a_{24}\\0&0&a_{33}&0\\a_{41}&a_{42}&0&0\end{bmatrix},\quad \omega=\exp(\pi i/3),$$

and also Eq. (14) is

$$\begin{bmatrix}a_{11}&a_{12}&0&0\\0&0&0&a_{24}\\0&0&a_{33}&0\\a_{41}&a_{42}&0&0\end{bmatrix}+\begin{bmatrix}0&0&0&-1\\0&0&-\omega^2&0\\0&-\omega&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&0&0\\0&0&0&a_{24}\\0&0&a_{33}&0\\a_{41}&a_{42}&0&0\end{bmatrix}$$

$$\times\begin{bmatrix}-i&0&0&0\\0&i&0&0\\0&0&0&-\omega\\0&0&-\omega^2&0\end{bmatrix}$$

$$=\begin{bmatrix}a_{11}+ia_{41}&a_{12}-ia_{42}&0&0\\0&0&0&a_{24}-a_{33}\\0&0&a_{33}-a_{24}&0\\a_{41}-ia_{11}&a_{42}+ia_{12}&0&0\end{bmatrix}.$$

In this matrix, the first column gives us CG coefficients of $\bar{D}_5\times D_6$ into $\bar{D}_3$, the second column into $\bar{D}_4$, and the third and the fourth columns into $\bar{D}_5$. Normalizing each column of the above matrix, we have

$$\begin{bmatrix}1/\sqrt{2}&1/\sqrt{2}&0&0\\0&0&0&1\\0&0&-1&0\\-i/\sqrt{2}&i/\sqrt{2}&0&0\end{bmatrix}$$

apart from a constant factor of absolute value unity. Thus we obtain Table II for the CG coefficients of $\bar{D}_5\times D_6$ of double point group $\bar{D}_3$ with respect to bases which transform according to Table I.

In some cases, we do not need all of the CG coefficients for the decomposition of $\bar{D}_5\times D_6$ into $\bar{D}_3+\bar{D}_4+\bar{D}_5$

TABLE II. Clebsch–Gordan coefficients for $\bar{D}_5\times D_6$ of point group $\bar{\mathbf{D}}_3$ with respect to bases which transform according to Table I.

	$\Psi(\bar{D}_3)$	$\Psi(\bar{D}_4)$	$\Psi_1(\bar{D}_5)$	$\Psi_2(\bar{D}_5)$
$\psi_1(\bar{D}_5)\psi_1(D_6)$	$1/\sqrt{2}$	$1/\sqrt{2}$	0	0
$\psi_1(\bar{D}_5)\psi_2(D_6)$	0	0	0	1
$\psi_2(\bar{D}_5)\psi_1(D_6)$	0	0	−1	0
$\psi_2(\bar{D}_5)\psi_2(D_6)$	$-i/\sqrt{2}$	$i/\sqrt{2}$	0	0

TABLE III. Irreducible representations of point group T. [a]

T	E	C_{2x}	C_{2y}	C_{2z}	$C_3(111)$	$C_3^2(111)$
D_1	(1)	(1)	(1)	(1)	(1)	(1)
D_2	(1)	(1)	(1)	(1)	(ϵ)	(ϵ^2)
D_3	(1)	(1)	(1)	(1)	(ϵ^2)	(ϵ)
D_4	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0&0\\0&1&0\\0&0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$	$\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$

[a] $\epsilon = \exp(2\pi i/3)$.

but only the coefficients into, say $\bar{D}_5$ (i.e., the part shown by dotted lines in Table II). In such cases, it is sufficient only to calculate a matrix

$$\sum_{r\in \mathbf{D}_3} [\bar{D}_5(r)\times D_6(r)]A\bar{D}_5(r)^\dagger, \qquad (15)$$

where A is a 4×2 rectangular matrix. First, carrying out the summation over the elements of $\bar{\mathbf{C}}_3$ we get

$$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\\a_{31}&a_{32}\\a_{41}&a_{42}\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

$$+\begin{bmatrix}-1&0&0&0\\0&-\omega^2&0&0\\0&0&\omega&0\\0&0&0&-1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\\a_{31}&a_{32}\\a_{41}&a_{42}\end{bmatrix}\begin{bmatrix}-\omega^2&0\\0&\omega\end{bmatrix}$$

$$+\begin{bmatrix}1&0&0&0\\0&-\omega&0&0\\0&0&\omega^2&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\\a_{31}&a_{32}\\a_{41}&a_{42}\end{bmatrix}\begin{bmatrix}-\omega&0\\0&\omega^2\end{bmatrix}$$

$$\approx\begin{bmatrix}0&0\\0&a_{22}\\a_{31}&0\\0&0\end{bmatrix},$$

where $\approx$ means that a numerical factor common to all the elements of the matrix is neglected. Then, augmenting this with the matrices for C_{2x}, we have

$$\begin{bmatrix}0&0\\0&a_{22}\\a_{31}&0\\0&0\end{bmatrix}+\begin{bmatrix}0&0&0&-1\\0&0&-\omega^2&0\\0&-\omega&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}0&0\\0&a_{22}\\a_{31}&0\\0&0\end{bmatrix}\begin{bmatrix}0&-\omega\\-\omega^2&0\end{bmatrix}$$

$$=\begin{bmatrix}0&0\\0&a_{22}-a_{31}\\a_{31}-a_{22}&0\\0&0\end{bmatrix};$$

and the part in Table II is obtained.

In this way one can obtain CG coefficients whenever irreducible representations concerned are known.

In the above example, we have considered the case in which an irreducible representation is contained only

TABLE IV. The Clebsch—Gordan coefficients of $D_4\times D_4$ into D_4 for point group **T**. The constants a, b, c, and d are determined in the text.

	$\Psi_1(D_4)$	$\Psi_2(D_4)$	$\Psi_3(D_4)$	$\Phi_1(D_4)$	$\Phi_2(D_4)$	$\Phi_3(D_4)$
$\psi_1(D_4)\phi_1(D_4)$	0	0	0	0	0	0
$\psi_1(D_4)\phi_2(D_4)$	0	0	a	0	0	c
$\psi_1(D_4)\phi_3(D_4)$	0	b	0	0	d	0
$\psi_2(D_4)\phi_1(D_4)$	0	0	b	0	0	d
$\psi_2(D_4)\phi_2(D_4)$	0	0	0	0	0	0
$\psi_2(D_4)\phi_3(D_4)$	a	0	0	c	0	0
$\psi_3(D_4)\phi_1(D_4)$	0	a	0	0	c	0
$\psi_3(D_4)\phi_2(D_4)$	b	0	0	d	0	0
$\psi_3(D_4)\phi_3(D_4)$	0	0	0	0	0	0

TABLE V. Group multiplication table of the double point group $\bar{\mathbf{D}}_4$.

	C_4	C_4^{-1}	C_2	C_{2a}	C_{2b}	C_{2x}	C_{2y}
C_4	C_2	E	$\bar{C}_4^{-1}$	$\bar{C}_{2y}$	$\bar{C}_{2x}$	$\bar{C}_{2a}$	C_{2b}
C_4^{-1}	E	$\bar{C}_2$	C_4	$\bar{C}_{2x}$	C_{2y}	$\bar{C}_{2b}$	$\bar{C}_{2a}$
C_2	$\bar{C}_4^{-1}$	C_4	$\bar{E}$	$\bar{C}_{2b}$	C_{2a}	C_{2y}	$\bar{C}_{2x}$
C_{2a}	$\bar{C}_{2x}$	$\bar{C}_{2y}$	C_{2b}	$\bar{E}$	$\bar{C}_2$	C_4	C_4^{-1}
C_{2b}	C_{2y}	$\bar{C}_{2x}$	$\bar{C}_{2a}$	C_2	$\bar{E}$	C_4^{-1}	$\bar{C}_4$
C_{2x}	$\bar{C}_{2b}$	$\bar{C}_{2a}$	$\bar{C}_{2y}$	C_4^{-1}	C_4	$\bar{E}$	C_2
C_{2y}	$\bar{C}_{2a}$	C_{2b}	C_{2x}	C_4	$\bar{C}_4^{-1}$	$\bar{C}_2$	$\bar{E}$

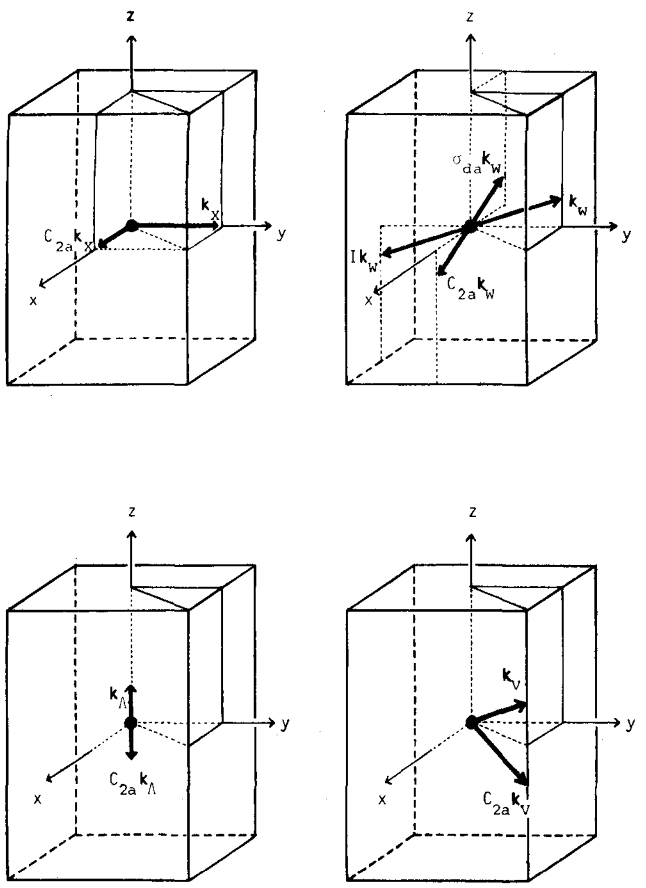

FIG. 1. Stars for the points X, W, Λ, and V of the group $\bar{\mathbf{D}}_{4h}^{14}$.

once in a direct product representation. Next, we shall consider a case where one and the same irreducible representation occurs in a direct product representation more than twice. Let us take the point group $\mathbf{T}$ as an example.

The point group $\mathbf{T}$ contains $\mathbf{D}_2$ as a subgroup and can be written $\mathbf{T}=\mathbf{D}_2+C_3(111)\mathbf{D}_2+C_3^2(111)\mathbf{D}_2$. In Table III the irreducible representations of $\mathbf{T}$ for four elements of $\mathbf{D}_2$, $C_3(111)$, $C_3^2(111)$ are given. A direct product representation $D_4\times D_4$ contains the irreducible representation D_4 twice: $D_4\times D_4=D_1+D_2+D_3+2D_4$. A similar calculation as in the above example gives us

$$\sum_{r\in\mathbf{T}}[D_4(r)\times D_4(r)AD_4(r)^\dagger$$

$$=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a_{23}+a_{72}+a_{61} \\ 0 & a_{32}+a_{81}+a_{43} & 0 \\ 0 & 0 & a_{32}+a_{81}+a_{43} \\ 0 & 0 & 0 \\ a_{23}+a_{72}+a_{61} & 0 & 0 \\ 0 & a_{23}+a_{72}+a_{61} & 0 \\ a_{32}+a_{81}+a_{43} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the CG coefficients of $D_4\times D_4$ to two D_4's turn out to be those shown in Table IV.

A remaining problem is to determine the constants a, b, and c, and d in Table IV. But these constants can take arbitrary values so long as they satisfy the orthonormality condition of CG coefficients. The reason for this is that, as easily be seen, the matrix $F^{\alpha\times\beta}(\mathbf{G})$ satisfies the relation

$$F^{\alpha\times\beta}(\mathbf{G})D(r)=[D^{(\alpha)}(r)\times D^{(\beta)}(r)]\,F^{\alpha\times\beta}(\mathbf{G}) \qquad (16)$$

irrespective of whether $D^{(\alpha)}\times D^{(\beta)}$ contains an irreducible representation only once or more than twice. Thus we can choose $a=1$, $b=0$, and so $c=0$, $d=1$; or, if the resulting basis functions are to be symmetric and antisymmetric product, we conveniently choose $a=b=1/\sqrt{2}$ and $c=-d=1/\sqrt{2}$, respectively.

5. APPLICATION TO SPACE GROUP

We will consider the nonsymmorphic double space group $\bar{\mathbf{D}}_{4h}^{14}$, the symmetry group for the rutile structure in paramagnetic phase (see Table V); the irreducible characters were given by Dimmock and Wheeler.[4] Unless otherwise noted, various notations in this section follow those of Dimmock and Wheeler.

Every irreducible representation of space group is specified by a star of reduced wave vector, and is easily induced from a small representation.[5] In Fig. 1 the stars of $\mathbf{k}_x$ and $\mathbf{k}_w$, the reduced wave vectors for points X and W, respectively, are shown together with the stars for points Λ and V. Looking at these stars, we see that a direct product representation of an irreducible representation for the point X and that for W is, in general, reducible to a direct sum of the irreducible representations for the point Λ and those for point V.

TABLE VI. Small representations for the point $X(0,\pi/a,0)$ of the group $\bar{\mathbf{D}}_{4h}^{14}$.

	$\Delta(X_1)$	$\Delta(X_2)$	$\bar{\Delta}(X_3)$	$\bar{\Delta}(X_4)$	
$\{E\mid\mathbf{t}\}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	
$\{C_2\mid\mathbf{t}\}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}i&0\\0&i\end{pmatrix}$	
$\{I\mid\mathbf{t}\}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\times\exp(i\mathbf{k}_X\cdot\mathbf{t})$
$\{\sigma_h\mid\mathbf{t}\}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-i&0\\0&i\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	
$\{C_{2x}\mid\mathbf{t}+\boldsymbol{\tau}\}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&i\\-i&0\end{pmatrix}$	
$\{C_{2y}\mid\mathbf{t}+\boldsymbol{\tau}\}$	$\begin{pmatrix}0&i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&-i\\i&0\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&1\\-1&0\end{pmatrix}$	
$\{\sigma_{vx}\mid\mathbf{t}+\boldsymbol{\tau}\}$	$\begin{pmatrix}0&i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\times\exp[i\mathbf{k}_X\cdot(\mathbf{t}+\boldsymbol{\tau})]$
$\{\sigma_{vy}\mid\mathbf{t}+\boldsymbol{\tau}\}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}0&i\\i&0\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&-1\\-1&0\end{pmatrix}$	

TABLE VII. Small representations for the point $W(0, \pi/a, \gamma)$ of the group $\mathbf{D}_{4h}^{14}$.[a]

	$\Delta(W_1)$	$\bar{\Delta}(W_2)$	$\bar{\Delta}(W_3)$	$\bar{\Delta}(W_4)$	$\bar{\Delta}(W_5)$	
$\{E\mid \mathbf{t}\}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	(1)	(1)	(1)	(1)	$\times\exp(i\mathbf{k}_W\cdot\mathbf{t})$
$\{C_2\mid \mathbf{t}\}$	$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$	(i)	$(-i)$	(i)	$(-i)$	
$\{\sigma_{vx}\mid \mathbf{t}+\boldsymbol{\tau}\}$	$\begin{bmatrix}0&i\exp(ic\gamma/2)\\-i\exp(-ic\gamma/2)&0\end{bmatrix}$	$(-i)$	(i)	(i)	$(-i)$	$\times\exp[i\mathbf{k}_W\cdot(\mathbf{t}+\boldsymbol{\tau})]$
$\{\sigma_{vy}\mid \mathbf{t}+\boldsymbol{\tau}\}$	$\begin{bmatrix}0&-i\exp(ic\gamma/2)\\-i\exp(-ic\gamma/2)&0\end{bmatrix}$	(-1)	(-1)	(1)	(1)	

[a] $0<\gamma<\pi/c$.

In Tables VI through IX the small representations for these points are listed, which are obtained by making use of solvability of space groups.[6] The ith small representation for a point P in the first Brillouin zone is to be denoted by $\Delta^{(P_i)}$ in these tables, where the barred representation $\bar{\Delta}^{(P_i)}$ is a small representation in which the matrix for r and that for $\bar{r}$ differ in sign.

An irreducible representation of space group is induced from each small representation.[5] Let us denote by $D^{(P_i)}$ the irreducible representation which is induced from $\Delta^{(P_i)}$. As an example, consider the CG coefficients into $\bar{D}^{(\Lambda_7)}$ of a direct product representation $D^{(X_1)}\times\bar{D}^{(W_2)}$, which is reducible to $\bar{D}^{(\Lambda_6)}+\bar{D}^{(\Lambda_7)}+\bar{D}^{(V_6)}+\bar{D}^{(V_7)}$ according to procedure described in Ref. 7. These coefficients are obtained by calculating the matrix

$$\sum_{\alpha}\sum_{\mathbf{t}}[D^{(X_1)}(\{\alpha\mid\mathbf{V}(\alpha)+\mathbf{t}\})\times\bar{D}^{(W_2)}(\{\alpha\mid\mathbf{V}(\alpha)+\mathbf{t}\})]A\cdot\bar{D}^{(\Lambda_7)}(\{\alpha\mid\mathbf{V}(\alpha)+\mathbf{t}\})^{\dagger}, \tag{17}$$

where α is the rotational part of the elements of space group, $\mathbf{v}(\alpha)$ the shortest nonprimitive translation vector associated with α, and $\mathbf{t}$ the primitive translation vector. The irreducible representations $D^{(X_1)}$, $\bar{D}^{(W_2)}$, and $\bar{D}^{(\Lambda_7)}$ are induced from the corresponding small representations. For instance $D^{(X_1)}(\{C_{2b}\mid\mathbf{t}\})$ is obtained as

$$\bar{D}^{(X_1)}(\{C_{2b}\mid\mathbf{t}\})$$

$$=\begin{pmatrix}0 & \Delta^{(X_1)}(\{E\mid 0\}^{-1}\{C_{2b}\mid\mathbf{t}\}\{C_{2a}\mid 0\})\\ \Delta^{(X_1)}(\{C_{2a}\mid 0\}^{-1}\{C_{2b}\mid\mathbf{t}\}\{0\}) & 0\end{pmatrix}$$

$$=\begin{pmatrix}0 & \Delta^{(X_1)}(\{C_2\mid\mathbf{t}\})\\ \Delta^{(X_1)}(\{C_2\mid C_{2a}^{-1}\mathbf{t}\}) & 0\end{pmatrix}$$

$$=\begin{pmatrix}0&0&e^{i\mathbf{k}_X\cdot\mathbf{t}}&0\\0&0&0&-e^{i\mathbf{k}_X\cdot\mathbf{t}}\\e^{iC_{2a}\mathbf{k}_X\cdot\mathbf{t}}&0&0&0\\0&-e^{iC_{2a}\mathbf{k}_X\cdot\mathbf{t}}&0&0\end{pmatrix}, \tag{18}$$

TABLE VIII. Small representations for the point $\Lambda(0,0,\gamma)$ of the group $\mathbf{D}_{4h}^{14}$.[a]

	$\Delta(\Lambda_1)$	$\Delta(\Lambda_2)$	$\Delta(\Lambda_3)$	$\Delta(\Lambda_4)$	$\Delta(\Lambda_5)$	$\bar{\Delta}(\Lambda_6)$	$\bar{\Delta}(\Lambda_7)$	
$\{E\mid\mathbf{t}\}$	(1)	(1)	(1)	(1)	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\times\exp(i\mathbf{k}_\Lambda\cdot\mathbf{t})$
$\{C_2\mid\mathbf{t}\}$	(1)	(1)	(1)	(1)	$\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$	$\begin{bmatrix}i&0\\0&-i\end{bmatrix}$	$\begin{bmatrix}i&0\\0&-i\end{bmatrix}$	
$\{\sigma_{da}\mid\mathbf{t}\}$	(1)	(-1)	(-1)	(1)	$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$	$\begin{bmatrix}0&-1\\1&0\end{bmatrix}$	$\begin{bmatrix}0&-1\\1&0\end{bmatrix}$	
$\{\sigma_{db}\mid\mathbf{t}\}$	(1)	(-1)	(-1)	(1)	$\begin{bmatrix}-1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}0&i\\i&0\end{bmatrix}$	$\begin{bmatrix}0&i\\i&0\end{bmatrix}$	
$\{C_4\mid\mathbf{t}+\boldsymbol{\tau}\}$	(1)	(1)	(-1)	(-1)	$\begin{bmatrix}0&-1\\1&0\end{bmatrix}$	$\begin{bmatrix}\eta&0\\0&-i\eta\end{bmatrix}$	$\begin{bmatrix}-\eta&0\\0&i\eta\end{bmatrix}$	$\times\exp[i\mathbf{k}_\Lambda\cdot(\mathbf{t}+\boldsymbol{\tau})]$
$\{C_4^{-1}\mid\mathbf{t}+\boldsymbol{\tau}\}$	(1)	(1)	(-1)	(-1)	$\begin{bmatrix}0&1\\-1&0\end{bmatrix}$	$\begin{bmatrix}-i\eta&0\\0&\eta\end{bmatrix}$	$\begin{bmatrix}i\eta&0\\0&-\eta\end{bmatrix}$	
$\{\sigma_{vx}\mid\mathbf{t}+\boldsymbol{\tau}\}$	(1)	(-1)	(1)	(-1)	$\begin{bmatrix}0&-1\\-1&0\end{bmatrix}$	$\begin{bmatrix}0&-i\eta\\-\eta&0\end{bmatrix}$	$\begin{bmatrix}0&i\eta\\\eta&0\end{bmatrix}$	
$\{\sigma_{vy}\mid\mathbf{t}+\boldsymbol{\tau}\}$	(1)	(-1)	(1)	(-1)	$\begin{bmatrix}0&1\\1&0\end{bmatrix}$	$\begin{bmatrix}0&\eta\\i\eta&0\end{bmatrix}$	$\begin{bmatrix}0&-\eta\\-i\eta&0\end{bmatrix}$	

[a] $0<\gamma<\pi/c$, $\eta=\exp(i\pi/4)$.

TABLE IX. Small representations for the point $V(\pi/a,\ \pi/a,\gamma)$ of the group $\bar{\mathrm{D}}_{4h}^{14}$.[a]

	$\Delta(V_1)$	$\Delta(V_2)$	$\Delta(V_3)$	$\Delta(V_4)$	$\Delta(V_5)$	$\bar{\Delta}(V_6)$	$\bar{\Delta}(V_7)$	
$\{E\,\vert\,\mathbf{t}\}$	(1)	(1)	(1)	(1)	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\times\exp(i\mathbf{k}_V\cdot\mathbf{t})$
$\{C_2\,\vert\,\mathbf{t}\}$	(1)	(1)	(1)	(1)	$\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$	$\begin{bmatrix}i&0\\0&-i\end{bmatrix}$	$\begin{bmatrix}i&0\\0&-i\end{bmatrix}$	
$\{\sigma_{da}\,\vert\,\mathbf{t}\}$	(1)	(1)	(-1)	(-1)	$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$	$\begin{bmatrix}0&-1\\1&0\end{bmatrix}$	$\begin{bmatrix}0&-1\\1&0\end{bmatrix}$	
$\{\sigma_{db}\,\vert\,\mathbf{t}\}$	(1)	(1)	(-1)	(-1)	$\begin{bmatrix}-1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}0&i\\i&0\end{bmatrix}$	$\begin{bmatrix}0&i\\i&0\end{bmatrix}$	
$\{C_4\,\vert\,\mathbf{t}+\boldsymbol{\tau}\}$	$(-i)$	(i)	$(-i)$	(i)	$\begin{bmatrix}0&-\theta\\-\theta^*&0\end{bmatrix}$	$\begin{bmatrix}-i\eta&0\\0&-\eta\end{bmatrix}$	$\begin{bmatrix}i\eta&0\\0&\eta\end{bmatrix}$	$\times\exp[i\mathbf{k}_V\cdot(\mathbf{t}+\boldsymbol{\tau})]$
$\{C_4^{-1}\,\vert\,\mathbf{t}+\boldsymbol{\tau}\}$	$(-i)$	(i)	$(-i)$	(i)	$\begin{bmatrix}0&\theta\\\theta^*&0\end{bmatrix}$	$\begin{bmatrix}-\eta&0\\0&-i\eta\end{bmatrix}$	$\begin{bmatrix}\eta&0\\0&i\eta\end{bmatrix}$	
$\{\sigma_{vx}\,\vert\,\mathbf{t}+\boldsymbol{\tau}\}$	$(-i)$	(i)	(i)	$(-i)$	$\begin{bmatrix}0&-\theta\\\theta^*&0\end{bmatrix}$	$\begin{bmatrix}0&-\eta\\i\eta&0\end{bmatrix}$	$\begin{bmatrix}0&\eta\\-i\eta&0\end{bmatrix}$	
$\{\sigma_{vy}\,\vert\,\mathbf{t}+\boldsymbol{\tau}\}$	$(-i)$	(i)	(i)	$(-i)$	$\begin{bmatrix}0&\theta\\-\theta^*&0\end{bmatrix}$	$\begin{bmatrix}0&-i\eta\\\eta&0\end{bmatrix}$	$\begin{bmatrix}0&i\eta\\-\eta&0\end{bmatrix}$	

[a] $0<\gamma<\pi/c,\ \ \theta=\exp(ic\gamma/2),\ \ \eta=\exp(i\pi/4)$.

where $\{E\,|\,\mathbf{0}\}$ and $\{C_{2a}\,|\,\mathbf{0}\}$ are chosen as coset representatives of the whole space group with respect to the group of $\mathbf{k}_X$. Multiplication of point group elements in (18) is carried out using the group multiplication table of $\bar{\mathbf{D}}_4$ given in Table V. If $\{\psi(\mathbf{k}_X,\ X_1, 1),\ \psi(\mathbf{k}_X,\ X_1,\ 2)\}$ is to be a basis for $\Delta^{(X_1)}$, then a basis for (18) is

$$\{\psi(\mathbf{k}_X,X_1,1),\ \psi(\mathbf{k}_X,X_1,2)\}+\{C_{2a}\,|\,\mathbf{0}\}\{\psi(\mathbf{k}_X,X_1,1),\ \psi(\mathbf{k}_X,\ X_1,2)\}$$

$$\equiv\{\psi(\mathbf{k}_X,X_1,1),\ \psi(\mathbf{k}_X,X_1,2),\ \psi(C_{2a}\mathbf{k}_X,X_1,1),\ \psi(C_{2a}\mathbf{k}_X,X_1,2)\},$$

where the plus sign stands for the direct sum. In a similar way we get

$$\bar{D}^{(W_2)}(\{C_{2b}\,|\,\mathbf{t}\})=\begin{pmatrix}0 & ie^{i\mathbf{k}_W\cdot\mathbf{t}} & 0 & 0\\ ie^{iC_{2a}\mathbf{k}_W\cdot\mathbf{t}} & 0 & 0 & 0\\ 0 & 0 & 0 & ie^{iI\mathbf{k}_W\cdot\mathbf{t}}\\ 0 & 0 & ie^{i\sigma_{da}\mathbf{k}_W\cdot\mathbf{t}} & 0\end{pmatrix}, \tag{19}$$

and

$$\bar{D}^{(\Lambda_7)}(\{C_{2b}\,|\,\mathbf{t}\})=\begin{pmatrix}0 & 0 & ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0\\ 0 & 0 & 0 & -ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}}\\ ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0\\ 0 & -ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0\end{pmatrix}. \tag{20}$$

The bases for these irreducible representations are

$$\{\psi(\mathbf{k}_W,\bar{W}_2),\ \psi(C_{2a}\mathbf{k}_W,\bar{W}_2),\ \psi(I\mathbf{k}_W,\bar{W}_2),\ \psi(\sigma_{da}\mathbf{k}_W,\bar{W}_2)\},$$

and

$$\{\psi(\mathbf{k}_\Lambda,\bar{\Lambda}_7,1),\ \psi(\mathbf{k}_\Lambda,\bar{\Lambda}_7,2),\ \psi(C_{2a}\mathbf{k}_\Lambda,\bar{\Lambda}_7,1),\ \psi(C_{2a}\mathbf{k}_\Lambda,\bar{\Lambda}_7,2)\},$$

respectively.

Substituting Eqs. (18), (19), and (20) into (17), there appear elements specified by the factor $\exp(i\mathbf{k}_V\cdot\mathbf{t})$ or $\exp(iC_{2a}\mathbf{k}_V\cdot\mathbf{t})$ in the 16×16 matrix $D^{(X_1)}(\{C_{2b}\,|\,\mathbf{t}\})$

TABLE X. Clebsch–Gordan coefficients of a direct product representation $D^{(X_1)}\times\bar{D}^{(W_2)}$ into $\bar{D}^{(\Lambda_7)}$ of the group $\bar{\mathrm{D}}_{4h}^{14}$.

	$\Psi(\mathbf{k}_\Lambda,\bar{\Lambda}_7,1)$	$\Psi(\mathbf{k}_\Lambda,\bar{\Lambda}_7,2)$	$\Psi(C_{2a}\mathbf{k}_\Lambda,\bar{\Lambda}_7,1)$	$\Psi(C_{2a}\mathbf{k}_\Lambda,\bar{\Lambda}_7,2)$
$\psi(\mathbf{k}_X,X_1,1)\psi(\mathbf{k}_W,\bar{W}_2)$	$1/\sqrt{2}$	0	0	0
$\psi(\mathbf{k}_X,X_1,1)\psi(I\mathbf{k}_W,\bar{W}_2)$	0	0	0	$-1/\sqrt{2}$
$\psi(\mathbf{k}_X,X_1,2)\psi(\mathbf{k}_W,\bar{W}_2)$	0	$\exp(-\pi i/4)/\sqrt{2}$	0	0
$\psi(\mathbf{k}_X,X_1,2)\psi(I\mathbf{k}_W,\bar{W}_2)$	0	0	$-\exp(-\pi i/4)\sqrt{2}$	0
$\psi(C_{2a}\mathbf{k}_X,X_1,1)\psi(C_{2a}\mathbf{k}_W,\bar{W}_2)$	0	0	$1/\sqrt{2}$	0
$\psi(C_{2a}\mathbf{k}_X,X_1,1)\psi(\sigma_{da}\mathbf{k}_W,\bar{W}_2)$	0	$1/\sqrt{2}$	0	0
$\psi(C_{2a}\mathbf{k}_X,X_1,2)\psi(C_{2a}\mathbf{k}_W,\bar{W}_2)$	0	0	0	$\exp(-\pi i/4)\sqrt{2}$
$\psi(C_{2a}\mathbf{k}_X,X_1,2)\psi(\sigma_{da}\mathbf{k}_W,\bar{W}_2)$	$\exp(-\pi i/4)\sqrt{2}$	0	0	0

$\times \bar{D}^{(W_2)}(\{C_{2b}|\mathbf{t}\})$. But we need not consider these elements, since the sum over $\mathbf{t}$ of these elements multiplied by any element of $\bar{D}^{(\Lambda_7)}(\{C_{2b}|\mathbf{t}\})^\dagger$ vanished. Therefore we can neglect in the summation over $\mathbf{t}$ the rows and columns containing the factor $\exp(i\mathbf{k}_V\cdot\mathbf{t})$ or $\exp(iC_{2a}\mathbf{k}_V\cdot\mathbf{t})$ in the 16×16 matrices of (17). Thus the terms for which $\alpha = C_{2b}$ of (17) are simplified to the form

$$\sum_{\mathbf{t}} \begin{bmatrix} 0 & 0 & 0 & 0 & ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} \\ ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -ie^{iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 & 0 \end{bmatrix} \cdot A \cdot \begin{bmatrix} 0 & 0 & -ie^{-iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 \\ 0 & 0 & 0 & ie^{-iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t}} \\ -ie^{-i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 \\ 0 & ie^{-i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 \end{bmatrix}, \quad (21)$$

where A is an arbitrary 8×4 rectangular matrix. For other α's than C_{2b} we can simplify the calculations in a similar way. The rows and columns to be neglected in the 16×16 matrices are common to all elements of the space group. In such a way we finally obtain for (17) the following matrix:

$$\begin{bmatrix} \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & \beta\exp(-\pi i/4) & 0 & 0 \\ 0 & 0 & -\beta\exp(-\pi i/4) & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & \beta\exp(-\pi i/4) \\ \beta\exp(-\pi i/4) & 0 & 0 & 0 \end{bmatrix}, \quad (22)$$

where $\beta = a_{11} + a_{53} - a_{24} + a_{62} + i(a_{81} - a_{43} + a_{74} + a_{32}) \times \exp(-\pi i/4)$ and the a_{ij}'s are elements of the matrix A. In calculating (22), we have made use of the fact that the space group $\bar{\mathbf{D}}_{4h}^{14}$ can be written as

$$\bar{\mathbf{D}}_{4h}^{14} = \mathbf{H} + \{I|0\}\mathbf{H} + \{C_4|\tau\}\mathbf{H} + \{S_4^{-1}|\tau\}\mathbf{H},$$

where the subgroup $\mathbf{H}$ is

$$\mathbf{H} \equiv [\{E|\mathbf{t}\}, \{\bar{E}|\mathbf{t}\}, \{C_2|\mathbf{t}\}, \{\bar{C}_2|\mathbf{t}\}, \{C_{2a}|\mathbf{t}\}, \{\bar{C}_{2a}|\mathbf{t}\}, \{C_{2b}|\mathbf{t}\}, \{\bar{C}_{2b}|\mathbf{t}\}].$$

We may take the normalizing constant β to be $1/\sqrt{2}$. Thus Table X is constructed.

It is to be noted that there is more simplified calculation to obtain Table X. First, obtain the C–G coefficients of $D^{(X_1)}\times\bar{D}^{(W_2)}$ into only the basis functions of small representation $\bar{\Delta}^{(\Lambda_7)}$, i.e., the part enclosed with dotted lines in Table X; it is sufficient to calculate simplified matrices. Taking the element $\{C_{2b}|\mathbf{t}\}$ as an example again, the matrix (21) is simplified to

$$\sum_{\mathbf{t}} \begin{bmatrix} 0 & 0 & ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 \\ 0 & 0 & 0 & -ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} \\ ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 & 0 \\ 0 & -ie^{i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 & 0 \end{bmatrix} A \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -ie^{-i\mathbf{k}_\Lambda\cdot\mathbf{t}} & 0 \\ 0 & ie^{-i\mathbf{k}_\Lambda\cdot\mathbf{t}} \end{bmatrix}, \quad (23)$$

where A is an arbitrary 4×4 matrix. The matrix (23) is obtained by deleting the rows and the columns of the 8×8 matrix in (21) which have the factor $\exp(iC_{2a}\mathbf{k}_\Lambda\cdot\mathbf{t})$, the third and the fourth columns of 4×4 matrix in (21). Such deletion is done for all other elements. The rows and the columns which should be deleted are common to all the elements of $\bar{\mathbf{D}}_h^{14}$. Carrying out the summation of (17) where matrices like (23) are substituted, the basis functions $\Psi(\mathbf{k}_\Lambda, \bar{\Lambda}_7, 1)$ and $\Psi(\mathbf{k}_\Lambda, \bar{\Lambda}_7, 2)$ are obtained. The remaining functions $\Psi(C_{2a}\mathbf{k}_\Lambda, \bar{\Lambda}_7, 1)$ and $\Psi(C_{2a}\mathbf{k}_\Lambda, \bar{\Lambda}_7, 2)$ are obtained by applying $\{C_{2a}|0\}$ to $\Psi(\mathbf{k}_\Lambda, \bar{\Lambda}_7, 1)$ and $\Psi(\mathbf{k}_\Lambda, \Lambda_7, 2)$, respectively.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Professor H. Watanabe and Dr. T. Iida for their helpful advices and discussions.

[1]The terminology of "basis-function generating machine" was used by J.H. Van Vleck, as refered to in M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964), p. 41.

[2]L.T. Klauder and J.G. Gay, J. Math. Phys. 9, 1488 (1968).

[3]Extension is straightforward to cases where there are more than three diagonal submatrices.

[4]J.O. Dimmock and R.J. Wheeler, Phys. Rev. 127, 391 (1962).

[5]G.F. Koster, Solid State Phys. 5, 173 (1957); see also Ref. 7.

[6]F. Seitz, Ann. Math. 37, 17 (1936). See also Ref. 7.

[7]J.F. Cornwell, *Selected Topics in Solid State Physics*, edited by E.P. Wholfarth (North-Holland, London, 1969), Vol. X, p. 181; C.J. Bradley and A.P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).

A general method for obtaining Clebsch-Gordan coefficients of finite groups. II. Extension to antiunitary groups

Isao Sakata

Department of Physics, Osaka City University, Osaka, Japan

(Received 7 September 1973)

A general method is presented for obtaining Clebsch–Gordan coefficients, in a matrix form, of finite antiunitary groups, as a direct extension of a general method for unitary groups. It is shown that there is an essential difference as well as apparent similarities between two methods for unitary and antiunitary groups.

1. INTRODUCTION

In a previous paper,[1] the author presented a general method for obtaining Clebsch–Gordan (or CG) coefficients of finite unitary groups and applied it to three examples. In the paper the starting point of our discussion was that a matrix, Eq. (3) of (I),

$$\sum_r D^{(\alpha)}(r) B D^{(\alpha)}(r)^\dagger$$

is a scalar matrix by Schur's lemma. From this we obtained a theorem by means of which a matrix connecting two equivalent irreducible representations of unitary groups can be found. And by extending the theorem to reducible representations, we reached at a general method for obtaining CG coefficients in a matrix form.

For antiunitary groups a theorem corresponding to Schur's lemma for unitary groups does not hold; that is, for any irreducible corepresentation a matrix $\boldsymbol{\alpha}$ satisfying

$$\boldsymbol{\alpha} D(u) = D(u)\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} D(a) = D(a)\boldsymbol{\alpha}^*$$

is not necessarily a scalar matrix, where u and a are unitary and antiunitary elements, respectively. Nevertheless, taking account of similarity relationships [(2) below] characteristic to antiunitary groups, the method will be extended so that CG coefficients for antiunitary groups may be obtained in a matrix form similar to the case of unitary groups. As in (I), every corepresentation appearing in this paper is assumed to consist of unitary matrices.

2. A GENERAL METHOD FOR OBTAINING CLEBSCH-GORDAN COEFFICIENTS OF ANTIUNITARY GROUPS

Corepresentation D of antiunitary groups are characterized by the multiplication rules

$$\begin{aligned} &D(u)D(u') = D(uu'), \quad D(u)D(a) = D(ua), \\ &D(a)D^*(u) = D(au), \quad D(a)D^*(a') = D(aa'); \end{aligned} \tag{1}$$

and two equivalent corepresentations D, D' are connected by

$$D(u) = \boldsymbol{\alpha}^\dagger D'(u)\boldsymbol{\alpha}, \quad D(a) = \boldsymbol{\alpha}^\dagger D'(a)\boldsymbol{\alpha}^*, \tag{2}$$

where $\boldsymbol{\alpha}$ is a unitary matrix.[2] In Eqs. (1) and (2) the elements u, u' are unitary and a, a' are antiunitary. If, in (2), D' is a product corepresentation and D is a corresponding completely reduced corepresentation, then $\boldsymbol{\alpha}$ is a matrix whose elements are CG coefficients. Consider a matrix

$$\begin{aligned} F &\equiv \sum_{u\in H} D'(u) A D^\dagger(u) + \sum_{a\in a_0 H} D'(a) A^* D^\dagger(a) \\ &= \sum_u D'(u) A D(u^{-1}) + \sum_a D'(a) A^* D^*(a^{-1}), \end{aligned} \tag{3}$$

where $\mathbf{H}$ is the unitary subgroup of the antiunitary group under consideration, a_0 is an antiunitary coset representative, and A is an arbitrary matrix. Then we have

$$\begin{aligned} FD(u') &= \sum_u D'(u) A D(u^{-1}u') + \sum_a D'(a) A^* D^*(a^{-1}u') \\ &= \sum_{u''} D'(u'u'') A D(u''^{-1}) + \sum_{a'} D'(u'a') A^* D^*(a'^{-1}) \\ &= D'(u')F, \end{aligned}$$

and

$$\begin{aligned} FD(a') &= D'(a')[\sum_a D'^*(a) A D(a^{-1}) + \sum_u D'^*(u) A^* D^*(u^{-1})] \\ &= D'(a')F^*. \end{aligned}$$

That is, the matrix F satisfies the same equations (2) as $\boldsymbol{\alpha}$. Accordingly CG coefficients of antiunitary groups are obtained by orthonormalizing the columns of the matrix F as in the case of unitary groups.

TABLE I. The irreducible corepresentations of antiunitary double point group $\mathbf{D}_4(\mathbf{D}_2)$.

	E	C_2	$\bar{E}$	$\bar{C}_2$	C_{2x}	θC_4
D_1	(1)	(1)	(1)	(1)	(1)	(1)
D_2	(1)	(1)	(1)	(1)	(−1)	(1)
D_3	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$
D_4	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$	$\begin{pmatrix}i&0&0&0\\0&-i&0&0\\0&0&i&0\\0&0&0&-i\end{pmatrix}$	$\begin{pmatrix}-1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}$	$\begin{pmatrix}-i&0&0&0\\0&i&0&0\\0&0&-i&0\\0&0&0&i\end{pmatrix}$	$\begin{pmatrix}0&-1&0&0\\1&0&0&0\\0&0&0&-1\\0&0&1&0\end{pmatrix}$	$\begin{pmatrix}0&0&0&-1\\0&0&-i&0\\0&1&0&0\\i&0&0&0\end{pmatrix}$

Copyright © 1974 American Institute of Physics

TABLE II. Clebsch—Gordan coefficients of $D_3\times D_3$ for the group $\mathrm{D}_4(\mathrm{D}_2)$ with respect to bases which transform according to Table I.

	$\Psi(D_1)$	$\Phi(D_1)$	$\Psi(D_2)$	$\Phi(D_2)$
$\psi(D_3,1)\phi(D_3,1)$	$1/\sqrt{2}$	$i/\sqrt{2}$	0	0
$\psi(D_3,1)\phi(D_3,2)$	0	0	$1/\sqrt{2}$	$i/\sqrt{2}$
$\psi(D_3,2)\phi(D_3,1)$	0	0	$-1/\sqrt{2}$	$i/\sqrt{2}$
$\psi(D_3,2)\phi(D_3,2)$	$1/\sqrt{2}$	$-i/\sqrt{2}$	0	0

If we put

$$F_u\equiv\sum_{u\in H}D'(u)AD(u^{-1}),\qquad(4)$$

we can write (3) as

$$F=F_u+D'(a_0)F_u^*D^\dagger(a_0).\qquad(5)$$

This equation simplifies calculation of F when we apply this method to practical problems. And also calculation of F_u can be further simplified when $\mathbf{H}$ has a subgroup [see Eq. (14) in (I)].

3. AN EXAMPLE

Let us apply the above-mentioned method to a practical problems. We take an antiunitary double point group $\bar{\mathbf{D}}_4(\bar{\mathbf{D}}_2)\equiv\bar{\mathbf{C}}_4+\theta(\bar{\mathbf{D}}_4-\bar{\mathbf{D}}_2)$, θ being time inversion operator. The group $\bar{\mathbf{D}}_4(\bar{\mathbf{D}}_2)$ can be written as

$$\bar{\mathbf{D}}_4(\bar{\mathbf{D}}_2)=(E,C_2,\ \bar{E},\bar{C}_2)+C_{2x}(E,C_2,\bar{E},\bar{C}_2)$$
$$+\theta C_4[(E,C_2,\bar{E},\bar{C}_2)+C_{2x}(E,C_2,\bar{E},\bar{C}_2)].$$

To know the irreducible corepresentations of $\bar{\mathbf{D}}_4(\bar{\mathbf{D}}_2)$, it is sufficient only to show matrices for the elements E, C_2, $\bar{E}$, $\bar{C}_2$, C_{2x}, and θC_4, as listed in Table I.

Consider a product corepresentation $D_3\times D_3$ which is reducible to $2D_1+2D_2$. The matrix F_u of (4) is, in this case,

$$F_u\approx\begin{pmatrix}a_{11}&a_{12}&0&0\\0&0&a_{23}&a_{24}\\0&0&a_{33}&a_{34}\\a_{41}&a_{42}&0&0\end{pmatrix},$$

where a_{ij} are elements of an arbitrary matrix A. The symbol $\approx$ means that a common numerical factor to all the elements of the matrix is neglected. Thus Eq. (5) becomes

$$F\approx\begin{pmatrix}a_{11}&a_{12}&0&0\\0&0&a_{23}&a_{24}\\0&0&a_{33}&a_{34}\\a_{41}&a_{42}&0&0\end{pmatrix}$$

$$+\begin{pmatrix}0&0&0&1\\0&0&-1&0\\0&-1&0&0\\1&0&0&0\end{pmatrix}\begin{pmatrix}a_{11}^*&a_{12}^*&0&0\\0&0&a_{23}^*&a_{24}^*\\0&0&a_{33}^*&a_{34}^*\\a_{41}^*&a_{42}^*&0&0\end{pmatrix}\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$=\begin{pmatrix}a_{11}+a_{41}^*&a_{12}+a_{42}^*&0&0\\0&0&a_{23}-a_{33}^*&a_{24}-a_{34}^*\\0&0&a_{33}-a_{23}^*&a_{34}-a_{24}^*\\a_{41}+a_{11}^*&a_{42}+a_{12}^*&0&0\end{pmatrix}.$$

We can arbitrarily choose four quantities $a_{11}+a_{41}^*$, $a_{12}+a_{42}^*$, $a_{23}-a_{33}^*$, and $a_{24}-a_{34}^*$ so long as unitary conditions for F are satisfied. One choice is

$$a_{11}+a_{41}^*=1/\sqrt{2},\ a_{12}+a_{42}^*=i/\sqrt{2},$$

$$a_{23}-a_{33}^*=1/\sqrt{2},\ a_{24}-a_{34}^*=i/\sqrt{2}.$$

In such a way we obtain Table II for CG coefficients of $D_3\times D_3$ for the group $\bar{\mathbf{D}}_4(\bar{\mathbf{D}}_2)$.

4. CONCLUSION

Equation (3) for antiunitary groups is a direct extension of Eq. (12) in (I) for unitary groups. But we must notice that there is an essential difference as well as the apparent similarity of these equations. Whereas the matrix

$$\sum_r D^{(\alpha)}(r)AD^{(\alpha)}(r)^\dagger$$

for an irreducible representation $D^{(\alpha)}$ of a unitary group is scalar, the matrix

$$\sum_u D^{(\alpha)}(u)AD^{(\alpha)}(u)^\dagger+\sum_a D^{(\alpha)}(a)A^*D^{(\alpha)}(a)^\dagger$$

for an irreducible corepresentation $D^{(\alpha)}$ of an antiunitary group is not necessarily a scalar matrix.

The method presented in this paper can be applied to find CG coefficients for antiunitary space groups also. The conservation law of the reduced wave vectors will simplify the calculations of the matrix (3) as in the case of unitary groups [see Sec. 5 of (I)].

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Professor H. Watanabe and Dr. T. Iida for their helpful advices and discussions.

[1] I. Sakata, J. Math. Phys. 15, xxx (1974), preceding paper. This paper is refered to as (I).
[2] E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 325.

Phase transitions of a multicomponent Widom–Rowlinson model*

L. K. Runnels† and J. L. Lebowitz

Belfer Graduate School of Science, Yeshiva University, New York, N.Y. 10033
(Received 10 April 1974)

We study a multicomponent version of the "$A-B$" model of Widom and Rowlinson, generalized in a symmetric way: There is an infinite repulsive interaction between any two unlike particles. We consider both lattice and continuum versions of the model and show that the "demixing" transition occurs for any finite number M of components, all having the same activity. No conclusion can be drawn about this transition in the limit $M\to\infty$. It is shown, however, that another transition, in which the density is greater on one of the sublattices, appears at a finite value of M which persists for all larger M at any fixed value of the activity. In the limit $M\to\infty$, $z\to 0$, $Mz=\zeta$, const, this system apparently becomes "equivalent" to a one-component system with activity ζ in which there is an exclusion for occupancy of nearest neighbor sites. The latter transition then becomes the "hard square" transition.

I. INTRODUCTION

Widom and Rowlinson introduced a model[1] for fluid systems which has been quite fruitful. The model postulated two types of particles (A and B), each of which had no interaction with other molecules of the same kind. Between unlike molecules, however, there was an infinite repulsive interaction. Widom and Rowlinson discussed the thermodynamics and symmetry of the demixing transition predicted to occur at high activities, and also showed the equivalence of the two-component model with a one-component model having many-body forces.

Lebowitz and Gallavotti[2] constructed a lattice version of the $A-B$ model in which the $A-B$ interaction was $+\infty$ for separations of one (or zero) lattice units, and vanished otherwise. (This was their Model 1; other variations were also discussed.) The Peierls' contour argument[3] was employed to prove rigorously that the lattice version of the $A-B$ model does in fact have a demixing transition.

Ruelle[4] then extended the proof to the original continuum model, where the only nonzero interaction was an infinite repulsion if the separation between an A particle and a B particle became less than some fixed distance R. Lebowitz and Lieb[5] then showed that Ruelle's proof could be modified to cover the continuum case with a soft $A-B$ repulsion, at sufficiently low temperature. Physically these $A-B$ systems are analogous to the ferromagnetic transition in Ising spin systems and many results (e.g., inequalities) proven for the latter can be carried over to the former.

In theories of liquid crystals it is often convenient to identify the various possible orientations of the asymmetric molecules with the components of a mixture.[6] As far as the elongated cores of the molecules are concerned, the interactions between the "components" are repulsive and greater (in the sense of the excluded covolume of two molecules) for molecules with more dissimilar orientations. We have thus been led to consider a caricature of this situation in the form of a multicomponent Widom—Rowlinson model. The number M of components is arbitrary; particles interact only with dissimilar particles, and then repulsively but symmetrically: the identity of the unlike species is unimportant. To produce a more realistic model for liquid crystals it would be necessary for the repulsive interaction to vary with some appropriate measure of the difference in orientations of the molecules.

For our purpose the various species will simply be numbered $1, 2, \ldots, M$. We will primarily be concerned with the lattice version of the model, with infinite repulsion between unlike molecules occupying neighboring sites. For simplicity we explicitly discuss the two-dimensional case. Some observations about the continuum version will also be offered.

We will first show that the "demixing" transition of Widom and Rowlinson persists for any finite number of components, for either the lattice or the continuum version. Not surprisingly, our upper bound on the common critical activity of the components tends to infinity as $M\to\infty$.

For the lattice model, however, there is another transition that appears for large but finite M and remains at finite activities as $M\to\infty$. In this transition, the symmetry between the two sublattices is broken, one of them having a higher density of particles. We call this the crystal (or "hard square") transition due to its apparent relationship with the phase transition of the hard square lattice gas.[7] This transition has no analog in the continuum system—at least none that is demonstrable at the present time.

A related but not equivalent model is the M-state Potts model.[8] In its simplest form the model postulates M states for each lattice site, nearest neighbor interactions being zero for like states and $W\neq 0$ for unlike neighboring states. The "ferromagnetic" case $W>0$ has been most studied; the expected "Curie point" in zero field has been located as the self-dual temperature of the dual transformation. The ordered phases of the Potts model at low temperature are probably analogous to the almost-one-component phases of the present model at high activity, but the absence of a vacuum state in the former, i.e., empty sites which do not interact with any component, prevents an exact isomorphism.

II. THE MODELS

In the following two sections we will describe the de-

 Copyright © 1974 American Institute of Physics

mixing transition and the crystal transition. It is first useful to establish some general terms which will be used in all cases, lattice or continuum and for either transition in the lattice case.

The interaction potential between a particle of type i and one of type j, at a separation r, is given by

$$\varphi_{ij}(\mathbf{r}) = \begin{cases} 0 & \text{for } |r| > R, \\ +\infty & \text{for } |r| \leq R, \end{cases} \tag{1a}$$

for $i \neq j$, $1 \leq i$, $j \leq M$. For $i = j$ we have

$$\varphi_{ij}(\mathbf{r}) = \begin{cases} 0 & \text{for } \mathbf{r} \neq 0, \\ +\infty & \text{for } \mathbf{r} = 0. \end{cases} \tag{1b}$$

In the continuum case R is the "hard core diameter" between unlike particles; in the nearest neighbor lattice case R is simply the lattice spacing.

For any configuration of particles there is a unique decomposition of the particles into groups which we call *clusters*: Two particles belong to the same cluster if the particle configuration *requires* the two to be of the same type. Equivalently, given the set of particle locations, each cluster contains particles all of the same type.

In each case to be discussed there is a particular way of defining an "outer contour piece," γ. Once that is done we will denote by *boundary cluster of* γ a cluster containing a particle interior to γ whose center is no farther than R from γ.

III. DEMIXING TRANSITION

A. Lattice model

We consider a rectangular region Λ of the two-dimensional square lattice. Each site can be occupied by any of the M components—all of which have the same activity z. According to Eqs. (1), neighboring occupied sites must carry the same type of particle. We represent the particles as squares whose centers reside at the centers of the sites of the square lattice. If the lattice is completely filled, the corners of these squares define the dual lattice.

We employ the Peierls argument to show that there is an activity $z'(M)$ such that a phase transition occurs for some $z < z'(M)$. The technique is to impose a homogeneous boundary condition—say a band of particles of type 1 all around the perimeter of Λ—and show that this boundary condition prejudices the equilibrium state throughout Λ. Specifically, we can show that for some $z'(M)$, the total density of all components $j \neq 1$ is a *decreasing* function of z, whereas we know the total density is an increasing function of z.

The proof is virtually already done in Ref. 2. On rereading that proof (for Model 1), wherever "A" is mentioned, we read "component-1"; wherever "B" occurs, we read "other-than-component-1."

The only change is in the definition of a "cluster" and the multiplicity of the configuration transformation. "Cluster" is defined in the preceding section; it is the same as in Ref. 2 except that the translation of "B" to "other-than-component-1" is not quite accurate. The bound on the multiplicity, $m^{|G|}$, in Ref. 2 for the present model becomes $M^{3|G|}$. Each boundary cluster after the transformation is composed wholly of component-1 particles, whereas prior to the transformation it was of some other component. Clearly there are no more than $M^{3|G|}$ boundary clusters of outer boundary G.

This dependence of the multiplicity on M (see Eq. (3.5), Ref. 2) means that z' depends on M and in fact tends to infinity as $M \to \infty$.[9] We cannot therefore, make any statement about a phase transition for the limiting case $M \to \infty$.

B. Continuum model

In the continuous case again there is very little that needs to be changed from the two component proof of Ruelle.[4] Again, we reread Ruelle's proof, inserting "other-than-component-1" wherever "B" occurs. There is a change needed in the configuration transformation, however. In the two component case, to "remove" an outer contour piece it suffices to simply interchange interior A and B particles. In the present case we modify only particles in boundary clusters. We need the observation that if outer piece γ has length l (in units of "little" square edge length d), then the number of boundary clusters of γ cannot exceed $l/3$.

The change in the contour transformation of Ref. 4 is in its step (a) which is changed to read: "*All* particles in any boundary cluster of γ are changed to component 1." As in the lattice case this introduces a multiplicity to the transformation and changes the estimate of the probability $p(\gamma)$ of outer piece γ to

$$p(\gamma) \leq M^{l/3} \exp(-ld^2 z/2).$$

This probability replaces Eq. (2) of Ref. 4 in the rest of the development. We can then show that if the activity z is sufficiently high (depending on M) the expectation value of the density of other-than-component-1 particles is strictly less than that of component 1. Again, however, no conclusion can be drawn for the limiting case $M \to \infty$.

IV. CRYSTAL TRANSITION

We turn now to the "new" transition for the multicomponent model. In the previous section we discussed the demixing transition induced by high activity and the packing requirement that particles be of the same type in order to achieve high densities. The "driving force" behind the present transition is somewhat different. The idea is that for modest activities and large M the chance is small that nearest neighbor sites will be occupied. Instead the particles will preferentially occupy one of the sublattices, since when only one sublattice is occupied there is no restriction on the species occupying any site, with a subsequent gain in entropy. In this way the "ordered" state of this model is similar to that of the nearest neighbor exclusion problem on the square lattice. The latter system has a well-known transition associated with sublattice ordering.[7] It must be shown, however, that this transition does actually occur at a bounded activity for finite M and persists in some well-defined sense as $M \to \infty$. We also suspect (for fixed large M) an upper activity limit on the stability of this sublattice ordering.

We use the technique employed by Dobrushin[7b] to prove the nonuniqueness of the equilibrium state for the nearest neighbor exclusion problem. Specifically, we shall show that for any positive activity z there is an $M_0 = M_0(z)$ such that the multicomponent lattice model with at least M_0 components has the crystal transition. The criterion is that M_0/z^4 be sufficiently large if $z \geq 1$ or that $M_0 z$ be sufficiently large if $z < 1$. The latter case has a limit which we believe represents the hard square system: $M \to \infty$, $z \to 0$, $Mz = \zeta =$ activity of the hard square gas. The existence and identification of this limit can be proven explicitly in one dimension. (There is, of course, no phase transition in one dimension.) Combining this result with that of the previous section we conclude that the multicomponent model with large but finite number of components has two quite different ordered phases at finite activities. We do not have a very useful estimate of the minimum value of M for which both may be observed.

A. Definitions

To facilitate the proof it is convenient to introduce the following definitions, which are illustrated in Fig. 1.

1. Contour segments: "bonds" of the square lattice, dual to the lattice of sites, separating two sites which are both empty or both occupied. (If both are occupied, then both particles must be of the same type.) This definition differs from that of Peierls (for the ferromagnetic Ising model) but is the same as used by Dobrushin[7b] for the antiferromagnetic Ising model and the hard square lattice gas.

2. Contour Γ: union of all contour segments, consisting of various connected components.

3. Pieces γ_i: smallest set of connected components of Γ, such that if two connected components are separated by a distance of no more than R (= lattice spacing) they belong to the same piece. Γ is then the union of the disjoint pieces $\gamma_1, \ldots, \gamma_n$.

4. Outer piece γ: one of the pieces such that there is a path from the boundary reaching a segment of γ without crossing Γ.

5. Interior site x: a lattice site such that a path from the boundary crosses γ an odd number of times before reaching x. Otherwise a site is exterior to γ.

6. Boundary conditions, with checkerboard coloring of square lattice with black and white squares: white boundary condition means black squares on outer *two* rows and columns are *vacant*. White squares on very outer-most rows and columns are populated arbitrarily, i.e. each site contains any one of the M species. Black boundary condition: white squares are vacant on two outer rows and columns and black squares on very outer-most rows and columns are populated arbitrarily. See the comment below in subsection B about these boundary conditions.

7. Bottom segment of piece γ: a horizontal segment adjacent to and beneath an interior site of γ. Any other horizontal segment is a top segment.

8. Distinguished sites of a configuration X producing contour Γ with outer piece γ:

a. A-site (annihilation): interior occupied site beneath a top segment.

b. L-site (liberated): an interior vacant site above a bottom segment (L_0-site); or an exterior occupied site adjacent to a bottom segment or a vertical segment of γ (L_1-site).

c. G-site (generator): interior occupied site adjacent to a bottom or vertical segment, but not adjacent to a top segment (G_0-site); or exterior occupied site adjacent to a top segment but not adjacent to a bottom or vertical segment (G_1-site). *Note*: Accordingly, every *occupied* site adjacent to a contour segment has exactly one kind of distinguished site designation. An interior occupied site is an A-site if adjacent to a top segment and is otherwise a G_0-site. An exterior occupied site is an L_1-site if adjacent to a bottom or vertical segment and is otherwise a G_1-site.

9. Cluster and boundary cluster: as defined in Sec. II.

In connection with these definitions we will need the following two observations, which we state as lemmas.

Lemma 1: Any A-site or L_1-site of an outer piece γ belongs to some boundary cluster with two or more particles.

Proof: By definition a particle on either type of site is adjacent to a contour segment which must have an occupied site on the other side.

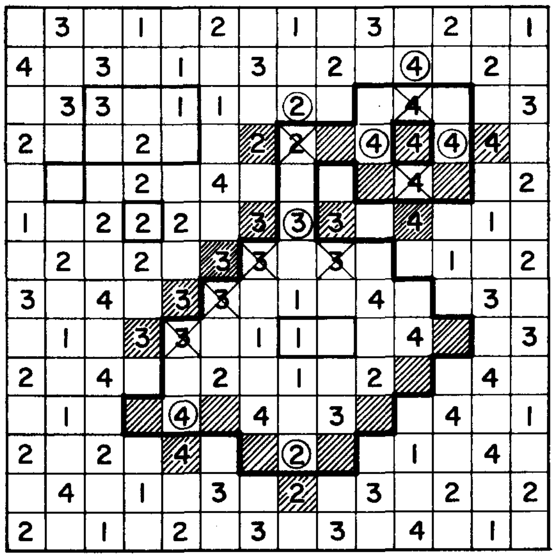

FIG. 1. A configuration whose contour consists of three pieces, two of which are outer pieces. The heaviest lines are the contour segments of one piece γ_1, consisting of two connected components. The numerals represent particle illustration. Distinguished sites associated with outer piece γ_1 are identified as follows: A-site particles are x'd, G-site particles are circled, and L-sites are shaded. The particle of type 4 at the center of the "square doughnut" portion enclosed by γ_1 is exterior to γ_1. The particle of type 1 contained in the small piece enclosed by γ_1 is interior to γ_1. The piece γ_1 has five boundary clusters; each must contain at least one G-site, according to Lemma 2. It should be noticed that the sites of one sublattice are vacant on the two outer rows and columns.

Lemma 2: Given an outer piece γ, one of whose interior sites belongs to a boundary cluster C, the cluster C must contain at least one G-site.

Proof: Regarding only the cluster C, we locate its highest site or highest set of contiguous sites. (In case of ties any highest site or highest set of contiguous sites will do.) If there is a highest single site, the site below it must be occupied and the other three neighboring sites empty. Hence the site in question must be either: (a) interior, adjacent to a bottom segment and not to a top segment, or (b) exterior, adjacent to a top segment and not to a bottom or vertical segment. In the first case the site is a G_0-site, and in the second case it is a G_1-site. If there is not a highest single site we consider the highest set of contiguous sites. The site above each must be vacant and alternate members of the set of contiguous sites must be interior sites, and adjacent to vertical contour segments. Each such interior site is a G_0-site.

B. Configuration transformation

We now define a one-to-many transformation among the allowed configurations on Λ. With configuration X producing contour Γ having outer piece γ we associate a class of X^* of configurations, in three steps:

a. particles at A-sites are annihilated;

b. all remaining particles at sites interior to γ are displaced upward by one unit;

c. L-sites are arbitrarily populated.

We notice that the inverse transformation is well defined by virtue of the original configurations at the G-sites and Lemmas 1 and 2. (The G_0-site particles have been displaced upward one unit by step *b*.) This means that for any configurations Y and X producing the same outer piece γ, $Y^* \cap X^* = \emptyset$ if $Y \neq X$. In the third step (arbitrary population of L-sites) alterations are made in the occupancy of some sites exterior to γ (the L_1-sites). According to our definitions these sites do not belong to boundary clusters of any different piece γ', since that would require γ' to be within one lattice spacing of γ and hence united into one piece. The boundary condition consists of vacant sites and thus an L_1-site is never part of the boundary condition.

Figure 2 shows the class of configurations produced by this transformation from the configuration shown in Fig. 1.

C. Probability of outer contour piece

We can now calculate a bound on the grand canonical probability of an outer piece γ, in the following steps.

1. Length. If γ contains l segments and c connected components, it can be traversed by a k-step lattice walk, where $k \leqslant l + 2(c-1)$. Since $l \geqslant 4c$ we have $k \leqslant 3l/2$.

2. Number of L-sites. Let n_V and n_H denote the number of vertical and horizontal segments, respectively, of γ. Let n_{L_0}, n_{L_1}, n_L denote the number of L_0-sites, L_1-sites, L-sites, respectively. Half of the horizontal segments produce an L-site (each of those at the bottom), so $n_L \geqslant n_H/2 \geqslant l/4$ *if* $n_H \geqslant n_V$. Notice that an L-site cannot thereby by counted twice. If, however $n_V > n_H$, we first notice that each vertical segment is followed (in a circuit around a connected component of γ) either (a) by another vertical segment or (b) by a horizontal segment. In case (a) one of the two vertical segments must produce an L_1-site, while in case (b) the horizontal segment produces "half" of an L-site (it might be a top segment, but there must be as many bottom segments as top segments). By this method of counting it is possible for some L_1-sites to be counted twice, so we can only conclude that $n_L \geqslant n_V/4 > l/8$. Regardless of the ratio n_H/n_V we can always assert that $n_L > l/8$.

3. Number of A-sites. Since A-sites occur only beneath top segments, which must be equaled in number by bottom segments, we clearly know that the number of A-sites, n_A, cannot exceed $l/2$.

4. Probability of outer piece γ. Let Z denote the partition function (with all M components having the same activity z), let $Z(\gamma)$ denote the partition function restricted to configurations X producing a contour with outer piece γ, and let Z_{X^*} denote the sum over configurations in class X^* derived from configuration X by the transformation defined above. We have then that the probability $p(\gamma)$ of outer piece γ is

$$p(\gamma) = Z(\gamma)/Z \leqslant \frac{\sum_{X \supset \gamma} z^{N(X)}}{\sum_{X \supset \gamma} Z_{X^*}}. \tag{2}$$

Here $N(X)$ is the total number of particles present in configuration X. As pointed out previously the inverse transformation $X^* \to X$ is unique so there is no overcounting. The L-sites are independent and in Z_{X^*} each contributes a factor $1 + Mz$; in the configuration X each L_1-site contributed a factor z while the L_0-sites contributed the factor 1. Finally each A-site had a factor z in X and has a factor 1 in Z_{X^*}.

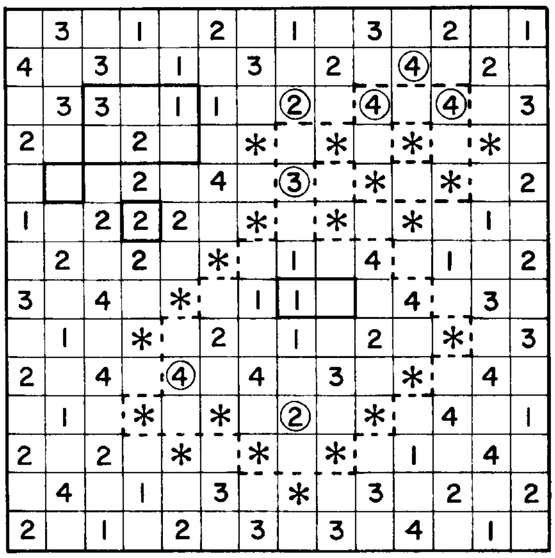

FIG. 2. The effect of the configuration transformation on the configuration shown in Fig. 1. Sites labeled $*$ are arbitrarily occupied by any of the species. Given the outer contour γ_1, here shown dotted, the configuration before the transformation had to be the one shown in Fig. 1; it can be reconstructed from the G-sites.

Thus we have

$$z^{N(X)}/z_{X^*} = z^{n_{L1}} z^{n_A}/(1+Mz)^{n_L}$$
$$\leqslant (1+Mz)^{-n_L}$$
$$\leqslant (Mz)^{-l/8} \quad \text{for } z<1. \tag{3a}$$

On the other hand, if $z \geqslant 1$ we have

$$z^{N(X)}/Z_{X^*} \leqslant z^{n_L} z^{n_A}/(1+Mz)^{n_L}$$
$$\leqslant M^{-n_L} z^{n_A}$$
$$\leqslant M^{-l/8} z^{l/2}$$
$$=(M/z^4)^{-l/8} \quad \text{for } z \geqslant 1. \tag{3b}$$

According to Eq. (2) the right-hand sides of Eq. (3) are also upper bounds on $p(\gamma)$.

D. Nonuniqueness

The standard arguments will be used to show that for white boundary conditions the probability that a black square is occupied may be made arbitrarily small, by choosing sufficiently large values of M. However, to complete the demonstration of nonuniqueness we must show that the *total* density is bounded below for fixed z as $M \to \infty$.

Imagine the lattice Λ paved with "Red Cross symbols" of five sites, and focus on one (K) consisting of the site (x, y) and the four neighboring sites $(x \pm 1, y)$ and $(x,\ y \pm 1)$.

Lemma 3: *For any configuration on* $\Lambda \setminus K$, the expected number of particles n_K in K is no less than $Mz/(1+Mz)$.

Proof: For any configuration on $\Lambda \setminus K$, the partition function on K has the form

$$\xi_K = 1 + \sum_{i=1}^{5} a_i z^i$$

where $a_i \geqslant 0$, and the expected number of particles on K is $n_K = z(\partial \ln \xi_K/\partial z)$. Now by algebra we show

$$n_K \geqslant a_1 z/(1+a_1 z) \geqslant \alpha^z/(1+\alpha z)$$

for any $\alpha \leqslant a_1$. We can always take $\alpha = M$ [from the configuration with (x, y) occupied and the other four sites empty]. Since this holds for any configuration $\Lambda \setminus K$ we know that in K the average density must not be less than $(1/5)Mz/(1+Mz)$. The same reasoning applies to each of the other "Red Cross symbols" paving Λ and so we obtain the lower bound on the total density p_t,

$$p_t \geqslant (1/5)Mz/(1+Mz). \tag{4}$$

Now with white boundary conditions if a black square is occupied it must be enclosed in some outer contour piece γ. Equations (3) give upper bounds on the probability $p(\gamma)$. There are no more than $(k/4)^2\, 3^{k-2}$ pieces of length l around any given site, where $k = 3l/2$ is the upper bound on the length of a lattice walk circumnavigating γ. This means that the probability p_b that a black site is occupied is bounded above:

$$p_b \leqslant \frac{1}{36} \sum_{j=2}^{\infty} j^2 y^j = \frac{y^2(4-3y+y^2)}{36(1-y)^3} \tag{5}$$

for $y<1$. Here $j = k/2$ and $y = 9/(Mz)^{1/6}$ for $z<1$, while $y = 9(z^2/M)^{1/6}$ for $z \geqslant 1$.

To demonstrate the influence of the white boundary conditions we must show that $p_b < p_t/2$ for sufficiently large M. For any $z>0$, Eqs. (4) and (5), together with $y<1$, yield a minimum value M_0 for which $p_b < p_t/2$ is satisfied. For $z<1$ the requirement is that the product $M_0 z$ be sufficiently large, while for $z \geqslant 1$ the requirement is that M_0/z^4 be sufficiently large. Numerically Eqs. (4) and (5) are not very helpful for determining the minimum M_0 for which this transition would be observed. (They show that $M_0 \approx 27^6$ is sufficiently large!)

It seems likely, but is not proven, that for fixed large M and increasing activity z, the "hard square" sublattice ordering will break down *before* the demixing phase separation occurs. That is, we expect the "phase diagram" to appear as shown schematically in Fig. 3. We have actually proven only that phase transition lines lie below the "Crystal" region and to the left of the "Demixing" region. If these phase transition lines have the same general shape as shown in Fig. 3, then the above assertion would be correct. With increasing activity, then, such a system would undergo *three* phase transitions.

E. The "hard square" limit

The case $z<1$ is particularly interesting. In this case the variables M and z enter Eqs. (4) and (5) only as the product Mz. This is consistent with the statement that the present model becomes isomorphic to the hard square lattice gas in the limit $M \to \infty$, $z \to 0$, $Mz = \zeta$ = activity of the one-component hard square lattice gas.

To be more precise we believe that in the above limit the thermodynamic properties of the system as well as its "equilibrium measure" defined on the set of "equivalence classes of configurations" $\tilde{\mathcal{A}}$ becomes the same as for the hard square system. Two configurations X and Y belong to the same equivalence class $\tilde{X} \in \tilde{\mathcal{A}}$ if they have the same set of occupied sites, i.e., if they differ only by the labeling of the species at each occupied site. To see how such an isomorphism would come about we note

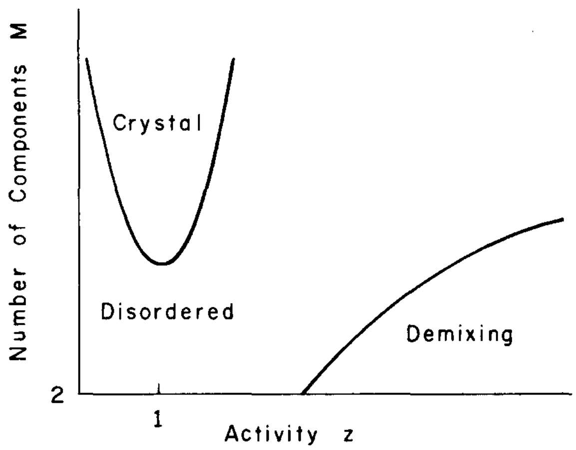

FIG. 3 ("Phase Diagram"). Shown schematically are the lines proven to lie within regions belonging to the two types of ordered phases: the crystal ("hard square") phase and the "demixed" phases of predominantly one component. The actual extent of the incursion of the disordered region into these two areas is not known.

first that a "fully restricted system" of M components in which no adjacent sites can be occupied, i.e., in which Eq. (1a) hold for all i and j, is obviously isomorphic, in the sense defined above to the one-component hard square system with fugacity $\zeta = Mz$. It seems reasonable to expect that in the limit $M\to\infty$, $z\to 0$, $Mz=\zeta$, the multicomponent Widom–Rowlinson model considered in this paper has the same property. We give an explicit computation of the thermodynamic properties for a one-dimensional system in the Appendix.

APPENDIX: ONE-DIMENSIONAL LATTICE SYSTEMS

Consider a one-dimensional lattice containing L sites, $L \geq 3$, with periodic boundary conditions. (Similar results hold for other boundary conditions.) Let $Z_\alpha(z,M;L)$ $\alpha = 0, 1, 2$, be the partition function for the "hard rod" system ($M\equiv 1$), the fully restricted system, and the Widom–Rowlinson model, considered in this paper, respectively. In all cases

$$Z_\alpha(z,M;L) = \operatorname{tr} T_\alpha^L = \sum_{k=1}^{M+1} \lambda_k^L(\alpha, z, M)$$

where T_α is the transfer matrix. T_α is a symmetric matrix of dimensionality $M+1$ (with $M\equiv 1$ for $\alpha = 0$) and $\lambda_k(\alpha, z, M)$ are its eigenvalues. The forms of these matrices are

$$T_0 = \begin{pmatrix} 1 & \sqrt{z} \\ \sqrt{z} & 0 \end{pmatrix}, \quad 2\times 2,$$

$$T_1 = \begin{pmatrix} 1 & \sqrt{z} & \cdots & \sqrt{z} \\ \sqrt{z} & & & \\ \cdot & & & \\ \cdot & & \bigcirc & \\ \cdot & & & \\ \sqrt{z} & & & \end{pmatrix}, \quad (M+1)\times(M+1)$$

$$T_2 = \begin{pmatrix} 1 & \sqrt{z} & \cdots & \sqrt{z} \\ \sqrt{z} & z & & \bigcirc \\ & & \ddots & \\ \sqrt{z} & \bigcirc & & z \end{pmatrix}, \quad (M+1)\times(M+1)$$

with eigenvalues

$$\lambda_{1,2}(0,z) = [1 \pm (1+4z)^{1/2}]/2,$$

$$\lambda_{1,2}(1,z,M) = [1 \pm (1+4Mz)^{1/2}]/2, \quad \lambda_k(1,z,M) = 0, \quad k = 3, \ldots, M+1,$$

$$\lambda_{1,2}(2,z,M) = \{1 + z \pm [(1+z)^2 + 4(M-1)z]^{1/2}\}/2, \quad \lambda_k(2,z,M) = z, \quad k = 3, \ldots, M+1.$$

In the limit $z\to 0$, $M\to\infty$, $Mz=\zeta$, we clearly have $Z_\alpha(z,M;L)\to Z_0(\zeta, L)$ for $\alpha = 1, 2$ (and we have omitted the M from Z_0). The same thing happens if we first take the thermodynamic limit $L\to\infty$ of the pressure $L^{-1}\ln Z_\alpha$ and then take the limit on z and M. The isomorphism of the equilibrium measures (as defined at the end of the paper) can presumably also be shown readily for the one-dimensional system and probably remains valid also in higher dimensions.

*Research supported in part by National Science Foundation Grant No. GP-3353X and by AFOSR Grant No. 73-2430A.
†Permanent address: Department of Chemistry, Louisiana State University, Baton Bouge, Louisiana 70803.
[1]B. Widom and J.S. Rowlinson, J. Chem. Phys. **52**, 1670 (1970).
[2]J.L. Lebowitz and G. Gallavotti, J. Math. Phys. **12**, 1129 (1971).
[3]See, for example, R.B. Griffiths in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M.S. Green (Academic, New York, 1972), Vol. I, p. 59.
[4]D. Ruelle, Phys. Rev. Lett. **27**, 1040 (1971).
[5]J.L. Lebowitz and E.H. Lieb, Phys. Lett. **A39**, 98 (1972); see also L.K. Runnels and B.C. Freasier, Phys. Rev. **A8**, 2126 (1973).
[6]L. Onsager, Ann. N.Y. Acad. Sci. **51**, 627 (1949).
[7](a) L.K. Runnels, Phys. Rev. Lett. **15**, 581 (1965); (b) R.L. Dobrushin, Funct. Anal. Appl. **2**, 302 (1968).
[8]R.J. Baxter, "Potts Model at the Critical Temperature," preprint; M.J. Stephen and L. Mittag, Phys. Lett. **A41**, 357 (1972); J.P. Straley and M.E. Fisher, J. Phys. **A6**, 1310 (1973).
[9]From our bound on the multiplicity and Eq. (3.5) of Ref. 2 it follows that a bound of the form $z' > (\text{const})\times M^{12}$ is sufficiently large. It is unlikely that an optimal bound would increase so rapidly with M.

Matrix mechanics approach to a nonlinear oscillator*

Francis R. Halpern and Thomas W. Yonkman

Department of Physics, University of California, San Diego, La Jolla, California 92037
(Received 31 January 1974)

The system with Hamiltonian p^2+x^4 is discussed. An approximation scheme is given for matrices x and p that satisfy the canonical commutation relation and diagonalize the Hamiltonian.

I. INTRODUCTION

The Hamiltonian to be treated is

$$H=p^2+x^4.$$

We look for a canonical pair $[p,x]=-i$ such that H is diagonal. Our scheme is the following. We assume that the most important matrix elements of x and p are those of the form $\langle n|x|n\mp 1\rangle$ and $\langle n|p|n\pm 1\rangle$. In our lowest approximation we treat only these matrix elements and we find these in the way that comes as close as possible to making H diagonal and the operators x and p a canonical pair. We treat three types of equations: (1) the diagonal commutator equation $C_{aa}=(px-xp)_{aa}$, (2) the off-diagonal commutator equations $C_{ab}=(px-xp)_{ab}$ with $a\neq b$, (3) the off-diagonal Hamiltonian equations H_{ab} with $a\neq b$.

We find the approximation leads to a double expansion of the matrix elements x_{ab}, p_{ab}. The first dimension of the expansion is the difference $a-b$. In this expansion we find that $x_{ab}=A_{a-b}(a+b)^{1/3}$, $p_{ab}=B_{a-b}(a+b)^{2/3}$. There is a second expansion involving terms of lower order in $(a+b)$. We have not explored these terms. We assume that x will have the form $A_{a-b}(a+b)^{1/3}+\bar{A}_{a-b}(a+b)^{-2/3}+\cdots$. We have found empirically that A_{a-b} and B_{a-b} are very rapidly decreasing functions of $a-b$.

In Sec. II the lowest approximation is carried out. In the Appendix the expansions of the commutator and Hamiltonian to leading order in $(a+b)$ are derived. These results are used in Sec. III to give the equations for the second, third, and fourth approximations. The solutions to these equations found by a computer are reported in this section. Section IV is our evaluation of the status of our work.

Anharmonic oscillators of the present type are of interest because they are the simplest examples of nonlinear problems in either classical or quantum mechanics. They serve as models in a restricted way for field theories and many body problems. They have been treated by a variety of methods since the earliest days of quantum mechanics.[1] Roughly speaking, methods can be divided up into a number of types:

I. Those methods which strive for numerical precision.[2] The principal technique is to truncate the problem and diagonalize a large but finite matrix. A variety of styles of truncation have been employed. These methods converge.

II. Development of the perturbation series and efforts to reorder it.[3] The perturbation series is divergent but rearrangements such as the Padé approximates converge and also give good numerical results.

III. Approaches like this one which attempt to determine the most significant matrix element of the canonical operators and their dependence on the matrix indices.[4] The WKB method probably fits most naturally with this class of approximation.

II. THE FIRST APPROXIMATION EQUATIONS

In this section the lowest approximation is carried out in a direct way. The equations of this approximation are rederived in the next section using the more formal procedures of the Appendix.

We assume that p_{ab} and x_{ab} are zero if $a-b\neq 1$. The formulas

$$x_{ab}=\xi_a\delta_{a,b+1}+\xi_{a+1}\delta_{a,b-1},$$
$$ip_{ab}=-\pi_a\delta_{a,b+1}+\pi_{a+1}\delta_{a,b-1}$$

include this selection rule. The real symmetric form of x and imaginary antisymmetric form for p follow from time reversal invariance. The commutator C and the Hamiltonian H are found to be

$$C_{ab}=-i[(\xi_a\pi_{a-1}-\xi_{a-1}\pi_a)\delta_{a-2,b}+2(\xi_{a+1}\pi_{a+1}-\xi_a\pi_a)\delta_{a,b} + (\xi_{a+2}\pi_{a+1}-\xi_{a+1}\pi_{a+2})\delta_{a+2,b}],$$

$$\begin{aligned}H_{ab}={}&\xi_{a-3}\xi_{a-2}\xi_{a-1}\xi_a\delta_{a-4,b}\\&+(-\pi_a\pi_{a-1}+\xi_{a-2}^2\xi_{a-1}\xi_a+\xi_{a-1}^3\xi_a+\xi_{a-1}\xi_a{}^3+\xi_{a-1}\xi_a\xi_{a+1}^2)\delta_{a-2,b}\\&+(\pi_a^2\pi_{a+1}^2+\xi_a^2\xi_{a-1}^2+(\xi_a^2+\xi_{a+1}^2)^2+\xi_{a+1}^2\xi_{a+2}^2)\delta_{a,b}\\&+(-\pi_{a+1}\pi_{a+2}+\xi_a^2\xi_{a+1}\xi_{a+2}+\xi_{a+1}^3\xi_{a+2}+\xi_{a+1}\xi_{a+2}^3\\&+\xi_{a+1}\xi_{a+2}\xi_{a+3}^2)\delta_{a+2,b}\\&+\xi_{a+1}\xi_{a+2}\xi_{a+3}\xi_{a+4}\delta_{a+4,b}.\end{aligned}$$

We focus our attention on the diagonal commutator equation

$$\xi_{a+1}\pi_{a+1}-\xi_a\pi_a=\tfrac{1}{2}$$

and the first off-diagonal Hamiltonian equation

$$-\pi_a\pi_{a-1}+\xi_a\xi_{a-1}\xi_{a-2}^2+\xi_a\xi_{a-1}^3+\xi_a^3\xi_{a-1}+\xi_{a+1}^2\xi_a\xi_{a-1}=0.$$

The commutator equation can be summed to give

$$\pi_a\xi_a=\tfrac{1}{2}a$$

Within the scope of our approximation we neglect the differences between π_a and π_{a-1} and between ξ_{a-2}, ξ_{a-1}, ξ_a, ξ_{a+1}. The equations of the first approximation become

$$\pi_a\xi_a=\tfrac{1}{2}a, \quad 4\xi_a^4=\pi_a^2.$$

The solution is

$$\xi_a=\tfrac{1}{2}(2a-1)^{1/3}, \quad \pi_a=\tfrac{1}{2}(2a-1)^{2/3}.$$

The -1 is included to make our later results neater. Since we will only work to leading order in $(2a-1)^{1/3}$ or $(2a-1)^{2/3}$ the change by 1 is not important.

Within this approximation the energy or diagonal

 Copyright © 1974 American Institute of Physics

matrix element of the Hamiltonian is given by

$$E_a = H_{aa} = 2\pi^2_{a+1/2} + 6\xi^4_{a+1/2}$$
$$= [2(\tfrac{1}{2})^2 + 6(\tfrac{1}{2})^4](2a)^{4/3}$$
$$= \tfrac{7}{8}(2a)^{4/3}.$$

III. EQUATIONS FOR THE SECOND, THIRD, AND FOURTH APPROXIMATIONS

In this section we use the expressions in the Appendix to find the equations of the next three orders of approximation.

In the second approximation we include all terms with A_1, A_3, B_1, and B_3. The equations we solve are $C_{aa} = -i$, $C_{aa+2} = 0$, $H_{aa+2} = 0$, $H_{aa+4} = 0$. From the Appendix the diagonal commutator equation is

$$-2i \sum_{s=a-3}^{a+3} (a-s)B_{a-s}A_{a-s} = -i$$

or

$$-2i(3A_3B_3 + A_1B_1 - B_{-1}A_{-1} - 3B_{-3}A_{-3}) = -i;$$

using the symmetry of the A's and B's this becomes

$$3A_3B_3 + A_1B_1 - \tfrac{1}{4} = 0.$$

Since ξ_a and π_a of Sec. II are related to A_1 and B_1 by

$$\xi_a = A_1(2a-1)^{1/3}, \quad \pi_a = B_1(2a-1)^{2/3},$$

the earlier equation

$$\xi_a \pi_a = \tfrac{1}{2}a$$

is equivalent to

$$A_1B_1(2a-1) = \tfrac{1}{2}a$$

or

$$A_1B_1 = \tfrac{1}{4}$$

the second order equation with $A_3 = B_3 = 0$.

The first off-diagonal commutator equation is

$$\sum [6s - 4(a+2) - 2a] B_{a-s} A_{s-a-2} = 0.$$

The appropriate limits on the sum are $a-1 \leqslant A \leqslant a+1$, which gives

$$-7B_1A_{-3} - B_{-1}A_{-1} + 5B_{-3}A_1 = 0$$

or

$$5B_3A_1 - A_1B_1 + 7B_1A_3 = 0.$$

The two Hamiltonian equations are $H_{aa+2} = 0$, $H_{aa+4} = 0$. If we substitute in from the Appendix, the required terms are

$$H_{a,a-2} = -(2a+2)^{4/3} \sum B_{a-s}B_{s-a-2} + (2a+2)^{4/3} \sum A_{a-q}A_{q-r}A_{r-s}A_{s-a-2},$$

$$H_{a,a+4} = -(2a+4)^{4/3} \sum B_{a-s}B_{s-a-4} + (2a+4)^{4/3} \sum A_{a-q}A_{q-r}A_{r-s}A_{s-a-4}.$$

The remaining problem is to determine the range of s in the first sum and of q, r, and s in the second sum so that only A_1A_3 and B_1 and B_3 occur. In this order it is simple enough to do this by inspection. We find that

$$H_{a,a+2} = -(2a+2)^{4/3}(B_1B_{-3} + B_{-1}B_{-1} + B_{-3}B_1) + (2a+2)^{4/3}(4A_1^3A_{-1} + 12A_{-1}^2A_1A_3 + 12A_{-3}A_{-1}A_3^2 + 12A_{-3}A_3A_1^2),$$

$$H_{a,a+4} = -(2a+4)^{4/3}(B_{-1}B_{-3} + B_{-3}B_{-1}) + (2a+4)^{4/3}(A_1^4 + 12A_1^2A_{-1}A_3 + 12A_3^2A_{-3}A_1 + 6A_{-1}^2A_3^2).$$

The equations of the second approximation are

$$A_1B_1 + 3A_3B_3 - \tfrac{1}{4} = 0,$$
$$5A_1B_3 - A_1B_1 + 7A_3B_1 = 0,$$
$$-B_1^2 + 2B_1B_3 + 4A_1^4 + 12A_1^3A_3 + 12A_1^2A_3^2 + 12A_1A_3^3 = 0,$$
$$-2B_1B_3 + A_1^4 + 12A_1^3A_3 + 6A_1^2A_3^2 + 12A_1A_3^3 = 0.$$

The numerically determined solution is:

$$A_1 = 0.461046, \quad B_1 = 0.531778$$
$$A_3 = 0.0230266, \quad B_3 = 0.0691887$$

The principal difficulty in developing the higher order equations comes from evaluating the x^4. The techniques of partition theory lead readily to a generating function that makes the calculation accessible. Consider the expression

$$X = (x_{-7} + x_{-5} + x_{-3} + x_{-1} + x_1 + x_3 + x_5 + x_7)^4.$$

There are 8^4 terms in the expansion of X. These terms can be segregated according to the sum of the subscripts which run from -28 to $+28$. A term such as $12x_{-7}x_7x_1^{\;2}$ with a subscript sum 2 corresponds to a term $12A_7^{\;2}A_1^{\;2}$ in H_{aa+2}. The same technique can also be employed in expanding p^2. Using this method the third and fourth order equations are:

Third approximation

$$A_1B_1 + 3A_3B_3 + 5A_5B_5 - \tfrac{1}{4} = 0,$$
$$13A_5B_3 + 7A_3B_1 - A_1B_1 + 5A_1B_3 + 11A_3B_5 = 0,$$
$$11A_5B_1 - 5A_3B_1 + A_1B_3 + 7A_1B_5 = 0,$$
$$-B_1^2 + 2B_1B_3 + 2B_3B_5 + 4A_1^4 + 12A_1^3A_3 + 12A_1^2A_3^2 + 12A_1A_3^3 + 4A_1^3A_5 + 24A_1^2A_3A_5 + 12A_1^2A_5^2 + 12A_1A_3^2A_5 + 24A_1A_3A_5^2 + 12A_3^3A_5 + 12A_3A_5^3 = 0,$$
$$-2B_1B_3 + 2B_1B_5 + A_1^4 + 12A_1^3A_3 + 6A_1^2A_3^2 + 12A_1A_3^3 + 12A_1^3A_5 + 12A_1^2A_3A_5 + 24A_1A_3^2A_5 + 24A_1A_3A_5^2 + 12A_1A_5^3 + 4A_3^3A_5 + 6A_3^2A_5^2 = 0,$$
$$-B_3^2 - 2B_1B_5 + 4A_1^3A_3 + 12A_1^2A_3^2 + 12A_1^3A_5 + 12A_1^2A_3A_5 + 24A_1A_3^2A_5 + 12A_1A_3A_5^2 + 12A_1A_5^3 + 12A_3^2A_5^2 + 4A_3^4 = 0$$

Fourth approximation

$$A_1B_1 + 3A_3B_3 + 5A_5B_5 + 7A_7B_7 - \tfrac{1}{4} = 0,$$
$$19A_7B_5 + 13A_5B_3 + 7A_3B_1 - A_1B_1 + 5A_1B_3 + 11A_3B_5 + 17A_5B_7 = 0,$$
$$17A_7B_3 + 11A_5B_1 - 5A_3B_1 + A_1B_3 + 7A_1B_5 + 13A_3B_7 = 0,$$
$$15A_7B_1 - 9A_5B_1 - 3A_3B_3 + 3A_1B_5 + 9A_1B_7 = 0,$$
$$-B_1^2 + 2B_1B_3 + 2B_3B_5 + 2B_5B_7 + 4A_1^4 + 12A_1^3A_3 + 12A_1^2A_3^2$$

$$+ 12A_1A_3^3 + 4A_1^3A_5 + 24A_1^2A_3A_5 + 12A_1^2A_5^2 + 12A_1A_3^2A_5$$
$$+ 24A_1A_3A_5^2 + 12A_3^3A_5 + 12A_3A_5^3 + 12A_1^2A_3A_7$$
$$+ 24A_1^2A_5A_7 + 12A_1^2A_7^2 + 12A_1A_3^2A_7 + 24A_1A_3A_5A_7$$
$$+ 24A_1A_3A_7^2 + 12A_1A_5^2A_7 + 4A_3^3A_7 + 24A_3^2A_5A_7$$
$$+ 24A_3A_5A_7^2 + 12A_5^3A_7 + 12A_5A_7^3 = 0,$$
$$- 2B_1B_3 + 2B_1B_5 + 2B_3B_7 + A_1 + 12A_1^3A_3$$
$$+ 6A_1^2A_3^2 + 12A_1A_3^3 + 12A_1^3A_5 + 12A_1^2A_3A_5$$
$$+ 24A_1A_3^2A_5 + 24A_1A_3A_5^2 + 12A_1A_5^3 + 4A_3^3A_5 + 6A_3^2A_5^2$$
$$+ 4A_1^3A_7 + 24A_1^2A_3A_7 + 12A_1^2A_5A_7 + 24A_1A_3A_5A_7$$
$$+ 12A_1A_5^2A_7 + 24A_1A_3A_7^2 + 24A_1A_5A_7^2 + 12A_3^3A_7$$
$$+ 12A_3^2A_5A_7 + 24A_3A_5^2A_7 + 12A_3A_7^3 + 6A_5^2A_7^2 = 0,$$
$$- B_3^2 - 2B_1B_5 + 2B_1B_7 + 4A_1^3A_3 + 12A_1^2A_3^2 + 12A_1^3A_5$$
$$+ 12A_1^2A_3A_5 + 24A_1A_3^2A_5 + 12A_1A_3A_5^2 + 4A_3^4 + 12A_3^2A_5^2$$
$$+ 12A_1^3A_7 + 12A_1^2A_3A_7 + 24A_1A_3^2A_7 + 12A_1A_5^3$$
$$+ 24A_1A_3A_5A_7 + 24A_1A_5^2A_7 + 24A_1A_5A_7^2 + 12A_1A_7^3 + 12A_3^2A_5A_7$$
$$+ 12A_3A_5^2A_7 + 12A_3^2A_7^2 + 12A_3A_5A_7^2 = 0,$$
$$- 2B_3B_5 - 2B_1B_7 + 6A_1^2A_3^2 + 4A_1A_3^3 + 4A_1^3A_5 + 6A_1^2A_5^2$$
$$+ 12A_1A_3A_5^2 + 12A_3^3A_5 + 12A_3A_5^3 + 24A_1^2A_3A_5$$
$$+ 12A_1^3A_7 + 12A_1^2A_3A_7 + 24A_1A_3^2A_7 + 24A_1A_3A_5A_7$$
$$+ 24A_1A_5^2A_7 + 12A_1A_5A_7^2 + 12A_1A_7^3 + 12A_3^2A_5A_7$$
$$+ 6A_3^2A_7^2 + 24A_3A_5A_7^2 + 4A_5^3A_7 = 0.$$

In Table I the computer solutions of these equations are recorded.

IV. CONCLUSIONS

The diagonalization of the Hamiltonian $p^2 + x^4$ has been carried out several steps. The numerical work indicates that the approximations are convergent. There are several directions for the further development of these ideas.

(1) Inclusion of a harmonic term: If the potential were $x^2 + \lambda x^4$ could the same procedure be applied? The lowest order equation can no longer be solved conveniently for a_n and b_n as we did in Sec. II. The equation becomes a nontrivial cubic. Suppose we call the solutions a_n and b_n. We may try and carry through the identical work with the $p_{rs} = B_{r-s}b_{r+s}$ and $x_{rs} = A_{r-s}a_{r+s}$. It seems possible to carry out the program without the specific simple forms $(r+s)^{1/3}$ and $(r+s)^{2/3}$.

(2) Terms of lower order than $(r+s)^{1/3}$ and $(r+s)^{2/3}$. To improve the approximations it should be possible to construct equations for terms proportional to $(r+s)^{-2/3}$ and $(r+s)^{-1/3}$. Although tedious it seems straightforward to include these contributions.

(3) Generalizations to systems of oscillators: The technique of using the diagonal commutator equation and first off-diagonal Hamiltonian equation to establish the form of the leading contributors has been carried out for the case of two oscillators with no great difficulty. The exact route to follow in adding more refined terms is not so clear in this case because there are six commutators, two coordinates, and two momenta. How to increase the number of equations and the number of variables at equal rates is not clear.

It is worth considering why the present method works. For example, we might have at the nth step solved $2n$ commutator equations and no Hamiltonian equations and expected the Hamiltonian equations to be automatically satisfied. This expectation seemed reasonable to us initially based on the argument that there is only one problem, namely $p^2 + x^4$ with $x^2 \sim p$. We were surprised when this approach did not work until we realized there is a whole class of problems such as $p^4 + x^8$ in which the relation between x and p is that given above. None of these is physical but they exist as mathematical examples. The procedure of taking one commutator and one Hamiltonian equation at a time is apparently successful in producing the correct x and the correct p.

APPENDIX: EXPRESSION FOR THE HAMILTONIAN AND COMMUTATOR

We assume that $X_{ab} = A_{a-b}\,(a+b)^{1/3}$ and that $p_{ab} = iB_{a-b}\,(a+b)^{2/3}$ based on our experience in the lowest approximation. We seek expressions for C_{ab} and H_{ab} that are correct to the leading terms in $(a+b)$. The coefficients A and B are chosen so that $A_{a-b} = A_{b-a}$ while $B_{a-b} = -B_{b-a}$. The commutator is given by

$$C_{ab} = \sum (p_{as}x_{sb} - x_{as}p_{sb})$$
$$= i\sum [B_{a-s}A_{s-b}(a+s)^{2/3}(s+b)^{1/3} - A_{a-s}B_{s-b}(a+s)^{1/3}(s+b)^{2/3}].$$

In the second sum change variables to $s' = a+b-s$ so that C_{ab} becomes

$$C_{ab} = i\sum B_{a-s}A_{s-b}[(a+s)^{2/3}(s+b)^{1/3} - (2a+b-s)^{1/3}(a+2b-s)^{2/3}].$$

The limits of the sum are not the same for the first and second terms but we shall only consider the common range of summation $0 \le s \le a+b$. Outside this range it will turn out that A and B are very small. We next expand the radicals about $(a+b)$ and retain terms to the second order. This gives

$$C_{ab} = i\sum B_{a-s}A_{s-b}\,(a+b)\left[\left(1+\frac{s-b}{a+b}\right)^{2/3}\left(1+\frac{s-a}{a+b}\right)^{1/3} - \left(1+\frac{a-s}{a+b}\right)^{1/3}\left(1+\frac{b-s}{a+b}\right)^{2/3}\right]$$
$$= i\sum (1/3)B_{a-s}A_{s-b}(6s-4b-2a) + O[(a+b)^{-2}] + \cdots.$$

The terms of first and third order vanish so this is our final expression for the commutator.

The diagonal commutator element is given by

$$C_{aa} = -2i\sum (a-s)B_{a-s}A_{a-s},$$

and the diagonal commutator equations in various orders can be found by including the appropriate values of s in the sum because by the parity selection rule s must have opposite odd—even parity from a.

The off-diagonal commutator equations are simply

$$\sum (6s-4b-2a)B_{a-s}A_{s-b} = 0$$

and s must again have odd—even parity opposite from that of a and b which have the same parity. The lowest order equation is found by only taking s as close to $\frac{1}{2}(a+b)$ as possible. Higher orders are found by taking s successively more remote from this central value.

Next we consider the expressions for p^2 and x^4 the two terms in the Hamiltonian. We treat p^2 first since it is simpler than x^4

$$\begin{aligned} p^2_{ab} &= i^2 \sum B_{a-s} B_{s-b} (a+s)^{2/3} (s+b)^{2/3} \\ &= -\sum B_{a-s} B_{s-b} (a+b)^{4/3} \left(1 + \frac{s-b}{a+b}\right)^{2/3} \left(1 + \frac{s-a}{a+b}\right)^{2/3} \\ &= -(a+b)^{4/3} \sum B_{a-s} B_{s-b} \left(1 + \frac{2}{3}\frac{2s-a-b}{a+b} + \cdots\right) \\ &= -(a+b)^{4/3} \sum B_{a-s} B_{s-b} \\ &\quad -\frac{2}{3}(a+b)^{1/3} \sum (2s-a-b) B_{a-s} B_{s-b} + \cdots . \end{aligned}$$

The second term vanishes since the terms with s and with $a+b-s$ contribute equal and opposite amounts so that

$$p^2_{ab} = -(a+b)^{4/3} \sum B_{a-s} B_{s-b} + O[(a+b)^{-2/3}].$$

We retain only the $(a+b)^{4/3}$ term. The vanishing of the second term is a general feature of our work.

Next we consider the x^4 term:

$$\begin{aligned} x^4_{ab} &= \sum A_{a-q} A_{q-r} A_{r-s} A_{s-b} (a+q)^{1/3} (q+r)^{1/3} (r+s)^{1/3} (s+b)^{1/3} \\ &= (a+b)^{4/3} \sum A_{a-q} A_{q-r} A_{r-s} A_{s-b} \left(1 + \frac{q-b}{a+b}\right)^{1/3} \\ &\quad \times \left(1 + \frac{q+r-a-b}{a-b}\right)^{1/3} \left(1 + \frac{r+s-a-b}{a+b}\right)^{1/3} \\ &\quad \times \left(1 + \frac{s-a}{a+b}\right)^{1/3} \\ &= (a+b)^{4/3} \sum A_{a-q} A_{q-r} A_{r-s} A_{s-b} + \tfrac{1}{3}(a+b)^{1/3} \\ &\quad \times \sum (2q+2r+2s-3a-3b) A_{a-q} A_{q-r} A_{r-s} A_{s-b} + \cdots . \end{aligned}$$

TABLE I. Solutions of the equations for the A's and B's in the various orders determined numerically.

	1st order	2nd order	3rd order	4th order
A_1	0.5	0.461087	0.460787	0.460786
A_3	---	0.0232112	0.0207894	0.0207712
A_5	---	---	0.001035	0.000898
A_7	---	---	---	0.0000462
B_1	0.5	0.531794	0.532751	0.532758
B_3	---	0.0688800	0.0719824	0.0720464
B_5	---	---	0.005005	0.0051872
B_7	---	---	---	0.0003028

Again the second term vanishes. To see this let $a-q=\theta_1$, $q-r=\theta_2$, $r-s=\theta_3$, and $s-b=\theta_4$. There are choices of q, r, and s such that all 24 permutations of θ_1, θ_2, θ_3, and θ_4 occur. If we write the second sum in terms of θ's it becomes

$$\sum A_{\theta_1} A_{\theta_2} A_{\theta_3} A_{\theta_4} (-3\theta_1 - \theta_2 + \theta_3 + 3\theta_4).$$

If this is summed over the 4! permutations of the θ's it vanishes so that

$$x^4_{ab} = (a+b)^{4/3} \sum A_{a-q} A_{q-r} A_{r-s} A_{s-b} + [O(a+b)^{-2/3}].$$

*Work supported in part by the United States Atomic Energy Commission.

[1]See, for example, Vladimir Rojansky, *Introductory Quantum Mechanics* (Prentice Hall, New York, 1938), Chap. V, for references to early work on this problem.
[2]S.N. Biswas, K. Datta, R.P. Saxena, P.K. Srivastava, and V.S. Varma, Phys. Rev. D **4**, 3617 (1971); C.A. Uzes and J.H. Henkel, Phys. Rev. D **8**, 1067 (1973); S. Graffi and V. Grecchi, preprint, June (1973).
[3]Carl M. Bender and Tai Tsun Wu, Phys. Rev. D **7**, 1620 (1973); S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. **32 B**, 631 (1970); Carl M. Bender and T.T. Wu, Phys. Rev. **184**, 1231 (1969); J.J. Loeffel, A. Martin, B. Simon, and A.S. Wightman, Phys. Lett. **30 B**, 656 (1969).
[4]A.Z. Capri, Lett. Nuovo Cimento **3**, 351 (1972); Francis R. Halpern, J. Math. Phys. **14**, 219 (1973); P. Lu, S.S. Wald, and B.L. Young, Phys. Rev. D **7**, 1701 (1973).

A new method for the evaluation of slowly convergent series

André Hautot*

University of Liège, Institute of Physics, Sart Tilman par 4000 Liège 1, Belgium
(Received 26 February 1974)

A new method is presented which sums certain slowly convergent series. It is based on the use of the Hankel integral transform and Schlömilch series. This method is applied with great success to the computation of lattice sums in ionic crystals. In particular, the Madelung constant is calculated with great accuracy through rather simple calculations: The final results only involve elementary functions so that the numerical evaluation is quite easy.

I. INTRODUCTION

It is commonly admitted that the interaction potential in an ionic crystal follows the law $qq'/r - A/r^s$. q and q' are the charges of the ions and r is the distance between them. The total interaction for an ion is therefore described by the following lattice sums:

$$\alpha = \sum\sum\sum{}'(\pm)r^{-1} \quad (=\text{Madelung constant of the crystal}), \tag{1}$$

$$M_s = \sum\sum\sum{}' r^{-s} \qquad (s>3). \tag{2}$$

The symbol ($\pm$) indicates that the signs of the ions are taken into account (and also the modulus of the charges if they are not equal for the various ions). The prime means that the summation is extended to all the ions in the crystal except that for which $r=0$.

Many solutions have been proposed for evaluating these sums. The natural method of counting (increasing r) is not interesting since the convergence is bad. Evjen[1] has modified the way of counting to improve the convergence. In spite of its success it must be recognized that the convergence remains poor. The method is almost interesting when one deals with very complicated multiple sums, for which no analytic method can be used. Madelung[2] calculated α by means of Fourier series. The convergence of the method is quite good. However it is not very elegant and Evjen[1] pointed out that the treatment lacked rigor in some places. The most powerful method with regard to the available accuracy is due to Ewald[3]. Unfortunately the method is far from simple. Born and Huang[4,5] have based another method on the properties of Jacobi's theta functions but the method loses its initial elegance when applied to numerical computations. Very recently[6] Glasser showed how it was possible to sum (1) and (2) when the lattice is even-dimensional but pointed that no extension seems to exist to the important three-dimensional case. Now one could ask: why a new method? Our answer lies in the two following points:

In spite of the existence of numerous summation methods there is some need for a simple method leading to very accurate values through accessible intermediate calculations.

Such a simple method exists and provides an interesting application of the so-called "Schlömilch series" in mathematical physics.

II. MATHEMATICAL PRELIMINARIES

A. A useful Laplace transform

Let us first recall a formula which shall play an important role:

$$(a^2+b^2)^{-s} = [2^{1-2s}\pi^{1/2}/\Gamma(s)]\int_0^\infty x^{2s-1}\exp(-ax) \times [J_{s-1/2}(bx)/(bx/2)^{s-1/2}]\,dx. \tag{3}$$

If we have to sum on both a and b, it might be very tempting to sum first with respect to a since the integrand is simply the general term of a geometric series. However there is a better method: it is possible to sum with respect to b. One obtains a Schlömilch series with very useful properties.

B. Some theorems about Schlömich series

These series were first investigated by Schlömilch[7] in the last century. Now this subject is classic and it is developed in advanced books dealing with the theory of Bessel functions.[8] We present some classical results about Schlömilch series which are interesting for our purpose. Schlömilch has investigated the problem of expanding an arbitrary function into a Schlömilch series:

$$f(x) = [a_0/2\Gamma(s+1)] + \sum_{m=1}^{\infty}[a_m J_s(mx) + b_m H_s(mx)]/(mx/2)^s$$

where J_s and H_s are Bessel and Struve functions, respectively.[8] Nielsen[9] has found the following results (all the functions below are even):

$$\begin{aligned} f_s(x) &= [1/2\Gamma(s+1)] + \sum_{m=1}^{\infty}(-1)^m J_s(mx)/(mx/2)^s \\ &= (1/2)\sum_{-\infty}^{+\infty}(-1)^m J_s(mx)/(mx/2)^s = 0 \quad \text{if } 0<x<\pi \\ &= [2\pi^{1/2}/x\Gamma(s+1/2)]\sum_{n=1}^{q}[1-(2n-1)^2\pi^2/x^2]^{s-1/2} \\ &\qquad \text{if } (2q-1)\pi < x < (2q+1)\pi. \end{aligned} \tag{4}$$

It is also possible to establish that:

$$\begin{aligned} g_s(x) &= [1/2\Gamma(s+1)] + \sum_{m=1}^{\infty} J_s(mx)/(mx/2)^s \\ &= (1/2)\sum_{-\infty}^{+\infty} J_s(mx)/(mx/2)^s \\ &= [\pi^{1/2}/x\Gamma(s+1/2)] \quad \text{if } 0<x<2\pi \\ &= [\pi^{1/2}/x\Gamma(s+1/2)] + [2\pi^{1/2}/x\Gamma(s+1/2)] \\ &\quad \times \sum_{n=1}^{q}[1-(2n\pi/x)^2]^{s-1/2} \quad \text{if } 2q\pi < x < 2(q+1)\pi. \end{aligned} \tag{5}$$

 Copyright © 1974 American Institute of Physics

From these two fundamental formulas we deduce other simple expressions:

$$[1/\Gamma(s+1)]+2\sum_{m=1}^{\infty} J_s(2mx)/(mx)^s=\sum_{-\infty}^{+\infty} J_s(2mx)/(mx)^s = f_s(x)+g_s(x), \quad (6)$$

$$\begin{aligned}\sum_{m=1}^{\infty} J_s[(2m-1)x]/[(2m-1)x/2]^s &= \sum_{-\infty}^{+\infty} J_s[(4p+1)x]/[(4p+1)x/2]^s\\ &= \sum_{-\infty}^{+\infty} J_s[(4p+3)x]/[(4p+3)x/2]^s\\ &= \frac{1}{2}\sum_{-\infty}^{+\infty} J_s[(2p+1)x]/[(2p+1)x/2]^s = \tfrac{1}{2}[g_s(x)-f_s(x)]. \quad (7)\end{aligned}$$

C. Hobson integral and its consequences

The modified Bessel function of the third kind K_s admits the following integral representation due to Hobson:

$$\int_a^{\infty} \exp(-bx)(x^2-a^2)^{s-1/2}\,dx=(2a/b)^s\pi^{-1/2}\Gamma(s+1/2)K_s(ab). \quad (8)$$

This formula enables us to calculate the following expressions:

$$P_s(b)=\int_0^{\infty}\exp(-bx)x^{2s}f_s(x)\,dx \quad (s\geq 0) \quad (9)$$

and

$$Q_s(b)=\int_0^{\infty}\exp(-bx)x^{2s}g_s(x)\,dx \quad (s>0). \quad (10)$$

One finds without difficulty through (4), (5), and (8) that

$$P_s(b)=2(2\pi/b)^s[K_s(\pi b)+3^sK_s(3\pi b)+5^sK_s(5\pi b)+\cdots], \quad (11)$$

$$Q_s(b)=2^{2s-1}b^{-2s}\Gamma(s)+2(2\pi/b)^s[2^sK_s(2\pi b)+4^sK_s(4\pi b)+\cdots]. \quad (12)$$

These expansions are very rapidly convergent. For example if in (11) we set $s=0$ and $b=1$, the first term in the brackets is $K_0(\pi)\sim 3.10^{-2}$ while the third term is $K_0(5\pi)\sim 5.10^{-8}$; the third term brings a relative correction less than 2.10^{-6}. The quick convergence is the consequence of the asymptotic behaviour of $K_s(z) \sim(\pi/2z)^{1/2}\exp(-z)$.

III. EVALUATION OF LATTICE SUMS

We shall apply the new method to the evaluation of α and M_s in the three fundamental cubic lattices: the NaCl structure, the CsCl structure and the ZnS structure. The method extends without difficulties to the noncubic systems.

A. The NaCl structure

The coordinates of the ions are integers m, n and p. The charge of each ion is $(-1)^{m+n+p+1}$.

1. The Madelung constant α (NaCl)

$$\begin{aligned}\alpha(\text{NaCl}) &= \sum\sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+p+1}(m^2+n^2+p^2)^{-1/2}\\ &= \sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+1}\int_0^{\infty}\exp[-x(m^2+n^2)^{1/2}]\\ &\quad\times\left[\sum_{-\infty}^{+\infty}(-1)^pJ_0(px)\right]dx+\sum_{-\infty}^{+\infty}{}'(-1)^{p+1}\int_0^{\infty}J_0(px)\,dx,\end{aligned}$$

where use has been made of (3). The Schlömilch series in the first term equals $2f_0(x)$. Therefore one has, with the aid of (9),

$$\begin{aligned}\alpha(\text{NaCl}) &= 2\ln 2+4\sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+1}\{K_0[\pi(m^2+n^2)^{1/2}]\\ &\quad+K_0[3\pi(m^2+n^2)^{1/2}]+\cdots\}\\ &= 2\ln 2+16[K_0(\pi)-K_0(\pi\sqrt{2})-K_0(2\pi)\\ &\quad+2K_0(\pi\sqrt{5})-K_0(\pi\sqrt{8})+2K_0(3\pi)-2K_0(\pi\sqrt{10})\\ &\quad+2K_0(\pi\sqrt{13})-K_0(4\pi)+\cdots].\end{aligned}$$

If four terms in the brackets are retained, one finds $\alpha=1.7479$. The relative error δ equals 2.10^{-4}. Nine terms give $1.747\,56$ ($\delta<2.10^{-6}$). This simple example shows how neat the method is. The same result might be obtained by using Poisson's simple summation formula but it almost appears as an accident.[13]

2. Calculation of M_{2s} (NaCl)

$$\begin{aligned}M_{2s}(\text{NaCl}) &= \sum\sum_{-\infty}^{+\infty}\sum{}'(m^2+n^2+p^2)^{-s}\\ &= [2^{1-2s}\pi^{1/2}/\Gamma(s)]\sum_{-\infty}^{+\infty}\sum{}'\int_0^{\infty}\exp[-x(m^2+n^2)^{1/2}]x^{2s-1}\\ &\quad\times\sum_{-\infty}^{+\infty}J_{s-1/2}(px)/(px/2)^{s-1/2}\,dx\\ &\quad+2\sum_1^{\infty}[\pi^{1/2}/\Gamma(s)]\int_0^{\infty}(x/2p)^{s-1/2}J_{s-1/2}(px)\,dx\\ &= 2\sum_1^{\infty}p^{-2s}+[2^{2-2s}\pi^{1/2}/\Gamma(s)]\\ &\quad\times\sum_{-\infty}^{+\infty}\sum{}'Q_{s-1/2}[(m^2+n^2)^{1/2}].\end{aligned}$$

The first term reduces to the Riemann zeta function; the second term splits into two parts in agreement with (12); the first part is written as

$$[2^{2-2s}\pi\Gamma(2s-1)/[\Gamma(s)]^2]\sum_{-\infty}^{+\infty}\sum{}'(m^2+n^2)^{1/2-s}.$$

The double series has been calculated by Glasser[6] who found that

$$\sum_{-\infty}^{+\infty}\sum{}'(m^2+n^2)^{-s}=4\zeta(s)\beta(s).$$

The final result is now immediate:

$$\begin{aligned}M_{2s}(\text{NaCl}) &= 2\zeta(2s)+[2^{4-2s}\pi\Gamma(2s-1)/[\Gamma(s)]^2]\\ &\quad\zeta(s-1/2)\,\beta(s-1/2)\\ &\quad+[2^{5/2-s}\pi^s/\Gamma(s)]\sum_{-\infty}^{+\infty}\sum{}'(m^2+n^2)^{(1-2s)/4}\\ &\quad\times\{2^{s-1/2}K_{s-1/2}[2\pi(m^2+n^2)^{1/2}]\\ &\quad+4^{s-1/2}K_{s-1/2}[4\pi(m^2+n^2)^{1/2}]+\cdots\}.\end{aligned}$$

Numerical examples:

$$\begin{aligned}M_{10} &= 2\zeta(10)+(35\pi/32)\zeta(9/2)\beta(9/2)\\ &\quad+(\pi^5/96\sqrt{2})\{4.\,2^{9/2}K_{9/2}(2\pi)+4.\,4^{9/2}K_{9/2}(4\pi)\\ &\quad+4.\,2^{9/4}K_{9/2}(2\pi\sqrt{2})+4.\,6^{9/2}K_{9/2}(6\pi)\end{aligned}$$

$$+4.4^{9/2}2^{-9/4}K_{9/2}(4\pi\sqrt{2})+8.4^{9/2}5^{-9/4}K_{9/2}(2\pi\sqrt{5})+\cdots\}$$

The series in the brackets converges quickly: three terms in the series give M_{10} with three significant figures; six terms give M_{10} with seven figures. One finds

$$M_{10}(\mathrm{NaCl})=6.426\,104.$$

B. The CsCl structure

The coordinates of the ions are $(m+1/2,\ n+1/2,\ p+1/2)$ = positive ions and (m,n,p) = negative ions.

1. The Madelung constant α (CsCl)

$$\alpha(\mathrm{CsCl})=\sum\sum_{-\infty}^{+\infty}\sum{}'\{[(m+1/2)^2+(n+1/2)^2+(p+1/2)^2]^{-1/2}-(m^2+n^2+p^2)^{-1/2}\}$$
$$=2\sum\sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+p+1}(m^2+n^2+p^2)^{-1/2}+6\sum\sum_{-\infty}^{+\infty}\sum\{[4m^2+(2n+1)^2+(2p+1)^2]^{-1/2}-[4m^2+(2n+1)^2+4p^2]^{-1/2}\}.$$

Under that form the expression is well prepared for the introduction of a Schlömilch series; using (3), (6), and (7) one finds

$$\alpha(\mathrm{CsCl})=2\alpha(\mathrm{NaCl})+6\sum_{-\infty}^{+\infty}\sum\int_0^\infty\exp\{-x[4m^2+(2n+1)^2]^{1/2}\}\sum_{-\infty}^{+\infty}\{J_0[(2p+1)x]-J_0(2px)\}\,dx,$$
$$\alpha(\mathrm{CsCl})=2\alpha(\mathrm{NaCl})-12\sum_{-\infty}^{+\infty}\sum P_0\{[4m^2+(2n+1)^2]^{1/2}\}$$
$$=2\alpha(\mathrm{NaCl})-24\sum_{-\infty}^{+\infty}\sum\{K_0(\pi[4m^2+(2n+1)^2]^{1/2})+K_0(3\pi[4m^2+(2n+1)^2]^{1/2})+\cdots\}$$
$$=2\alpha(\mathrm{NaCl})-48[K_0(\pi)+2K_0(\pi\sqrt{5})+2K_0(3\pi)+2K_0(\pi\sqrt{13})+2K_0(\pi\sqrt{17})+4K_0(5\pi)+2K_0(\pi\sqrt{29})+\cdots]$$
$$=2.035\,35.$$

2. Calculation of M_{2s}(CsCl)

$$M_{2s}(\mathrm{CsCl})=\sum\sum_{-\infty}^{+\infty}\sum{}'\{[(m+1/2)^2+(n+1/2)^2+(p+1/2)^2]^{-s}+(m^2+n^2+p^2)^{-s}\}=M_{2s}(\mathrm{NaCl})+2^{2s}\sum\sum_{-\infty}^{+\infty}\sum[(2m+1)^2+(2n+1)^2+(2p+1)^2]^{-s}.$$

The triple series is easily calculated by using the method which is now familiar to the reader; one finds

$$[2^{1-2s}\pi^{1/2}/\Gamma(s)]\sum\sum_{-\infty}^{+\infty}(Q_{s-1/2}\{[(2m+1)^2+(2n+1)^2]^{1/2}\}-P_{s-1/2}\{[(2m+1)^2+(2n+1)^2]^{1/2}\}).$$

Using (11) and (12) one finds a first contribution of the type $\sum_{-\infty}^{+\infty}\sum[(2m+1)^2+(2n+1)^2]^{-s}$. Its value is given by Glasser[6]: $2^{2-s}(1-2^{-s})\zeta(s)\beta(s)$. Finally one finds

$$M_{2s}(\mathrm{CsCl})=M_{2s}(\mathrm{NaCl})+2^{s+3/2}\pi^{1/2}[\Gamma(s-1/2)/\Gamma(s)]\times(1-2^{1/2-s})\zeta(s-1/2)\beta(s-1/2)-[2^{s+7/2}\pi^s/\Gamma(s)]\sum_0^\infty\sum[(2m+1)^2+(2n+1)^2]^{(1-2s)/4}(K_{s-1/2}(\pi[(2m+1)^2+(2n+1)^2]^{1/2})-2^{s-1/2}K_{s-1/2}\{2\pi[(2m+1)^2+(2n+1)^2]^{1/2}\}+3^{s-1/2}\ldots).$$

Numerical examples:

$$M_{10}(\mathrm{CsCl})=M_{10}(\mathrm{NaCl})+(105\pi/96)(16\sqrt{2}-1)\zeta(\tfrac{9}{2})\beta(\tfrac{9}{2})-(32\pi^5\sqrt{2}/3)[2^{-9/4}K_{9/2}(\pi\sqrt{2})-2^{9/4}K_{9/2}(2\pi\sqrt{2})+2^{-9/4}3^{9/2}K_{9/2}(3\pi\sqrt{2})-2^{27/4}K_{9/2}(4\pi\sqrt{2})+2.10^{-9/4}K_{9/2}(\pi\sqrt{10})+\cdots]$$
$$=40.3043.$$

C. The ZnS structure

The negative ions lie at the sites $(m/2,\ n/2,\ p/2)$ with $m+n+p$ even. The positive ions lie at the sites $(m/2+1/4,\ n/2+1/4,\ p/2+1/4)$ with the same condition.

1. The Madelung constant α (ZnS)

$$\alpha(\mathrm{ZnS})=\sum\sum_{-\infty}^{+\infty}\sum{}'\{12[(4m+1)^2+(4n+3)^2+(4p+3)^2]^{-1/2}+4[(4m+1)^2+(4n+1)^2+(4p+1)^2]^{-1/2}-(m^2+n^2+p^2)^{-1/2}-6[4m^2+(2n+1)^2+(2p+1)^2]^{-1/2}\}.$$

The two first terms can be transformed together into $\sum\sum_0^\infty\sum 16[(2m+1)^2+(2n+1)^2+(2p+1)^2]^{-1/2}=2\sum\sum_{-\infty}^{+\infty}\sum[(2m+1)^2+(2n+1)^2+(2p+1)^2]^{-1/2}$ through simple arithmetical devices. We find that

$$\alpha(\mathrm{ZnS})=\alpha(\mathrm{CsCl})-6\sum\sum_{-\infty}^{+\infty}\sum[4m^2+(2n+1)^2+(2p+1)^2]^{-1/2}.$$

The triple series will be evaluated in Sec. III.C2 for a general exponent s. Here we take the limit when s tends to 1/2. We find:

$$\alpha(\mathrm{ZnS})=\alpha(\mathrm{CsCl})+3\ln 2-48[K_0(\pi\sqrt{2})+K_0(2\pi\sqrt{2})+2K_0(\pi\sqrt{10})+2K_0(\pi\sqrt{18})+2K_0(\pi\sqrt{26})+\cdots]$$
$$=3.782\,926.$$

This simple formula gives α with seven significant figures!

2. Calculation of M_{2s} (ZnS)

Using the arithmetical devices used in Sec. III C. 1, M_{2s} is easily brought into the form

$$M_{2s}(\mathrm{ZnS})=\sum\sum_{-\infty}^{+\infty}\sum{}'\{2^{4s-1}[(2m+1)^2+(2n+1)^2+(2p+1)^2]^{-s}+(m^2+n^2+p^2)^{-s}+3.2^{2s}[4m^2+(2n+1)^2+(2p+1)^2]^{-s}\}$$

$$= 2^{2s-1}M_{2s}(\mathrm{CsCl}) - (2^{2s-1} - 1)M_{2s}(\mathrm{NaCl}) + 3.\,2^{2s}\sum\sum_{-\infty}^{+\infty}\sum[4m^2 + (2n+1)^2 + (2p+1)^2]^{-s}.$$

The triple series can be evaluated as above. One finds

$$[2^{1-2s}\pi^{1/2}/\Gamma(s)]\sum_{-\infty}^{+\infty}\sum (Q_{s-1/2}\{(2n+1)^2 + (2p+1)^2]^{1/2}\} + P_{s-1/2}\{[(2n+1)^2 + (2p+1)^2]^{1/2}\}).$$

Finally, one has

$$\begin{aligned}M_{2s}(\mathrm{ZnS}) = {} & 2^{2s-1}M_{2s}(\mathrm{CsCl}) - (2^{2s-1} - 1)M_{2s}(\mathrm{NaCl}) \\ & + 3\pi^{1/2}2^{s+3/2}(1 - 2^{1/2-s}) \\ & [\Gamma(s-1/2)/\Gamma(s)]\zeta(s-1/2)\beta(s-1/2) \\ & + 3[2^{s+7/2}\pi^s/\Gamma(s)]\sum_0^\infty\sum[(2n+1)^2 \\ & + (2p+1)^2]^{(1-2s)/4}(K_{s-1/2}\{\pi[(2n+1)^2 \\ & + (2p+1)^2]^{1/2}\} + 2^{s-1/2}K_{s-1/2}\{2\pi[(2n+1)^2 \\ & + (2p+1)^2]^{1/2}\} + \cdots).\end{aligned}$$

Numerical example:

$$\begin{aligned}M_{10}(\mathrm{ZnS}) = {} & 512M_{10}(\mathrm{CsCl}) - 511M_{10}(\mathrm{NaCl}) \\ & + (105\pi/32)(16\sqrt{2} - 1)\zeta(9/2)\beta(9/2) \\ & + 32\pi^5\sqrt{2}[2^{-9/4}K_{9/2}(\pi\sqrt{2}) + 2^{9/4}K_{9/2}(2\pi\sqrt{2}) \\ & + 2^{-9/4}3^{9/2}K_{9/2}(3\pi\sqrt{2}) + 2^{27/4}K_{9/2}(4\pi\sqrt{2}) \\ & + 2.\,10^{-9/4}K_{9/2}(\pi\sqrt{10}) + \cdots] \\ = {} & 17\,740.\end{aligned}$$

D. Refinement of the above results

The evaluation of M_{2s} and α has been performed in a satisfactory way: the calculations are neat and the final results are expressed in the form of very quickly convergent series. However tables of the K_s functions are needed. When $s = n + \frac{1}{2}$ (n integer), the tabulation is easily performed since $K_{n+1/2}$ is an elementary function (product of an exponential by a polynomial). When $s = n$ (integer), the problem is less simple. If a relative accuracy of about 10^{-6} is judged sufficient, one can use Watson's table[8] (with seven figures). In practice, this accuracy is quite sufficient. However it is possible to refine the results by expressing α and M_{2s} in terms of elementary functions only. This statement is obvious in the case of M_{2s} provided $s = n$ is an integer. If $s = n + \frac{1}{2}$ we shall see that this is also true. Now we present the refined method and we apply it to the evaluation of α. If $s \neq n$ or $n + \frac{1}{2}$, the problem is not soluble in terms of elementary functions; since K_s is not tabulated in these cases the evaluation of M_{2s} would require further investigation. Fortunately the two possibilites $s = n$ or $s = n + \frac{1}{2}$ are in practice quite sufficient. So we try to refine the previous result:

$$\alpha(\mathrm{NaCl}) = 2\ln 2 + 4\sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+1}\{K_0[\pi(m^2+n^2)^{1/2}] + K_0[3\pi(m^2+n^2)^{1/2}] + \cdots\}.$$

First, we calculate:

$$S(z) = \sum_{-\infty}^{+\infty}\sum{}'(-1)^{m+n+1}K_0[z(m^2+n^2)^{1/2}]. \tag{13}$$

We show that the use of Schlömilch series allows us to transform (13). We have

$$S(z) = \sum_{-\infty}^{+\infty}\sum{}' = S_1 + S_2 = 2\sum_{m=1}^{\infty}\sum_{n=-\infty}^{+\infty} + 2\sum_{n=1}^{\infty} \quad (m=0).$$

To calculate S_1, we start with the formula

$$K_0[z(m^2+n^2)^{1/2}] = m\int_0^\infty(t^2+z^2)^{-1/2}J_0(nt)K_1[m(t^2+z^2)^{1/2}]t\,dt.$$

which is introduced in the definition of S_1: a Schlömilch series immediately appears which is summed accordingly to (4):

$$S_1 = 2\sum_{m=1}^{\infty}(-1)^{m+1}\,2m\int_0^\infty f_0(t)(t^2+z^2)^{-1/2}K_1[m(t^2+z^2)^{1/2}]t\,dt.$$

Using Eq. (4), we find a development with integrals of the type:

$$\int_0^\infty(u^2+z^2)^{-1/2}K_1[m(u^2+z^2)^{1/2}]\,du = (\pi/2mz)\exp(-mz).$$

We get:

$$\begin{aligned}\sum_{n=-\infty}^{+\infty}(-1)^nK_0[z(m^2+n^2)^{1/2}] = {} & 2\pi\{(z^2+\pi^2)^{-1/2} \\ & \times\exp[-m(z^2+\pi^2)^{1/2}] \\ & + (z^2+9\pi^2)^{-1/2}\exp[-m(z^2+9\pi^2)^{1/2}] + \cdots\}\end{aligned}$$

and finally

$$S_1(z) = 4\pi\sum_{k=0}^{\infty}[z^2+(2k+1)^2\pi^2]^{-1/2}\{\exp[z^2+(2k+1)^2\pi^2]^{1/2} + 1\}^{-1}.$$

$S_2(z)$ is evaluated by means of a similar technique (see Appendix A). The final result expresses $\alpha(\mathrm{NaCl})$ in terms of elementary functions [except for the use of $\zeta(1/2)$ and $\beta(1/2)$ which are tabulated]:

$$\begin{aligned}\alpha(\mathrm{NaCl}) = {} & 4(1-2^{1/2})\zeta(1/2)\beta(1/2) \\ & + 16\sum_{k,l=0}^{\infty}\sum[(2l+1)^2+(2k+1)^2]^{-1/2} \\ & \{\exp[(2l+1)^2+(2k+1)^2]^{1/2}\pi + 1\}^{-1}.\end{aligned} \tag{14}$$

This expansion exhibits remarkable convergence; eight terms give α with twelve figures!:

$$\alpha(\mathrm{NaCl}) = 1.747\,564\,594\,63.$$

Note that one term gives α correct with four figures:

$$\alpha(\mathrm{NaCl}) \approx 4(1-2^{1/2})\zeta(1/2)\beta(1/2) + 16.\,2^{-1/2}[\exp(\pi 2^{1/2}) + 1]^{-1} = 1.747.$$

Of course the same procedure gives the values of $\alpha(\mathrm{CsCl})$ and $\alpha(\mathrm{ZnS})$ (see Appendix B for more details):

$$\begin{aligned}\alpha(\mathrm{CsCl}) = {} & 2\alpha(\mathrm{NaCl}) - 12\sum_{l=1}^{\infty}(2l-1)^{-1}\operatorname{csch}(2l-1)\pi \\ & - 24\sum_{k,l=1}^{\infty}\sum[2l-1)^2+k^2]^{-1/2}\operatorname{csch}\pi[(2l-1)^2+k^2]^{1/2} \\ = {} & 2.035\,361\,509\,45,\end{aligned} \tag{15}$$

$$\alpha(\mathrm{ZnS}) = \alpha(\mathrm{CsCl}) + 3\ln 2 - 6\sum_{l=1}^{\infty} l^{-1}\,\mathrm{csch}(l\pi)$$
$$+ 12\sum_{k,l=1}^{\infty}\sum (-1)^{k+1}(k^2+l^2)^{-1/2}\,\mathrm{csch}[\pi(k^2+l^2)^{1/2}]$$
$$= 3.782\,926\,104\,08.$$

In the special case of the NaCl structure, the refined result might be derived from Poisson's double summation formula.[13]

E. The $\exp(-ar)/r$ potential

The same method applies when more complicated lattice sums must be evaluated. Let us examine the important case where the interaction is of the type $\exp(-ar)/r$. We must calculate ($a>0$):

$$S = \sum\sum_{-\infty}^{+\infty}\sum{}' r^{-1}\exp(-ar).$$

We calculate this sum in the NaCl structure. We start with the formula

$$\int_0^{\infty} t(t^2+a^2)^{-1/2} J_0(xt)\exp[-y(t^2+a^2)^{1/2}]\,dt$$
$$= (x^2+y^2)^{-1/2}\exp[-a(x^2+y^2)^{1/2}].$$

We set $x=p$ and $y=(m^2+n^2)^{1/2}$ (with the notation of Sec. III.A.1). We obtain

$$S = \sum\sum_{-\infty}^{+\infty}\sum{}'\int_0^{\infty} t(t^2+a^2)^{-1/2}J_0(pt)\exp[-(m^2+n^2)^{1/2}$$
$$\times (t^2+a^2)^{1/2}]\,dt.$$

The sum splits into two parts:

$$\sum\sum_{-\infty}^{+\infty}\sum = \sum_{-\infty}^{+\infty}\sum{}'\sum_{p=-\infty}^{+\infty} + 2\sum_{p=1}^{\infty} \quad (m=n=0).$$

In the first term a Schlömilch series appears which is summed in accordance with (5). The second term is easily summed by elementary manipulations on geometric progressions. We find

$$S = -2\ln[1-\exp(-a)] + 2\sum\sum_{-\infty}^{\infty}{}'\{K_0[a(m^2+n^2)^{1/2}]$$
$$+ 2K_0[(a^2+4\pi^2)^{1/2}(m^2+n^2)^{1/2}]$$
$$+ 2K_0[(a^2+16\pi^2)^{1/2}(m^2+n^2)^{1/2}] + \cdots\}.$$

This series quickly converges through the whole range of a values. The use of the K_0 function may be avoided by using the procedure described in Sec. III.D. One finds

$$S = (4\pi/a)[\exp a - 1]^{-1} + 16\pi\sum_1^{\infty}\sum_0^{\infty}[a^2+(2k\pi)^2+(2l\pi)^2]^{-1/2}$$
$$\times\{\exp[a^2+(2k\pi)^2+(2l\pi)^2]^{1/2} - 1\}^{-1}$$
$$+ 4\sum_1^{\infty}\sum (m^2+n^2)^{-1/2}\exp[-a(m^2+n^2)^{1/2}]$$

When a is small, the behavior of the last term has been studied by Glasser[6] who gives its approximate value. The other terms are easily evaluated since they involve only elementary functions.

IV. BRIEF DISCUSSION OF THE NUMERICAL RESULTS

It is interesting to compare the various numerical α values occurring in the literature since they do not always coincide! Let us consider the most important example: α(NaCl). Most of the authors give the value 1.7476 in their textbooks on solid state physics. Kittel[11] and Dekker[12] give more accurate values: 1.747 558. They obtained that value from the classical paper of Shermann.[13] Comparing with our result, we note a discrepancy of 6.10^{-6}. Sakamoto[14] and earlier Emersleben[15] have calculated the same quantity by Ewald's method; they have found a value in agreement with ours. The same remark holds for CsCl: the traditional value[11,12,13] is 2.035 356 but we find 2.035 361. For ZnS the literature is less accurate (3.782 92) so that the discrepancy does not exist.

V. CONCLUSIONS

It is possible to reformulate the above theory by using the language of the theory of integral transforms.[10] Having to sum the series $S=\sum_z(\pm)u(z)$, we introduce the Hankel transform (or order s) of the function $z^s u(z)$:

$$F(t) = \int_0^{\infty} z\,J_s(zt)z^s u(z)\,dz.$$

The inversion theorem tells us that

$$z^s u(z) = \int_0^{\infty} t\,J_s(zt)F(t)\,dt.$$

After slight manipulation we can write

$$S = 2^{-s}\int_0^{\infty} t^{s+1}\left[\sum_z(\pm)J_s(zt)/(zt/2)^s\right]F(t)\,dt.$$

A Schlömilch series appears which is summed according to (4) or (5). Performing the integration, the final result takes the form of a new series whose convergence may be improved with respect to the convergence of $\sum(\pm)u(z)$. This paper has shown by several classical examples that the method is effective and useful. It furnishes a very good method for computing lattice sums in ionic crystals. No other method gives simple results as in Eqs. (14)–(16) with such an accuracy. Among all the existing methods leading to the evaluation of very accurate lattice sums, this method appears to be one of the simplest.

Very recently we have further refined the above results. In particular, the use of Schlömilch series allows us to find numerous summation formulas for K_s functions like those described in Appendix B. Calculations and related applications will be reported in a future paper. A possible application is the expression of α in term of elementary functions only (without reference to the zeta and the beta function of Riemann).

Example: One has the curious formula

$$\alpha(\mathrm{NaCl}) = (9/2)\ln 2 - (\pi/2)$$
$$+ 12\sum_1^{\infty}\sum\{[(2j-1)^2+(2k-1)^2]^{-1/2}$$
$$\times\mathrm{csch}\,\pi[(2j-1)^2+(2k-1)^2]^{1/2}$$
$$-(4j^2+4k^2)^{-1/2}\,\mathrm{csch}\,\pi(4j^2+4k^2)^{1/2}\}.$$

Four terms give:

$$(9/2)\ln 2-(\pi/2)+(12\sqrt{2})\operatorname{csch}\pi\sqrt{2}-(12/\sqrt{8})\operatorname{csch}\pi\sqrt{8}+(24\sqrt{10})\operatorname{csch}\pi\sqrt{10}=1.74756(28)$$

accurate to 10^{-6}.

Similar formulas hold for the other crystallographic structures. They will be reported in a future paper with other possible applications.

APPENDIX A

Certain double series containing K_s functions can be summed exactly in terms of Riemann zeta and beta functions. If $s>0$ one has

$$\sum_{l,m=1}^{\infty}\sum(-1)^{m+1}m^{1/2-s}(2l-1)^{s-1/2}K_{s-1/2}[(2l-1)m\pi]=\pi^{-s}2^{s-5/2}\Gamma(s)[2(1-2^{1-s})\zeta(s)\beta(s)-(1-2^{1-2s})\zeta(2s)].$$

The proof of this formula is left to the reader. He will start with the formula[6]

$$\sum_{m,n=1}^{\infty}\sum(-1)^{m+n}(m^2+n^2)^{-s}=(1-2^{1-2s})\zeta(2s)-(1-2^{1-s})\beta(s)\zeta(s).$$

He will evaluate the double series by the new method. The result will follow. This series occurs in the evaluation of α(NaCl) (with $s=1/2$).

APPENDIX B

Using the method presented in Sec. III.D, the reader will have no difficulty to prove that

$$\sum_{-\infty}^{+\infty}\sum K_0\{z[4m^2+(2n+1)^2]^{1/2}\}=(\pi/2z)\operatorname{csch}z+\pi\sum_{k=1}^{\infty}(z^2+k^2\pi^2)^{-1/2}\operatorname{csch}(z^2+k^2\pi^2)^{1/2}$$

and that

$$\sum_{0}^{\infty}\sum K_0\{z[(2m+1)^2+(2n+1)^2]^{1/2}\}=(\pi/8z)\operatorname{csch}z-(\pi/4)\sum_{k=1}^{\infty}(-1)^{k+1}(z^2+k^2\pi^2)^{-1/2}\times\operatorname{csch}(z^2+k^2\pi^2)^{1/2}.$$

The first equation leads to the refined value of α(CsCl) while the second leads to α(ZnS).

* Presently Professor at the National University of Zaire, Kinshasa.

[1]H. M. Evjen, Phys. Rev. **39**, 680 (1932).
[2]E. Madelung, Phys. Z. **19**, 524 (1918).
[3]P. E. Ewald, Ann. Phys. **64**, 253 (1921).
[4]M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1956).
[5]R. Dh. Misra, Proc. Camb. Phil. Soc. **36**, 173 (1940).
[6]M. L. Glasser, J. Math. Phys. **14**, 409 (1973).
[7]Schlömilch, Z. für Math. und Phys. II, 155 (1857).
[8]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1966).
[9]C. Nielsen, Ann. di Mat. (3) VI, 301 (1901).
[10]I. N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).
[11]C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1956).
[12]A. J. Dekker, *Solid State Physics* (Prentice-Hall, Englewood Cliffs, N. J., 1959).
[13]J. Shermann, Chem. Rev. **11**, 93 (1932).
[14]Y. Sakamoto, J. Sci. U. Hiroshima **A16**, 569 (1953).
[15]O. Emersleben, Z. Phys. Chem. **199**, 170 (1952).

Lie theory and separation of variables. 5. The equations $iU_t + U_{xx} = 0$ and $iU_t + U_{xx} - c/x^2 U = 0$

E. G. Kalnins and W. Miller Jr.

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, P.Q., Canada
(Received 18 March 1974)

A detailed study of the group of symmetries of the time-dependent free particle Schrödinger equation in one space dimension is presented. An orbit analysis of all first order symmetries is seen to correspond in a well-defined manner to the separation of variables of this equation. The study gives a unified treatment of the harmonic oscillator (both attractive and repulsive), Stark effect, and free particle Hamiltonians in the time dependent formalism. The case of a potential c/x^2 is also discussed in the time dependent formalism. Use of representation theory for the symmetry groups permits simple derivation of expansions relating various solutions of the Schrödinger equation, several of which are new.

INTRODUCTION

The present paper is one of a series investigating the connection between separation of variables and Lie symmetry groups. In this work we make a detailed study of the free particle Schrödinger equation in the time-dependent formalism, i.e., the equation

$$(*)\quad u_{xx} + iu_t = 0,$$

and of the radial equation for a free particle,

$$(**)\quad u_{xx} - \frac{c}{x^2}u + iu_t = 0.$$

Anderson *et al.*[1] (with some errors) and Boyer[2] have classified all equations of the form

$$(***)\quad u_{xx} - V(x)u + iu_t = 0$$

which admit a nontrivial symmetry algebra of first order differential operators. It is known, e.g., Neiderer,[3] that among these equations, those corresponding to the harmonic oscillator and the linear potential are actually equivalent to (*). Here we show in a very explicit manner that every equation (***) admitting symmetries is equivalent to either (*) or (**). The equations (***) are exactly those obtained from (*) and (**) by taking all possible separations of variables.

In Sec. 1 we rederive the known six-parameter symmetry group G of equation (*).[1,2,4,5] Here G is a semidirect product of the three-parameter Weyl group W and $SL(2,R)$. We determine the global action of G and compute the orbit structure of its Lie algebra under the adjoint representation.

In Sec. 2 we classify all coordinate systems such that variables separate in equation (*) and relate them one-to-one with the G orbits. It is found necessary to include R separation as well as ordinary separation in this analysis. The orbits are essentially labelled by the attractive and repulsive harmonic oscillator, linear potential, and free particle Hamiltonians. Although all our coordinates systems are already known,[4] the proof that they are exhaustive and their explicit relation to orbits appears to be new.

In Secs. 3 and 4 we give the basis in a one-parameter model for a representative of each G orbit. The calculation of the basis functions in the Hilbert space of functions depending on x and t, and the overlap functions between the various bases are also given. We show that our knowledge of the G structure of (*) greatly simplifies the derivation of the spectral representations of various associated Hamiltonians as well as expansion theorems relating different solutions of (*). Several of the overlap functions are new and our proofs of the L_2-expansion theorems for parabolic cylinder and Airy functions are much simpler than the standard derivations. This work can be considered as the Hilbert space analogy of Weisner's work[6] on analytic expansions in Hermite functions. The papers of Whittaker[7] and Erdélyi[8] are also related to our procedure.

Finally, in Sec. 5 we give a corresponding analysis of the equation (**). The methods of Barut[9] for computing the spectra of Hamiltonians through the use of representation theory are closely related to our approach.

The analysis presented in this paper is preliminary to the treatment of the time-dependent Schrödinger equations in two and three space variables, which admit symmetries. There the theory is much richer. In particular, degenerate eigenvalues appear and it is necessary to associate separable coordinates with both first and second order symmetry operators. Nevertheless, as we shall show in forthcoming papers, the same general approach can be utilized.

All special functions appearing in this work are normalized as in the Bateman project.[10]

1. SYMMETRIES OF THE EQUATION $iu_t + u_{xx} = 0$

Let X be the differential operator

$$X = i\partial_t + \partial_{xx} \tag{1.1}$$

acting on the space $\mathcal{F}$ of locally C^∞ functions of the real variables x, t. We wish to find the maximal symmetry algebra of the equation

$$iu_t = -u_{xx}, \tag{1.2}$$

i.e., we wish to compute all linear differential operators

$$L = a(x,t)\partial_x + b(x,t)\partial_t + c(x,t), \quad a, b, c \in \mathcal{F} \tag{1.3}$$

such that $Lu(x,t)$ satisfies (1.2) whenever u does. As is well known[1,2,11] a necessary and sufficient condition for L to be a symmetry is

$$[L, X] = r(x,t)X \tag{1.4}$$

for some $r \in \mathcal{F}$. By equating coefficients of $\partial_{xx}, \partial_t, \partial_x$, and 1 on both sides of (1.4), one obtains a system of

 Copyright © 1974 American Institute of Physics

differential equations for $a, b, c,$ and r. We omit the details which can be found in several references.[1,2,4] The final result is that the allowable L form a six-dimensional complex Lie algebra $\mathcal{G}^c$ with basis

$$K_2 = -t^2\partial_t - tx\partial_x - t/2 + ix^2/4, \quad K_1 = -t\partial_x + ix/2,$$
$$K_0 = i, \quad K_{-1} = \partial_x, \quad K_{-2} = \partial_t, \quad K^3 = x\partial_x + 2t\partial_t + \tfrac{1}{2} \tag{1.5}$$

and commutation relations

$$[K^3, K_j] = jK_j, \quad j = \pm 2, \pm 1, 0, \quad [K_{-1}, K_1] = \tfrac{1}{2}K_0,$$
$$[K_{-1}, K_2] = K_1, \quad [K_{-2}, K_1] = -K_{-1}, \quad [K_{-2}, K_2] = -K^3. \tag{1.6}$$

In this paper we will be concerned only with the *real* Lie algebra $\mathcal{G}$ whose basis is (1.5). A second convenient basis for $\mathcal{G}$ is S_j, L_k, E, where

$$S_1 = K_{-1}, \quad S_2 = K_1, \quad L_3 = K_{-2} - K_2,$$
$$L_1 = K^3, \quad L_2 = K_{-2} + K_2, \quad E = K_0. \tag{1.7}$$

The commutation relations become

$$[L_1, L_2] = -2L_3, \quad [L_3, L_1] = 2L_2, \quad [L_2, L_3] = 2L_1,$$
$$[S_1, S_2] = \tfrac{1}{2}E, \quad [L_3, S_1] = S_2, \quad [L_3, S_2] = -S_1,$$
$$[L_2, S_1] = [S_2, L_1] = -S_2, \quad [L_1, S_1] = [L_2, S_2] = -S_1$$

where E generates the center of $\mathcal{G}$. Clearly, the operators L_1, L_2, L_3 form a basis for a subalgebra of $\mathcal{G}$ isomorphic to $sl(2,R)$ and the operators S_1, S_2, E form a basis for the Weyl algebra $\mathcal{W}$. Furthermore, $\mathcal{G}$ is the semidirect product of $sl(2,R)$ and $\mathcal{W}$.

Using standard results from Lie theory,[12] one can exponentiate the differential operators of $\mathcal{G}$ to obtain a local Lie group G of operators acting on $\mathcal{F}$. The action of the Weyl group W is given by operators

$$T(u, v, \rho) = \exp[\rho + (uv/4)]E \exp(uS_2)\exp(vS_1) \tag{1.8}$$

with multiplication

$$T(u, v, \rho)T(u', v', \rho') = T(u + u', v + v', \rho + \rho' + (vu' - uv')/4) \tag{1.9}$$

where

$$[T(u, v, \rho)f](x, t) = \exp\{i[\rho + (uv + 2ux - u^2t)/4]\} \times f(x + v - ut, t), \quad f \in \mathcal{F}.$$

The action of $SL(2,R)$ is given by operators

$$[T(A)f](x, t) = \exp\left[i\left(\frac{x^2\beta/4}{\delta + t\beta}\right)\right](\delta + t\beta)^{-1/2} \times f\left[\frac{x}{\delta + t\beta}, \frac{\gamma + t\alpha}{\delta + t\beta}\right] \tag{1.10}$$

where

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL(2,R),$$

i.e., A is a real matrix with determinant $+1$. Furthermore,

$$T\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \exp(\beta K_2), \quad T\begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} = \exp(\beta K_{-2}),$$
$$T\begin{pmatrix} e^{\alpha} & 0 \\ 0 & e^{-\alpha} \end{pmatrix} = \exp(\alpha K^3), \quad T\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \exp(\theta L_3) \tag{1.11}$$
$$T\begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} = \exp(\phi L_2).$$

Finally, the action of $SL(2,R)$ on W via the adjoint representation is

$$T^{-1}(A)T(u, v, \rho)T(A) = T(u\delta + v\beta, u\gamma + v\alpha, \rho). \tag{1.12}$$

This defines G as a semidirect product of $SL(2,R)$ and W:

$$g = (A, w) \in G, \quad A \in SL(2,R), \quad w = (u, v, \rho) \in W,$$
$$T(g) = T(A)T(w), \tag{1.13}$$
$$T(g)T(g') = T(AA')[T(A')^{-1}T(w)T(A')]T(w') = T(gg').$$

It follows from general Lie theory that $T(g)$ maps solutions of (1.2) into solutions.[11]

The group G acts on the Lie algebra $\mathcal{G}$ of differential operators K via the adjoint representation:

$$K \to K^g = T(g)KT^{-1}(g).$$

This action splits $\mathcal{G}$ into G orbits. For our purposes the operator $K_0 = i$ is trivial so we will merely study the orbit structure of the factor algebra $\mathcal{G}' = \mathcal{G}/\{K_0\}$ where $\{K_0\}$ is the center of $\mathcal{G}$.

This computation was carried out by Weisner[6] for the complexification of $\mathcal{G}$ and needs only minor modification to adopt it to $\mathcal{G}$. Let

$$K = A_2K_2 + A_1K_1 + A_{-1}K_{-1} + A_{-2}K_{-2} + A_3K^3$$

be a nonzero element of $\mathcal{G}'$ and set $\alpha = A_2A_{-2} + A_3^2$. It is straightforward to show that α is invariant under the adjoint representation. In the table below we give a complete set of orbit representatives. That is, K lies on the same G orbit as a real multiple of exactly one of the five operators in the list.

$$\text{Case 1}(\alpha < 0):\ K_{-2} - K_2 = L_3,$$
$$\text{Case 2}(\alpha > 0):\ K_3, \tag{1.14}$$
$$\text{Case 3}(\alpha = 0):\ K_2 + K_{-1},\ K_{-2},\ K_{-1}.$$

Note that there are essentially five orbits.

It is well-known that knowledge of the symmetry algebra of a differential equation permits one to obtain solutions of the equation via separation of variables.[13,14] Indeed, in our case for given $K \in \mathcal{G}$ and $\lambda \in R$ the system of equations

$$Ku = i\lambda u, \quad Xu = 0 \tag{1.15}$$

leads to a separation of variables in the Schrödinger equation. It is clear that two operators K, K' on the same G orbit lead to equivalent separation of variables via (1.15). Furthermore, since $K_{-2}u = iK_{-1}^2u$ whenever $Xu = 0$, the orbits containing K_{-1} and K_{-2} lead to essentially equivalent separations. Thus Eqs. (1.15) lead to separation of variables in four distinct coordinate systems associated with the orbit representatives K_3, L_3, $K_2 + K_{-1}$, and K_{-1}. In Sec. 2 we shall classify all coordinate systems in which variables separate for $Xu = 0$ and show that there exist only the four obtainable from (1.15). Thus separation of variables for $Xu = 0$ is explainable in terms of the symmetry algebra alone. (Note that for equations such as $u_{xx} + u_{yy} + k^2u = 0$ and $-iu_t = u_{xx} + u_{yy}$ it is necessary to use quadratic elements in the enveloping algebra of the symmetry algebra to describe separation of variables.[15,16]

The real six-dimensional symmetry algebra $\mathcal{G}''$ of the heat equation

$$u_t = u_{xx} \tag{1.16}$$

can be obtained by a computation analogous to that for the free-particle Schrödinger equation.[4] One finds that the operators

$$\begin{aligned} &K_2' = t^2\partial_t + tx\partial_x + t/2 + x^2/4, \quad K_1' = t\partial x + x/2, \\ &K_0' = 1, \quad K_{-1}' = \partial_x, \quad K_{-2}' = \partial_t, \quad K'^3 = x\partial_x + 2t\partial_t + \tfrac{1}{2} \end{aligned} \tag{1.17}$$

form a basis for $\mathcal{G}''$ where K_0' spans the center of $\mathcal{G}''$ and

$$[K'^3, K_j'] = jK_j', \quad j = \pm 2, \pm 1, 0, \quad [K_1', K_2'] = [K_{-1}', K_{-2}'] = 0$$

$$[K_{-1}', K_2'] = K_1', \quad [K_{-1}', K_1'] = \tfrac{1}{2}K_0, \quad [K_{-2}', K_1'] = K_{-1}',$$

$$[K_{-2}', K_2'] = K'^3.$$

There are five orbits in $\mathcal{G}''/\{K_0'\}$ under the adjoint representation with corresponding orbit representatives $K_3', K_2' + K_{-2}', K_2' + K_1', K_{-2}', K_{-1}'$. Since $K_{-2}' = (K_{-1}')^2$ for solutions of the heat equation, only four coordinate systems in which variables separate are associated with the five orbits.

2. SEPARATION OF VARIABLES FOR THE EQUATION $XU = 0$ AND THE HEAT EQUATION $u_t = u_{xx}$

In this section we examine the problem of the separation of variables for Eq. (1.2). As opposed to the corresponding problem for the Helmholtz equation there is no established method of approach here (i.e., no associated differential form and corresponding obvious, group of motions as in the case of, say, the Euclidean plane.[17]) We therefore proceed directly and examine the possibilities.

Choosing a new set of real variables v_1 and v_2 where

$$x = G(v_1, v_2), \quad t = H(v_1, v_2) \tag{2.1}$$

and G, H are real invertable functions, Eq. (1.2) can be written in the form

$$(a_{11}\partial_{11} + a_{12}\partial_{12} + a_{22}\partial_{22} + a_1\partial_1 + a_2\partial_2)u = 0, \tag{2.2}$$

where

$$a_{11} = \left(\frac{H_2}{D}\right)^2, \quad a_{12} = -\frac{2H_1H_2}{D^2}, \quad a_{22} = \left(\frac{H_1}{D}\right)^2,$$

and $D = G_1H_2 - H_1G_2$ (subscripts denote differentiation with respect to v_i), a_1 and a_2 are complicated functions whose explicit form we do not need for general G and H. From the form of (2.2) we see that a necessary condition for a separable solution (see definition below) of the form $u = A(v_1)B(v_2)$ is that at least one of the coefficients a_{11}, a_{12}, a_{22} be zero, i.e., either H_1 or H_2 is zero. Without loss of generality we can take $H_1 = 0$ and write $t = v_2$ (as H cannot then be a constant function). With these assumptions (1.2) assumes the form (2.2) where

$$a_{11} = \frac{1}{G_1^2}, \quad a_1 = \frac{iG_2}{G_1} - \frac{G_{11}}{G_1^3}, \quad a_2 = i \tag{2.3}$$

and all other coefficients zero. In order that this equation separate we have the additional constraints

$$\frac{1}{G_1} = f(v_2)g(v_1), \quad \frac{G_2}{G_1} = f^2(v_2)h(v_1). \tag{2.4}$$

From these equations we have

$$G_{12} = \frac{1}{g}\partial_2\left(\frac{1}{f}\right) = f\partial_1\left(\frac{h}{g}\right) \tag{2.5}$$

and hence

$$\frac{1}{f}\partial_2\left(\frac{1}{f}\right) = \tfrac{1}{2}b \tag{2.6}$$

with b a constant real number. There are two cases to consider:

(i) $b \neq 0$. Then $1/f = \sqrt{bv_2 + c}$. Without loss of generality we can take $c = 0$ as our defining equation is translation invariant. The function G then has the form $G = \bar{g}(v_1)v_2^{1/2}$ where $\bar{g}$ is a nonconstant real function. Accordingly we can define $\bar{g}(v_1) = v_1$. The system of coordinates is then

$$t = v_2, \quad x = v_1v_2^{1/2}. \tag{2.7}$$

(ii) $b = 0$. From the equation $G_2 = f(h/g)$ we see that $G = cv_2 + \bar{g}(v_1)$ and hence the coordinate system in this case is

$$t = v_2, \quad x = cv_2 + v_1. \tag{2.8}$$

One point that should be mentioned here is that the full equation does admit a separable solution when the functions A and B are exponentials and the new variables are given by

$$t = av_1 + bv_2, \quad x = cv_1 + dv_2 \tag{2.9}$$

with $ad - bc \neq 0$. In our definition of separation, however, we require that in the associated coordinate system the Eq. (1.2) can be replaced by two ordinary (nontrivial) differential equations in each of the separable variables. Then only the subclass of coordinates given by (2.8) is admissible as strictly separable. We accordingly make no further comment on the choice of variables (2.9).

In addition to considering separable coordinates for (1.2) it is also of interest to consider R-separable solutions of this equation. These are coordinates which admit solutions of the form $\exp[Q(v_1, v_2)]A(v_1)B(v_2)$ where Q is not expressible in the form $g(v_1) + h(v_2)$ and is not a constant. With the inclusion of such a multiplier term e^Q, Eq. (1.2) for the product $A(v_1)B(v_2)$ assumes the form (2.2) with an extra term a_0u added to the left-hand side. The conditions for R-separability are the same as for strict separability so that $a_{22} = a_{12} = 0$.

The nonzero coefficients are given by

$$\begin{aligned} &a_{11} = \frac{1}{G_1^2}, \quad a_1 = \frac{2Q_1}{G_1^2} - i\frac{G_2}{G_1} - \frac{G_{11}}{G_1^3}, \quad a_2 = i, \\ &a_0 = \frac{(Q_{11} + Q_1^2)}{G_1^2} - Q_1\left(i\frac{G_2}{G_1} + \frac{G_{11}}{G_1^3}\right) + iQ_2. \end{aligned} \tag{2.10}$$

The conditions for separability then become upon writing $Q = R + iS$ (R and S real)

$$1/G_1 = f(v_2)/g_1(v_1), \tag{2.11a}$$

$$2R_1/G_1^2 = f^2(v_2)w(v_1), \tag{2.11b}$$

$$(2S_1/G_1^2) - (G_2/G_1) = f^2(v_2)K(v_1). \tag{2.11c}$$

Equation (2.11b) allows us to take $R=0$, since its solution is of the form $r_1(v_1)+r_2(v_2)$. The remaining conditions simplify to

$$\frac{S_1^2}{G_1^2}-S_1\frac{G_2}{G_1}+S_2=f^2(v_2)q(v_1)+p(v_2), \tag{2.12a}$$

$$\frac{S_{11}}{G_1^2}-S_1\frac{G_{11}}{G_1^3}=f^2(v_2)\,r(v_1)+s(v_2). \tag{2.12b}$$

[Note: $g_1(v_1)=\partial_1 g(v_1)$ for some g.] From (2.11a) the form of G is $G=g/f+h(v_2)$ and $g\neq$ const. We are then free to take $g=v_1$. From (2.11c) we see that

$$2S_1=-\frac{f_2}{f^3}v_1+\frac{h_2}{f}+K. \tag{2.13}$$

We can therefore write the form of S as

$$S=-\frac{f_2}{4f^3}v_1^2+\frac{h_2}{2f}v_1. \tag{2.14}$$

[Remember that terms of the form $\bar{g}(v_1)+\bar{h}(v_2)$ in the expression for S can be dropped as they do not contribute to strict R-separation.] We now evaluate the possibilities.

(i) $f=$ const. Then we can put $f=1$. Equation (2.12a) implies $h_{22}=2a\neq 0$. Without loss of generality we can then take $h=av_2^2$. The corresponding coordinate system is

$$t=v_2,\quad x=v_1+av_2^2,\quad a>0, \tag{2.15}$$

and $S=av_1v_2$.

(ii) $f_2/f^3=-\frac{1}{2}a\neq 0$. In this case we can take $f=v_2^{-1/2}$, the constant a being absorbed in the definition of the variable v_1. Substitution into (2.12a) then requires $h_{22}=-\frac{1}{4}bv_2^{-3/2}$ for some constant b, so that

$$h=bv_2^{1/2}+cv_2. \tag{2.16}$$

We may take $b=0$ by redefining v_1. The resulting coordinate system is then

$$t=v_2,\quad x=v_1v_2^{1/2}+cv_2 \tag{2.17}$$

with

$$S=\tfrac{1}{2}cv_1v_2^{1/2}.$$

This is seen to be a generalization of the coordinate system (2.7).

(iii) $f_2/f^3\neq$ const. In this case, substituting into (2.12a) we obtain the equations given below as requirements for the functions f and h:

$$ff_{22}-2f_2^2=\alpha f^6, \tag{2.18a}$$

$$h_{22}=\beta f^3 \tag{2.18b}$$

with α,β real constants. We consider two possibilities.

(1) $\alpha=0$. In this case $f=av_2^{-1}$ and $h=b/v_2+cv_2$. In particular, we can take $a=1$ and $c=0$ effectively absorbing c into the definition of v_1. The resulting coordinate system is

$$t=v_2,\quad x=v_1v_2+\frac{b}{v_2},\quad b\geq 0, \tag{2.19}$$

with

$$S=\tfrac{1}{4}v_2v_1^2-bv_1/2v_2.$$

(2) $\alpha\neq 0$. In this case (2.18a) has the solution

$$f=(av_2^2+b)^{-1/2} \tag{2.20}$$

and h has a solution of the form $h=c(av_2^2+b)^{1/2}+dv_2$. We can put $c=0$, effectively absorbing this term in the definition of v_1. This results in two distinct types of coordinates depending on the relative sign of a and b.

(a) $$t=v_2,\quad x=v_1\sqrt{1+v_2^2}+dv_2 \tag{2.21}$$

where

$$S=\tfrac{1}{4}v_1^2v_2+\tfrac{1}{2}dv_1\sqrt{1+v_2^2}$$

and

(b) $$t=v_2,\quad x=v_1\sqrt{1-v_2^2}+dv_2 \tag{2.22a}$$

with

$$S=-\tfrac{1}{4}v_1^2v_2+\tfrac{1}{2}dv_1\sqrt{1-v_2^2},$$

$$t=v_2,\quad x=v_1\sqrt{v_2^2-1}+dv_2 \tag{2.22b}$$

with

$$S=\tfrac{1}{4}v_1^2v_2+\tfrac{1}{2}dv_1\sqrt{v_2^2-1}.$$

The coordinate system (b) is the only system which requires two distinct parametrizations to cover the entire range of variation of v_2. This then exhausts the classification of all coordinate systems which are R-separable and separable for (1.2). In particular, it is to be noticed that in each case the operator $X=\partial_{xx}+i\partial_t$ can be written $X=f(v_1,v_2)(L+K)$ where L and K are operators in v_1 and v_2, respectively. In particular, K is a first order operator such that $XKu=0$ and so can always be expressed as a linear combination of the generators K_i. In Table I we give all the coordinate systems we have found together with the associated operators K. It is clear that in this classification we have not made use of the full invariance group of (1.2) apart from translational invariance. If we do include this group in our definition of equivalence all the coordinate systems we have found are equivalent to ones whose representative basis defining operators are one of the forms (1.14). In particular, we see that under this equivalence more than one coordinate system may be on the same orbit. This is a consequence of the fact that the group action has not been accounted for in the classification of sep-

TABLE I. Separable coordinate systems for the Schrödinger equation $Xu=0$ and their associated basis defining operators. (Note only the x coordinate is given as we always have $t=v_2$.)

Coordinate system	Multiplier e^{iS}	Basis operator K
1. $x=cv_2+v_1,\ c\geq 0$	$S=0$	$K=K_{-2}+cK_{-1}$
2. $x=v_1+av_2^2,\ a>0$	$S=av_1v_2$	$K=K_{-2}-2aK_1$
3. $x=v_1v_2^{1/2}+cv_2,$ $c\in R$	$S=\frac{1}{2}cv_1v_2^{1/2}$	$K=K^3-cK_1$
4. $x=v_1v_2$ $+b/v_2,\ b\geq 0$	$S=\frac{1}{4}v_2v_1^2-bv_1/2v_2$	$K=K_2+2bK_{-1}$
5. $x=v_1\sqrt{1+v_2^2}$ $+dv_2,\ d\geq 0$	$S=\frac{1}{4}v_1^2v_2$ $+\frac{1}{2}dv_1\sqrt{1+v_2^2}$	$K=K_2-K_{-2}-dK_{-1}$
6. $x=v_1\sqrt{1-v_2^2}+dv_2$	$S=-\frac{1}{4}v_1^2v_2$ $+\frac{1}{2}dv_1\sqrt{1-v_2^2}$	$K=K_2+K_{-2}+dK_{-1}$
$x=v_1\sqrt{v_2^2-1}+dv_2,$ $d\geq 0$		$S=\frac{1}{4}v_1^2v_2+\frac{1}{2}dv_1\sqrt{v_2^2-1}$

TABLE II. Separable coordinate systems for the heat equation $U_t = U_{xx}$ (for all multipliers $S=0$).

Coordinate system	Multiplier	Operator
1. $x=v_1$	0	K'_{-2}
2. $x=v_1 v_2{}^{1/2}$	0	K'^3
3. $x=v_1\sqrt{1+v_2^2}$	$R=-\frac{1}{4}v_2 v_1^2$	$K'_2+K'_{-2}$
4. $x=v_1+\frac{1}{2}v_2^2$	$R=-\frac{1}{2}v_1 v_2$	$K'_{-2}+K'_1$

arable systems. In the next section we deal with those bases corresponding to inequivalent orbits. In that section we give the solutions of (1.2) in the corresponding coordinates.

Finally, in this section we list in Table II the separable coordinate systems for the heat equation (1.16) corresponding to representatives of the inequivalent orbits of basis defining symmetry operators.

3. ONE AND TWO-VARIABLE MODELS

We now show that the operators (1.5) can be interpreted as a Lie algebra of skew-Hermitian operators on the Hilbert space $L_2(R)$ of complex-valued Lebesgue square-integrable functions on the real line. To do this we consider t as a fixed parameter and, in view of (1.2), replace ∂_t by $i\partial_{xx}$ in expressions (1.5). It is easy to show that the resulting operators restricted to the domain of C^∞-functions with compact support and multiplied by i are symmetric and essentially self-adjoint. Indeed the operators (1.5) are real linear combinations of the operators

$$\mathcal{K}_2=ix^2/4,\quad \mathcal{K}_1=ix/2,\quad \mathcal{K}_0=i,\quad \mathcal{K}_{-1}=\partial_x,\quad \mathcal{K}_{-2}=i\partial_{xx}$$
$$\mathcal{K}^3=x\partial_x+\tfrac{1}{2} \tag{3.1}$$

and $i\mathcal{K}_j$, $i\mathcal{K}^3$ are essentially self-adjoint. Moreover, when the parameter t is set equal to zero, K_j becomes $\mathcal{K}_j$ and K^3 becomes $\mathcal{K}^3$. It follows that the operators $\mathcal{K}_j$, $\mathcal{K}_3$ satisfy the commutation relations (1.6).

From Stone's theorem[18] we know that to each skew-Hermitian $\mathcal{H}\in\mathcal{G}$ there corresponds a one-parameter group $U(\alpha)=\exp(\alpha\mathcal{H})$ of unitary operators on $L_2(R)$. This group in turn acts on $\mathcal{G}$ via $\mathcal{K}\to U(\alpha)\mathcal{K}U(-\alpha)$. In particular, one can easily verify that

$$\begin{aligned}&[\exp(t\mathcal{K}_{-2})]\mathcal{K}_j[\exp(-t\mathcal{K}_{-2})]=K_j,\\ &[\exp(t\mathcal{K}_{-2})]\mathcal{K}^3[\exp(-t\mathcal{K}_{-2})]=K^3.\end{aligned} \tag{3.2}$$

Thus if $f\in L_2(R)$ then $u=\exp(t\mathcal{K}_{-2})f$ satisfies $u_t=\mathcal{K}_{-2}u$ or $iu_t=-u_{xx}$ (for almost every t) whenever f is in the domain of $\mathcal{K}_{-2}$, and $u(0)=f$. Also it is easy to show that the unitary operators $\exp(\alpha K)$ $=\exp(t\mathcal{K}_{-2})\exp(\alpha\mathcal{K})\exp(-t\mathcal{K}_{-2})$ map such a u into v $=\exp(\alpha\mathcal{K})u$ which also satisfies $v_t=\mathcal{K}_{-2}v$. Thus the unitary operators $\exp(\alpha K)$ are symmetries of (1.2).

Later we will show that the operators $\mathcal{K}_j$, $\mathcal{K}^3$ generate a global unitary representation of the group G on $L_2(R)$. Assuming this for the moment, let $U(g)$, $g\in G$, be the corresponding unitary operators and set $T(g)$ $=\exp(t\mathcal{K}_{-2})U(g)\exp(-t\mathcal{K}_{-2})$. Again it is easy to demonstrate that the $T(g)$ are unitary symmetries of (1.2) and that the associated infinitesimal operators are $K=\exp(t\mathcal{K}_{-2})\mathcal{K}\exp(-t\mathcal{K}_{-2})$.

Now consider the operator $\mathcal{L}_3=\mathcal{K}_{-2}-\mathcal{K}_2=i\partial_{xx}-ix^2/4$ $\in\mathcal{G}$. If $f\in L_2(R)$ then $u(t)=\exp(t\mathcal{L}_3)f$ satisfies $u_t=\mathcal{L}_3 u$ or $iu_t=-u_{xx}+x^2u/4$ and $u(0)=f$. Similarly, the unitary operators $V(g)=\exp(t\mathcal{L}_3)U(g)\exp(-t\mathcal{L}_3)$ are symmetries of this equation, the Schrödinger equation for the harmonic oscillator, and one can verify that the associated infinitesimal operators $\exp(t\mathcal{L}_3)\mathcal{K}\exp(-t\mathcal{L}_3)$ can be expressed as first order differential operators in t and x. Continuing in this manner we consider the operator $\mathcal{L}_2=\mathcal{K}_{-2}+\mathcal{K}_2=i\partial_{xx}-ix^2/4\in\mathcal{G}$. If $f\in L_2(R)$ then $u(t)$ $=\exp(t\mathcal{L}_2)f$ satisfies $u_t=\mathcal{L}_2 u$ or $iu_t=-u_{xx}-x^2u/4$ and $u(0)=f$. The operators $W(g)=\exp(t\mathcal{L}_2)U(g)\exp(-t\mathcal{L}_2)$ form the unitary symmetry group of this equation, repulsive harmonic oscillator potential, and the associated infinitesimal operators $\exp(t\mathcal{L}_2)\mathcal{K}\exp(-t\mathcal{L}_2)$ are first order in x and t. Finally, we consider the operator $\mathcal{H}$ $=\mathcal{K}_{-2}-\mathcal{K}_1=i\partial_{xx}-ix/2\in\mathcal{G}$. If $f\in L_2(R)$ then $u(t)=\exp(t\mathcal{H})$ satisfies $u_t=\mathcal{H}_u$ or $iu_t\pm-u_{xx}+xu/2$ and $u(0)=f$. The unitary operators $X(g)=\exp(t\mathcal{H})\mathcal{K}\exp(-t\mathcal{H})$ are symmetries of this Schrödinger equation for the linear potential and the infinitesimal operators $\exp(t\mathcal{H})\mathcal{K}\exp(-t\mathcal{H})$ are first order in x and t.

Note further from (1.14) the operators $\mathcal{K}_{-2}$, $\mathcal{L}_3$, $\mathcal{L}_2$, and $\mathcal{K}_{-2}-\mathcal{K}_1$ corresponding to the free particle, attractive and repulsive harmonic oscillator, and linear potential Hamiltonians, lie on the same G orbits as the four representatives $\mathcal{K}_{-2}$, $\mathcal{L}_3$, $\mathcal{K}_3$ and $\mathcal{K}_2+\mathcal{K}_{-1}$, respectively. Thus these four Hamiltonians correspond exactly to the four systems of coordinates in which Eq. (1.2) separates. We see that these Hamiltonians form a complete set of orbit representatives in $\mathcal{G}$ in the sense explained following Eq. (1.15).

Note that if two operators lie on the same G orbit then the first operator is unitary equivalent to a real constant times the second operator. Thus two suitably normalized operators on the same orbit necessarily have the same spectrum. In particular, if $\mathcal{K},\mathcal{K}'\in\mathcal{G}$ with $\mathcal{K}'=U(g)\mathcal{K}U(g^{-1})$ and the self-adjoint operator $i\mathcal{K}$ has a complete set of (possibly generalized) eigenvectors $f_\lambda(x)$ with

$$i\mathcal{K}f_\lambda=\lambda f_\lambda,\quad (f_\lambda,f_\mu)=\delta_{\lambda\mu} \tag{3.3}$$

where

$$(h_1,h_2)=\int_{-\infty}^{\infty}h_1(x)\overline{h_2(x)}\,dx,\quad h_j\in L_2(R), \tag{3.4}$$

then for $f'_\lambda=U(g)f_\lambda$ we have

$$i\mathcal{K}'f'_\lambda=\lambda f'_\lambda,\ (f'_\lambda,f'_\mu)=\delta_{\lambda\mu} \tag{3.5}$$

and the f'_λ form a complete set of eigenvectors for $i\mathcal{K}'$.[19] These remarks imply that, if we wish to compute the spectrum corresponding to each operator $\mathcal{K}\in\mathcal{G}$, it is enough to determine the spectra of the four Hamiltonians listed above. Moreover, we may be able to choose another operator $\mathcal{K}$ on the same G orbit as a given Hamiltonian such that the spectral decomposition of $\mathcal{K}$ is especially easy. The spectral decomposition of the Hamiltonian and the corresponding eigenfunction expansions then follow from those of $\mathcal{K}$ by application of a group operator $U(g)$.

As a special case of these remarks consider the operator $\mathcal{K}_{-2}=i\partial_{xx}$. If $\{f_\lambda\}$ is the basis of generalized eigenvectors for some operator $\mathcal{K}\in\mathcal{G}$, then $\{f'_\lambda(t)$

$=\exp(t\mathcal{K}_{-2})f_\lambda\}$ is the basis of generalized eigenvectors for $K=\exp(t\mathcal{K}_{-2})\mathcal{K}\exp(-t\mathcal{K}_{-2})$ and the $f'_\lambda(t)$ satisfy the equation $iu_t=-u_{xx}$. Similar remarks hold for the other Hamiltonians.

We begin our explicit computations by determining the spectral resolution of the operator $\mathcal{L}_3=\mathcal{K}_{-2}-\mathcal{K}_2$. The results are well-known.[18] The eigenfunction equation is

$$i\mathcal{L}_3 f=\lambda f,\quad (-\partial_{xx}+x^2/4)f=\lambda f,$$

and the normalized eigenfunctions are

$$f^{(1)}_{\lambda_n}(x)=[n!\,\sqrt{2\pi}\,2^n]^{-1/2}\exp(-x^2/4)H_n(x2^{-1/2}),$$
$$\lambda_n=n+\tfrac{1}{2},\quad n=0,1,2,\ldots,\quad (f^{(1)}_{\lambda_n},f^{(1)}_{\lambda_m})=\delta_{nm} \tag{3.6}$$

where $H_n(x)$ is a Hermite polynomial.

It is now easy to show that the $\mathcal{K}$ operators exponentiate to a global unitary irreducible representation of G. Indeed, from the known recurrence formulas for the Hermite polynomials one can check that the operators $\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3$ acting on the $f^{(1)}$-basis define a reducible representation of $sl(2,R)$ belonging to the discrete series. The value of the Casimir operator is $\frac{1}{4}(\mathcal{L}_1^2+\mathcal{L}_2^2-\mathcal{L}_3^2)=-3/16$. As first shown by Bargmann,[20] this Lie algebra representation extends to a global unitary reducible representation of $SL(2,R)$. Similarly, the operators $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{L}_3$, acting on the $f^{(1)}$-basis define the irreducible representation $(\lambda,l)=(-\frac{1}{2},1)$ of the Lie algebra of the harmonic oscillator group S.[21] Again this Lie algebra representation is known to generate a global unitary irreducible representation of S.[21,22] Finally, since every unitary operator from $SL(2,R)$ can be written in the form $\exp(\alpha\mathcal{L}_3)\exp(\beta\mathcal{L}_1)\exp(\gamma\mathcal{L}_3)$,[20] where $\exp(\alpha\mathcal{L}_3)$ also belongs to S, and since $\mathcal{L}_1$ is a first order operator whose exponential is easily determined, we can check that the identity (1.12) holds in general. Thus our representation of G extends to a global unitary representation U of G which is irreducible since $U|S$ is already irreducible. The matrix elements of the operators $U(g)$ in the $f^{(1)}$-basis can be found in numerous references, e.g., Refs. 20, 22, 23.

The unitary operators $U(g)$ on $L_2(R)$ are easily computed. The operators

$$U(u,v,\rho)=\exp[\rho+(uv/4)]\;\mathcal{E}\exp(u\mathcal{S}_2)\exp(v\mathcal{S}_1)$$

defining an irreducible representation of W take the form

$$[U(u,v,\rho)f](x)=\exp\left[i\left(\rho+\frac{uv}{4}+\frac{ux}{2}\right)\right]f(x+v),\quad f\in L_2(R). \tag{3.7}$$

The operators $U(A)$, $A\in SL(2,R)$, are more complicated. From Ref. 24 (p. 493) we have

$$\exp(a\mathcal{K}_{-2})f(x)$$
$$=\text{l.i.m.}\,\frac{1}{\sqrt{4\pi i a}}\int_{-\infty}^{\infty}\exp[-(x-y)^2/4ia]f(y)\,dy, \tag{3.8}$$

and it is elementary to show

$$\exp(b\mathcal{K}^3)f(x)=\exp(b/2)f(e^b x),$$
$$\exp(c\mathcal{K}_2)f(x)=\exp(icx^2/4)f(x). \tag{3.9}$$

Relations (1.11) imply

$$\exp(\phi\mathcal{L}_2)=\exp(\tanh\phi\,\mathcal{K}_2)\exp(\sinh\phi\cosh\phi\,\mathcal{K}_{-2})$$
$$\times\exp(-\ln\cosh\phi\,\mathcal{K}_3),$$

so (3.8) and (3.9) yield

$$\exp(\phi\mathcal{L}_2)f(x)=\frac{\exp[(ix^2/4)\tanh\phi]}{(4\pi i\sinh\phi)^{1/2}}$$
$$\times\,\text{l.i.m.}\int_{-\infty}^{\infty}\exp[-(x-y\cosh\phi)^2/4i\sinh\phi\cosh\phi]f(y)\,dy. \tag{3.10}$$

A similar computation for $\exp(\theta\mathcal{L}_3)$ gives

$$\exp(\theta\mathcal{L}_3)f(x)=\frac{\exp[(ix^2/4)\cot\theta]}{(4\pi i\sin\theta)^{1/2}}$$
$$\times\,\text{l.i.m.}\int_{-\infty}^{\infty}\exp[-(y^2\cos\theta-2xy)/4i\sin\theta]f(y)\,dy. \tag{3.11}$$

Using (3.8) we see that the basis functions $f^{(1)}_{\lambda_n}(x)$ map to the ON basis functions $F^{(1)}_{\lambda_n}(x,t)=\exp(t\mathcal{K}_{-2})f^{(1)}_{\lambda_n}(x)$ or

$$F^{(1)}_{\lambda_n}(x,t)=[n!\,2^n\sqrt{2\pi(1+t^2)}]^{-1/2}\exp\left(\frac{i}{4}\frac{x^2t}{1+t^2}-\frac{x^2}{4(1+t^2)}\right.$$
$$\left.-i\lambda_n\arctan t\right)H_n[x/\sqrt{2(1+t^2)}] \tag{3.12}$$

which are solutions of (1.2).

Next we study the spectral theory for the orbit containing the operators $\mathcal{K}_{-2}+\mathcal{K}_2$ (repulsive oscillator) and $\mathcal{K}_3$. Since the spectral analysis for $\mathcal{K}_3$ is elementary we study it first. [The corresponding results for $\mathcal{K}_{-2}+\mathcal{K}_2$ then follow by application of an appropriate group operators $U(g)$.] The eigenfunction equation is

$$i\mathcal{K}^3 f=\lambda f,\quad \mathcal{K}^3=x\partial_x+\tfrac{1}{2}.$$

The spectral resolution for this operator is well-known.[25] It is obtained by considering $L_2(R)$ as the direct sum $L_2(R+)\oplus L_2(R-)$ of square-integrable functions on the positive and negative reals, respectively, and taking the Mellin transform of each component. Then $i\mathcal{K}_3$ transforms into multiplication by the transform variable. The spectrum is continuous and covers the real axis with multiplicity two. The generalized eigenfunctions are

$$f^{(2)\pm}_\lambda(x)=\frac{1}{\sqrt{2\pi}}\,x_\pm^{-i\lambda-1/2},\quad \lambda\in R, \tag{3.13}$$
$$(f^{(2)\pm}_\lambda,f^{(2)\pm}_\mu)=\delta(\mu-\lambda),\quad (f^{(2)\pm}_\lambda,f^{(2)\mp}_\mu)=0,$$

where

$$x_+^\alpha=\begin{cases}x^\alpha & \text{if } x>0\\ 0 & \text{if } x<0,\end{cases}\qquad x_-^\alpha=\begin{cases}0 & \text{if } x>0\\ (-x)^\alpha & \text{if } x<0\end{cases}.$$

From (3.8) we find $\exp(t\mathcal{K}_{-2})f^{(2)\pm}_\lambda=F^{(2)\pm}_\lambda(x,t)$ where

$$F^{(2)\pm}_\lambda(x,t)=\exp\left(\frac{x^2}{4it}+\frac{\pi\lambda}{4}+\frac{i\pi}{8}\right)$$
$$\times\frac{(2t)^{-i\lambda/2+1/4}}{\sqrt{8\pi^2 it}}\,\Gamma(\tfrac{1}{2}-i\lambda)D_{i\lambda-\frac{1}{2}}\left(\frac{-xe^{-i\pi/4}}{\sqrt{2t}}\right),$$
$$t>0, \tag{3.14}$$

$\Gamma(z)$ is a gamma function, and $D_\nu(z)$ is a parabolic cylinder function.[10] [These results follow from (3.8) by moving the integration contour from the positive real axis to a ray making an angle of $\pi/4$ with the real axis. We can also use the fact that we know the differential equations characterizing the function (3.14).] Also, we have

$$\text{(a) } F_\lambda^{(2)+}(x,t)=F_{-\lambda}^{(2)+}(x,-t), \qquad \text{(b) } F_\lambda^{(2)-}(x,t)=F_\lambda^{(2)+}(-x,t). \tag{3.15}$$

It follows immediately from (3.13) that

$$(F_\lambda^{(2)\pm},F_\mu^{(2)\pm})=\delta(\mu-\lambda), \quad (F_\lambda^{(2)\pm},F_\mu^{(2)\mp})=0. \tag{3.16}$$

Application of these orthogonality and completeness relations to expand an arbitrary $f\in L_2(R)$ yields the Hilbert space version of Cherry's theorem,[10,26] which is an expansion in terms of parabolic cylinder functions. Note that our expansion is simply related to the spectral resolution of the operator $K^3=2t\partial_t+x\partial_x+\frac{1}{2}=2it\partial_{xx}+x\partial_x+\frac{1}{2}$.

The next orbit we consider contains the operators $\mathcal{K}_{-2}+\mathcal{K}_1$ (linear potential) and $\mathcal{K}_2+\mathcal{K}_{-1}$. Since the spectral analysis for the second operators is simpler, we study it. The eigenfunction equation is

$$i(\mathcal{K}_2+\mathcal{K}_{-1})f=\lambda f, \quad \mathcal{K}_2+\mathcal{K}_{-1}=ix^2/4+\partial_x.$$

The spectral resolution is easily obtained from the Fourier integral theorem. The spectrum is continuous and covers the real axis, and the generalized eigenfunctions are

$$f_\lambda^{(3)}(x)=\frac{1}{\sqrt{2\pi}}\exp[-i(\lambda x+x^3/12)], \quad \lambda\in R, \tag{3.17}$$

$$(f_\lambda^{(3)},f_\mu^{(3)})=\delta(\mu-\lambda).$$

We find that

$$F_\lambda^{(3)}(x,t)=\exp(-i\pi/4)\,2^{1/6}\exp\left[\frac{i}{4}\left(-\frac{1}{8v_2^2}+v_2v_1^2-\frac{v_1}{v_2}\right)-\frac{i\lambda}{v_2}\right]\times\mathrm{Ai}[2^{2/3}(\tfrac{1}{2}v_1+\lambda)] \tag{3.18}$$

with v_1 and v_2 as in Table I, system 4 with $b=\frac{1}{2}$.

Ai(z) is a Airy function. These are the basis functions for the operator $K_2+K_{-1}=-it^2\partial_{xx}+(1-tx)\partial_x-t/2+ix^2/4$. For the orbit containing K_{-1} the complete set of eigenfunctions is

$$f^{(4)}=\frac{1}{\sqrt{2\pi}}\exp(-i\lambda x), \quad \lambda\in R, \tag{3.19}$$

with the usual orthogonality properties. It is not hard to show that

$$F_\lambda^{(4)}(x,t)=\frac{1}{\sqrt{2\pi}}\exp[i(\lambda^2t-\lambda x)]. \tag{3.20}$$

The case of the remaining orbit K_{-2} differs so little from this last case that we do not treat it here.

If $\{f_\lambda(x)\}$ is a basis of (generalized) eigenfunctions of some $\mathcal{K}\in\mathcal{G}$ and $F_\lambda(x,t)=\exp(t\mathcal{K}_{-2})f_\lambda(x)$ then $F_\lambda(x,\tau)=\exp([\tau-t]\mathcal{K}_{-2})\,F_\lambda(x,t)$ and we have the Hilbert space expansions

$$k(x-y,t)=\int F_\lambda(x,t)\overline{f_\lambda(y)}\,d\lambda, \tag{3.21}$$

$$k(x-y,\tau-t)=\int F_\lambda(x,\tau)\overline{F_\lambda(y,t)}\,d\lambda$$

where the integration domain is the spectrum of $i\mathcal{K}$ and

$$k(x,t)=\frac{1}{\sqrt{4\pi it}}\exp(-x^2/4it)$$

is the kernel of the integral operator $\exp(t\mathcal{K}_{-2})$. These expansions are known as continuous generating functions.[7,8]

4. OVERLAP FUNCTIONS

In this section we compute the overlap functions $(f_\lambda^{(i)},f_\mu^{(j)})$ which allow us to expand eigenfunctions $f_\lambda^{(i)}$ in terms of eigenfunctions $f_\mu^{(j)}$. Since $(U(g)f_\lambda^{(i)},U(g)f_\mu^{(j)})=(f_\lambda^{(i)},f_\mu^{(j)})$, the same expressions allow us to expand eigenfunctions $U(g)f_\lambda^{(i)}$ in terms of eigenfunctions $U(g)f_\mu^{(j)}$. We give here then those overlap functions corresponding to bases $f_\lambda^{(i)}$ that we have taken as standard:

$$(f_{\lambda_n}^{(1)},f_\lambda^{(2)\pm})=\frac{(\pm 2)^{n+i\lambda-1/2}\,\Gamma(i\lambda/2+\frac{1}{4}+\frac{1}{2}n)}{2\pi\sqrt{2^n n!}}\times{}_2F_1(-\tfrac{1}{2}n,\tfrac{1}{2}-\tfrac{1}{2}n,\tfrac{3}{4}-i\lambda/2-\tfrac{1}{2}n;\tfrac{1}{2}). \tag{4.1}$$

For the calculation of the overlap functions $(f_{\lambda_n}^{(1)},f_\lambda^{(3)})$ it is convenient to give a generating function rather than an explicit expression. The result is

$$2^{2/3}\exp[-i(\tfrac{1}{6}+\lambda+\sqrt{2y})]\,\mathrm{Ai}[2^{2/3}(\tfrac{1}{4}-i\lambda-i\sqrt{2y})] = \sum_{n=0}^{\infty}\frac{(\sqrt{2}iy)^n}{\sqrt{n!}}(f_{\lambda_n}^{(1)},f_\lambda^{(3)}). \tag{4.2}$$

This expression follows from the form of the generating function of Hermite polynomials given by Ref. 10.

$$(f_{\lambda_n}^{(1)},f_\lambda^{(4)})=[n!(-2)^n\pi]^{-1/2}\exp(-\lambda^2)H_n(\sqrt{2}\lambda), \tag{4.3}$$

$$(f_{\lambda'}^{(3)},f_\lambda^{(2)+}) = \frac{1}{2\pi}(12i)^{(1/6-i\lambda/3)}\sum_{n=0}^{\infty}\frac{\Gamma\{[(n-i\lambda)/3]+\frac{1}{6}\}}{n!}\times[\exp(5i\pi/6)\lambda']^n(12)^{n/3}, \tag{4.4}$$

where

$$(f_{\lambda'}^{(3)},f_\lambda^{(2)-})=(-1)^{i\lambda-1/2}\overline{(f_{\lambda'}^{(3)},f_{-\lambda}^{(2)+})}. \tag{4.5}$$

$$(f_{\lambda'}^{(3)},f_\lambda^{(4)})=2^{2/3}\mathrm{Ai}(2^{2/3}[\lambda-\lambda']),$$

$$(f_{\lambda'}^{(4)},f_{\lambda'}^{(2)\pm})=\frac{\pm 1}{2\pi}\exp(\mp i\lambda'\pi/2)\,\Gamma(-\lambda'+1)(\lambda\pm i0)^{-\lambda'-1}. \tag{4.6}$$

The general overlap function relating an eigenbasis on one orbit to an eigenbasis on another orbit is of the form $(U(g)f_\lambda^{(i)},f_\mu^{(j)})$. Indeed, a general eigenbasis $\{h_\lambda^{(i)}\}$ on orbit i can be expressed as $h_\lambda^{(i)}=U(g_h)f_\lambda^{(i)}$. Thus, $(h_\lambda^{(i)},k_\mu^{(j)})=(U(g_h)f_\lambda^{(i)},U(g_k)f_\mu^{(j)})=(U(g_k^{-1}g_h)f_\lambda^{(i)},f_\mu^{(j)})$. These expressions are known as "mixed basis matrix elements."[27] Their knowledge allows us to expand any eigenfunction of an operator in $\mathcal{G}$ in terms of eigenfunctions of any other operator in $\mathcal{G}$. Since the inner product is invariant under the unitary operators $U(g)$, the knowledge of the matrix elements for fixed i,j, and g can lead to a variety of different expansions. We shall not tabulate these elements here but merely note that they are of some interest. Indeed, they yield Hilbert space analogies of the analytic function expansions derived by Weisner in Ref. 6. However, the Hilbert space theory is richer and more complicated since one can derive expansions in all bases, not just Hermite function bases as used by Weisner.

As an example we give the mixed basis elements:

$$(\exp(t\mathcal{K}_{-2})f^{(1)}_{\lambda_n},\ f^{(2)\pm}_{\mu})=(f^{(1)}_{\lambda_n},\ \exp(-t\mathcal{K}_{-2})f^{(2)\pm}_{\mu})$$

$$=\frac{(\pm 2)^{n+i\mu-1/2}(1+it)^{i\mu/2}\exp(-i\lambda_n \arctan t)}{2\pi\sqrt{2^n n!}\,(1-it)^{n/2+i\mu/2+1/4}}$$

$$\times\Gamma\left(\frac{i\mu}{2}+\frac{1}{4}+\frac{n}{2}\right)$$

$$\times{}_2F_1\left(-\frac{n}{2},\frac{1}{2}-\frac{n}{2},\frac{3}{4}-\frac{i\mu}{2}-\frac{n}{2};\frac{1-it}{2}\right).$$

These elements allow us to expand Hermite polynomials as an integral over parabolic cylinder functions and parabolic cylinder functions in series of Hermite polynomials.

5. THE EQUATION $iu_t + u_x - cu/x^2 = 0$

Here we apply the methods discussed in the previous sections to the differential operator

$$Y=i\partial_t+\partial_{xx}-c/x^2,\quad c\neq 0. \tag{5.1}$$

We first compute the maximal symmetry algebra of the equation $Yu=0$. Thus, we find all operators L, Eq. (1.3), such that $Y(Lu)=0$ whenever $Yu=0$. A straightforward calculation shows that the symmetry algebra $\mathcal{H}^c$ is three-dimensional with basis

$$K_{-2}=\partial_t,\quad K_2=-t^2\partial_t-tx\partial_x-t/2+ix^2/4,\qquad(5.2)$$
$$K^3=2t\partial_t+x\partial_x+1/2$$

and commutation relations

$$[K^3,K_{\pm2}]=\pm 2K_{\pm2},\quad [K_2,K_{-2}]=K^3.$$

For the basis L_j where

$$L_1=K^3,\quad L_2=K_{-2}+K_2,\quad L_3=K_{-2}-K_2,$$

we have the relations

$$[L_1,L_2]=-2L_3,\quad [L_3,L_1]=2L_2,\quad [L_3,L_2]=-2L_1. \tag{5.3}$$

It is clear that the real Lie algebra generated by these basis elements is $sl(2,R)$. The corresponding group action of $SL(2,R)$ on functions $f(x,t)$ is given by the operators (1.10), and the explicit relation between the group and Lie algebra operators by (1.11).

The group $SL(2,R)$ acts on $sl(2,R)$ via the adjoint representation and splits the Lie algebra into orbits. Let

$$K=A_2K_2+A_{-2}K_{-2}+A_3K^3\in sl(2,R)$$

and set $\alpha=A_2A_{-2}+A_3^2$. It is straightforward to check that α is invariant under the adjoint representation and that K lies on the same $SL(2,R)$ orbit as a real multiple of exactly one of the three operators in the following list:

$$\begin{aligned}&\text{Case } 1(\alpha<0):\ K_{-2}-K_2=L_3,\\ &\text{Case } 2(\alpha>0):\ K^3,\\ &\text{Case } 3(\alpha=0):\ K_2.\end{aligned}\tag{5.4}$$

We see that there are essentially three orbits.

The evaluation of all separable coordinate systems proceeds as for the free particle case except that now we have the added restriction that $G_1/G=h(u_1)$. The resulting coordinate systems, multipliers, and basis defining operator are then listed in Table III.

In analogy with our argument in Sec. 3 we can interpret the operators (5.2) as a Lie algebra of skew-Hermitian operators on the Hilbert space $L_2(R+)$ of complex-valued Lebesgue square-integrable functions $f(x)$ on the positive real line, $0<x<\infty$. This is accomplished by considering t as a fixed parameter and replacing ∂_t by $i\partial_{xx}-ic/x^2$ in expressions (5.2). The resulting operators when multiplied by i and restricted to the domain of C^∞ functions with compact support in $R+$ are via Weyl's lemma,[28] easily seen to be essentially self-adjoint provided $c\geq 2$. In the remainder of this paper we assume that the constant c satisfies this inequality. The operators $K_{\pm2},K^3$ are real linear combinations of the skew-Hermitian operators

$$\mathcal{K}_{-2}=i\partial_{xx}-ic/x^2,\quad \mathcal{K}_2=ix^2/4,\quad \mathcal{K}^3=x\partial_x+1/2 \tag{5.5}$$

to which they reduce when $t=0$. Similarly, the skew-Hermitian operators

$$\mathcal{L}_1=\mathcal{K}^3=x\partial_x+\tfrac{1}{2},\quad \mathcal{L}_2=\mathcal{K}_{-2}+\mathcal{K}_2=i\partial_{xx}-ic/x^2+ix^2/4, \tag{5.6}$$

$$\mathcal{L}_3=\mathcal{K}_{-2}-\mathcal{K}_2=i\partial_{xx}-ic/x^2-ix^2/4$$

satisfy relations (5.3) and the L_j reduce to $\mathcal{L}_j$ when $t=0$.

In analogy with Sec. 3, one finds

$$\begin{aligned}&\exp(t\mathcal{K}_{-2})\mathcal{K}_j\exp(-t\mathcal{K}_{-2})=K_j,\\ &\exp(t\mathcal{K}_{-2})\mathcal{L}_j\exp(-t\mathcal{K}_{-2})=L_j.\end{aligned}\tag{5.7}$$

Thus for any $f\in L_2(R+)$ the vector $u(t)=\exp(t\mathcal{K}_{-2})f$ satisfies $u_t=\mathcal{K}_{-2}u$ or $iu_t=-u_{xx}+cu/x^2$ and $u(0)=f$. Also the unitary operators $\exp(\alpha K)$ $=\exp(t\mathcal{K}_{-2})\exp(\alpha\mathcal{K})\exp(-t\mathcal{K}_{-2})$, $\mathcal{K}\in sl(2,R)$, map solutions of the equation $u_t=\mathcal{K}_{-2}u$ into other solutions.

We will soon demonstrate that the operators $\mathcal{K}_{\pm2}$, $\mathcal{K}^3$ generate a global unitary irreducible representation of the universal covering group J of $SL(2,R)$ by operators $U(g)$, $g\in J$, on $L_2(R+)$. Assuming this we see that the operators $T(g)=\exp(t\mathcal{K}_{-2})U(g)\exp(-t\mathcal{K}_{-2})$ define a group of unitary symmetries of the equation $Yu=0$, with associated infinitesimal operators $K=\exp(t\mathcal{K}_{-2})\mathcal{K}\exp(-t\mathcal{K}_{-2})$. This discussion shows the relationship between our Lie algebra of $\mathcal{K}$-operators and the Schrödinger equation for the radial free particle.

Next consider the operator $\mathcal{L}_3\in sl(2,R)$. If $f\in L_2(R+)$ then $u(t)=\exp(t\mathcal{L}_3)f$ satisfies $u_t=\mathcal{L}_3u$ or $iu_t=-u_{xx}+cu/x^2+x^2u/4$, the Schrödinger equation for the radial harmonic oscillator. The unitary operators $V(g)$ $=\exp(t\mathcal{L}_3)U(g)\exp(-t\mathcal{L}_3)$ are symmetries of this equation and the associated infinitesimal operators

TABLE III. Separable coordinate systems for the equation $Yu=0$.

Coordinate	Multiplier e^{iS}	Basis operator
1. $x=v_1$	$S=0$	K_{-2}
2. $x=v_1v_2^{1/2}$	$S=0$	K^3
3. $x=v_1v_2$	$S=\frac{1}{4}v_2v_1^2$	K_2
4. $x=v_1\sqrt{1+v_2^2}$	$S=\frac{1}{4}v_2v_1^2$	K_2-K_{-2}
5. $x=v_1\sqrt{\pm(1-v_2^2)}$	$S=\pm\frac{1}{4}v_2v_1^2$	K_2+K_{-2}

$\exp(t\mathcal{L}_3)\mathcal{K}\exp(-t\mathcal{L}_3)$ are first order linear differential operators in x and t. Similarly, if $f\in L_2(R+)$ then $u(t)=\exp(t\mathcal{L}_2)f$ satisfies $u_t=\mathcal{L}_2u$ or $iu_t=-u_{xx}+cu/x^2-x^2u/4$, the Schrödinger equation for the repulsive radial oscillator. The operators $W(g)=\exp(t\mathcal{L}_2)U(g)\exp(-t\mathcal{L}_2)$ determine the symmetry group of this equation and the associated infinitesimal operators $\exp(t\mathcal{L}_2)\mathcal{K}\exp(-t\mathcal{L}_2)$ are first order in x and t.

From (5.4) it follows that the operators $\mathcal{K}_{-2},\mathcal{L}_3,\mathcal{L}_2$ corresponding to the radial free particle, attractive and repulsive harmonic oscillator Hamiltonians lie on the same J orbits, as the three orbit representatives $\mathcal{K}_2$, $\mathcal{L}_3$ and $\mathcal{K}^3$, respectively. Our three Hamiltonians correspond to the three J orbits of $sl(2,R)$. The remarks concerning expressions (3.3)–(3.5) and the invariance of spectra for operators on an orbit carry over without change to this case except that the inner product is now

$$(h_1,h_2)=\int_0^\infty h_1(x)\overline{h_2(x)}\,dx,\quad h_j\in L_2(R+). \tag{5.8}$$

Note that if $\{f_\lambda\}$ is the basis of generalized eigenvectors for some $\mathcal{K}\in sl(2,R)$ then $\{f'_\lambda(t)=(\exp t\mathcal{K}_{-2})f_\lambda\}$ is the basis of eigenvectors for $K=\exp(t\mathcal{K}_{-2})\mathcal{K}\exp(-t\mathcal{K}_{-2})$ and the $f'_\lambda(t)$ satisfy the Schrödinger equation for the radial free particle. Similar remarks hold for the other Hamiltonians.

We first present the well-known results for the spectrum of $\mathcal{L}_3$. The eigenfunction equation is

$$i\mathcal{L}_3f=\lambda f,\quad (-\partial_{xx}+c/x^2+x^2/4)f=\lambda f$$

and the normalized eigenfunctions are

$$f^{(1)}_{\lambda_n}(x)=\left(\frac{n!\,2^{-\mu/2}}{\Gamma(n+1+\mu/2)}\right)^{1/2}e^{-x^2/4}x^{(\mu+1)/2}L_n^{(\mu/2)}(x^2/2), \tag{5.9}$$

$$\lambda_n=-2n-\mu/2-1,\quad c=(\mu^2-1)/4,\quad \mu\geq 3,$$

$$n=0,1,2,\cdots,$$

where $L_n^{(\alpha)}(z)$ is a generalized Laguerre polynomial. The $\{f^{(1)}_{\lambda_n}\}$ form an ON basis for $L_2(R+)$.

Using the recurrence relations for the Laguerre polynomials one can check that the operators $\mathcal{L}_j$ acting on the $f^{(1)}$ basis define an irreducible representation of $sl(2,R)$ belonging to the discrete series. The Casimir operator is $\frac14(\mathcal{L}_1^2+\mathcal{L}_2^2-\mathcal{L}_3^2)=-3/16+c/4$. As is well-known,[20,23] this Lie algebra representation extends to a global unitary irreducible representation of J. The matrix elements of the operators $U(g)$ in a $f^{(1)}$ basis can be found in Refs. 23 or 29.

We now compute the operators $U(g)$ directly. Clearly,

$$\exp(a\mathcal{K}^3)f(x)=\exp(a/2)f(e^ax),$$

$$\exp(\alpha\mathcal{K}_2)f(x)=\exp(i\alpha x^2/4)f(x).$$

Furthermore,

$$\exp(\beta\mathcal{L}_3)f(x)=\frac{\exp(\mp i\pi(\mu+2)/4}{2|\sin\beta|}\ \text{l.i.m.}\int_0^\infty(xy)^{1/2}\times\exp\left(\pm\frac{i}{4}(x^2+y^2)\left|\cot\beta\right|\right)\times J_{\mu/2}\left(\frac{xy}{2|\sin\beta|}\right)f(y)\,dy,\quad 0<|\beta|<\pi, \tag{5.10}$$

where we take the upper sign for $\beta>0$ and the lower for $\beta<0$. [Here $J_\mu(z)$ is a Bessel function.] The additional relation $\exp(\pi\mathcal{L}_3)=\exp[-i\pi(1+\mu/2)]$ allows us to determine $\exp(\beta\mathcal{L}_3)$ for any β. To prove these results we apply the integral operator (5.10) to an $f^{(1)}$ basis element, and use the Hille–Hardy formula[22] and the fact that $\exp(\beta\mathcal{L}_3)f^{(1)}_{\lambda_n}=\exp[-i(2n+\mu/2+1)\beta]f^{(1)}_{\lambda_n}$ to check its validity. Since (5.10) is valid on an ON basis and $\exp(\beta\mathcal{L}_3)$ is unitary, the expression must be true for all $f\in L_2(R+)$.

The group multiplication formula

$$\exp\gamma\mathcal{K}_{-2}=\exp(-\sin\theta\cos\theta\mathcal{K}_2)\exp(\ln\cos\theta\mathcal{K}^3)\exp(\theta\mathcal{L}_3)$$

with $\gamma=\tan\theta$ and expressions (5.9), (5.10) easily yield

$$\exp(\gamma\mathcal{K}_{-2})f(x)=\frac{\exp[\mp(i/4)\pi(\mu+2)]}{2|\gamma|}\ \text{l.i.m.}\int_0^\infty(xy)^{1/2}\times\exp\left(\frac{i(x^2+y^2)}{4\gamma}\right)J_{\mu/2}\left(\frac{xy}{2|\gamma|}\right)f(y)\,dy, \tag{5.11}$$

where we take the upper sign for $\gamma>0$ and the lower for $\gamma<0$. A similar group theoretic calculation gives

$$\exp(\phi\mathcal{L}_2)f(x)=\frac{\exp[\mp(i/4)\pi(\mu+2)]}{2|\sinh\phi|}\ \text{l.i.m.}\int_0^\infty(xy)^{1/2}\times\exp\left(\frac{i}{4}(x^2+y^2)\coth\phi\right)\times J_{\mu/2}\left(\frac{xy}{2|\sinh\phi|}\right)f(y)\,dy. \tag{5.12}$$

From (5.11) we find that the basis functions $f^{(1)}_{\lambda_n}(x)$ map to the ON basis functions $F^{(1)}_{\lambda_n}(x,t)=\exp(t\mathcal{K}_{-2})f^{(1)}_{\lambda_n}(x)$

$$F^{(1)}_{\lambda_n}(x,t)=2(-1)^n\exp[\pm(i/4)\pi(\mu+2)]\left(\frac{x^2}{1+t^2}\right)^{(\mu+1)/4}\times(t-i)^{-\mu/4-3/4-n}(t+i)^{\mu/4+1/4+n}\times\exp\left(\frac14\frac{x^2}{1+t^2}(-1+i\gamma)\right)L_n^{\mu/2}\left(\frac12\frac{x^2}{1+t^2}\right)\quad\text{for } t\lessgtr 0 \tag{5.13}$$

which are solutions F of $YF=0$.

The J orbit containing the operator $\mathcal{L}_2$ (repulsive radial oscillator) also contains $\mathcal{K}^3$ so we merely study the spectral theory for $\mathcal{K}^3$. The results are well-known.[35] The eigenfunction equation is

$$i\mathcal{K}^3f=\lambda f,\quad \mathcal{K}^3=x\partial_x+\tfrac12.$$

The spectrum is continuous and covers the real axis with multiplicity one. The generalized eigenfunctions are

$$f^{(2)}_\lambda(x)=\frac{1}{\sqrt{2\pi}}x^{-i\lambda-1/2},\quad \lambda\in R, \tag{5.14}$$

$$(f^{(2)}_\lambda,f^{(2)}_\mu)=\delta(\mu-\lambda).$$

Again using (5.11) we find $F^{(2)}_\lambda(x,t)=\exp(t\mathcal{K}_2)f^{(2)}_\lambda(x)$ where

$$F_\lambda^{(2)}(x,t)=\frac{1}{\sqrt{2\pi}}\,\frac{\Gamma(i\lambda/2+\mu/4+\frac{1}{2})}{\Gamma(1+\mu/2)}\,\exp[\mp(\pi/4)(i\mu+i+\lambda)]$$

$$\times t^{i\lambda/2-1/4}(x^2/t)^{-1/4}\exp\left(\frac{ix^2}{8t}\right)M_{(i\lambda/2),(\mu/4)}\left(\frac{ix^2}{t}\right) \tag{5.15}$$

for $t \gtrless 0$. Here $M_{\chi,\eta}(z)$ is a solution of Whittaker's equation.[10] If follows from our procedure that the basis functions satisfy

$$(F_\lambda^{(2)},\ F_\mu^{(2)})=\delta(\mu-\lambda)$$

and can be used to expand any $f\in L_2(R+)$.

Finally, the orbit containing $\mathcal{K}_{-2}$, corresponding to the radial free particle, also contains $\mathcal{K}_2$. The spectral theory for $\mathcal{K}_2$ is elementary because $\mathcal{K}_2$ is already diagonalized in our realization. The generalized eigenfunctions are (symbolically)

$$f_\lambda^{(3)}=\delta(x-\lambda),\quad i\mathcal{K}_2 f_\lambda^{(3)}=(\lambda^2/4)f_\lambda^{(3)},\quad \lambda\geq 0.$$

The spectrum is continuous and covers the positive real axis with multiplicity one. We have

$$F_\lambda^{(3)}(x,t)=\exp(t\mathcal{K}_{-2})f_\lambda^{(3)}(x)$$

or

$$F_\lambda^{(3)}(x,t)=\frac{\exp(\mp i(\pi/4)(\mu+2)}{2|t|}\,(x\lambda)^{1/2}$$

$$\times\exp\left(\frac{i(x^2+\lambda^2)}{4t}\right)J_{\mu/2}\left(\frac{x\lambda}{2|t|}\right) \tag{5.17}$$

with $(F_\lambda^{(3)},\ F_\mu^{(3)})=\delta(\mu-\lambda)$. Expansions in the basis $\{F_\lambda^{(3)}\}$ are equivalent to the inversion theorem for the Hankel transform. The $F_\lambda^{(3)}$ are basis functions for the operator K_2.

Each of our bases has continuous generating functions of the form (3.19) where now

$$k(x,y,t)=\frac{\exp(\pm i(\pi/4)(\mu+2)}{2|t|}\,(xy)^{1/2}$$

$$\times\exp\left(\frac{i(x^2+y^2)}{4t}\right)J_{\mu/2}\left(\frac{xy}{2|t|}\right) \tag{5.18}$$

(see Ref. 8).

The overlap functions $(f_\lambda^{(i)},\ f_\mu^{(j)})$ have the same significance as in Sec. 4. Because of the simplicity of the basis $f_\lambda^{(3)}$ the only overlap of interest is

$$(f_{\lambda_n}^{(1)},\ f_\lambda^{(2)})=\tfrac{1}{2}\left(\frac{\Gamma(n+1+\frac{1}{2}\mu)2^{i\lambda}}{\pi n!}\right)^{1/2}\frac{\Gamma(i\lambda/2+\mu/4+\frac{1}{2})}{\Gamma(1+\frac{1}{2}\mu)}$$

$$\times {}_2F_1\left(-n,\frac{i\lambda}{2}+\frac{\mu}{4}+\frac{1}{2};\,1+\frac{1}{2}\mu;2\right). \tag{5.19}$$

In particular, we notice that the overlap functions are dependent on the representatives f_λ^{i}, $f_\mu^{(j)}$ that have been chosen on each orbit. From this we see that the most general way to define an overlap function is as the mixed basis matrix element $(f_\lambda^{(i)}, U(g)f_\mu^{(j)})$ where g is a general group element. This problem has been treated for the group $SL(2,R)$, Ref. 27, where a corresponding group parametrization has been given for each choice of $i\neq j$ in the above expression. In particular, the resulting expressions for the mixed basis matrix elements proved quite tractable to calculate and amounted to the calculation of the mixed basis matrix element of a one parameter subgroup in each case. We refer to the original article[27] for further details.

ACKNOWLEDGMENTS

We wish to thank Dr. T. Hida and Dr. C. Boyer for helpful discussions.

*Permanent address: School of Mathematics, University of Minnesota, Minneapolis, Minn. 55455.

[1]R. L. Anderson, S. Kumei, and C. E. Wulfman, Rev. Mex. Fis. 21, 1 (1972).
[2]C. P. Boyer, "The maximal kinematical invariance group for an arbitrary potential," preprint CIMAS 1973/53, Mexico.
[3]U. Niederer, Helv, Phys. Acta 46, 191 (1973); U. Niederer, "The group theoretical equivalence of the free particle, the Harmonic oscillator and the free fall," *Proceedings of the 2nd International Colloquium on Group Theoretical Methods in Physics*, 1973, University of Nijmegen, The Netherlands.
[4]G. W. Blumen and J. D. Cole, J. Math. Mech. 18, 1025 (1969).
[5]L. V. Ovsjannikov, *Gruppovye svoystva differentsialny uravneni*, (Novosibirsk, Moscow, 1962).
[6]L. Weisner, Can. J. Math. 11, 141 (1959).
[7]E. Whittaker, Proc. Roy. Soc. Edin. 61, A, 1 (1941).
[8]A. Erdélyi, Proc. Roy. Soc. Edin. 61, 61 (1941).
[9]A. O. Barut, SIAM J. Appl. Math. 25, 247 (1973); A. O. Barut and G. Bornzin, J. Math. Phys. 12, 841 (1971).
[10]A. Erdélyi *et al.*, *Higher Trancendental Functions* (McGraw-Hill, New York, 1953), Vols. 1, 2.
[11]W. Miller, Jr., SIAM J. Math. Anal. 4, 314 (1973).
[12]W. Miller, Jr., *Symmetry Groups and their Applications* (Academic, New York, 1973).
[13]G. Birkhoff, *Hydrodynamics* (Princeton U. P. Princeton, 1960).
[14]A. J. A. Morgan, Quart. J. Math. Oxford Ser. (2), 3, 250 (1952).
[15]P. Winternitz, I. Lukǎc, and Y. Smorodinskiĭ, Sov. J. Nucl. Phys. 7, 139 (1968).
[16]W. Miller, Jr., SIAM J. Math. Anal. (to appear).
[17]E. G. Kalnins, "On the separation of variables for the Laplace equation $\Delta\psi+k^2\psi=0$ in two and three dimensional Minkowski space," Université de Montréal preprint CRM-319.
[18]N. Dunford and J. Schwartz, *Linear Operators*. Part II (Wiley-Interscience, New York, 1963).
[19]K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (Polish Scientific, Warsaw, 1968).
[20]V. Bargmann, Ann. Math. 48, 569 (1947).
[21]W. Miller, Jr., J. Math. Phys. 13, 648 (1972).
[22]W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[23]P. Sally, "Analytic Continuation of the Irreducible Unitary Representations of the Universal Covering Group of SL(2,*R*), AMS Mem., No. 69, Providence, R.I., 1967.
[24]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[25]N. Y. Vilenkin *Special Functions and the Theory of Group Representations* (AMS Transl., Providence, R.I., 1968).
[26]T. M. Cherry, Proc. Edin. Math. Soc. (2), 8, 50 (1949).
[27]E. G. Kalnins, J. Math. Phys. 14, 654 (1973).
[28]I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1967), Vol. 1.
[29]W. Miller, Jr., J. Math. Phys. 13, 827 (1972).

The distribution of the zeroes of the Jost function: The s-wave attractive exponential potential

J. P. Lavine

*Physique Nucléaire Théorique, Institut de Physique, Université de Liège, Sart Tilman, 4000 Liège 1, Belgium**

Département de Physique, Université Laval, Québec, G1K 7P4, Canada†

(Received 19 March 1974)

We show how the zeroes of the Jost function for an s-wave attractive exponential potential are distributed. In particular, we use known results, especially some of Coulomb's, on the zeroes of Bessel functions to demonstrate that there are no zeroes for complex momentum $k=k_1+ik_2$ ($k_1\neq 0$, $k_2\neq 0$).

During a recent numerical investigation of the inverse scattering formalism of Gel'fand and Levitan,[1] we searched for efficient methods of evaluating the driving term of their integral equation for the kernel. The driving term is related to an integral[2] whose integrand involves the Jost function, $f_l(k)$, as $|f_l(k)|^{-2}$. A desire to deform the contour led us to investigate how the zeroes of the Jost function are distributed in the complex k plane. A finite-range potential's distribution is discussed by Newton,[3] where references to the original papers may be found. Sartori[4] has considered the case of s-wave potentials which vanish at infinity faster than any exponential, but his approach does not appear to be generalizable to potentials which vanish slower at infinity.

In this paper we determine the zero distribution for a s-wave attractive exponential potential. Since the Schrödinger equation is analytically solvable in this case,[5] it continues to be of interest in scattering theory.[6]

It is well known that the S matrix may be expressed in terms of the Jost function as

$$S(k)=e^{2i\delta(k)}=f(k)/f(-k). \tag{1}$$

We follow the convention used by Newton[3] here and the angular-momentum subscript is suppressed since we only deal with s waves.

Let us write the exponential potential as

$$V(r)=-V_0\exp(-r/a),\quad V_0>0. \tag{2}$$

Then, for the s wave the Jost function is[7]

$$f(k)=\exp[-iak\log(a^2V_0)]\,\Gamma(1+2iak)J_{2iak}(2aV_0^{1/2}). \tag{3}$$

We define $Z\equiv 2aV_0^{1/2}$ and we note that Z is real.

The poles of $f(k)$ come from the gamma function and occur when

$$1+2iak=0,-1,-2,\cdots$$

or

$$k=-in/(2a)\quad\text{for } n=-1,-2,\cdots. \tag{4}$$

These are the so-called redundant poles.[8] Equation (3) shows us that, since the gamma function is never equal to zero, the zeroes of $f(k)$ are the zeroes of

$$J_{2iak}(Z)=0. \tag{5}$$

Our task is to use what is known about Bessel functions to find these zeroes. We use as a reference the book by Gray and Mathews[9] and we consider the different sections of the complex k plane. Some useful relations involving Bessel functions are contained in the appendix.

First, let $\nu=2iak$ and let $k=-i\gamma$ with $\gamma>0$. Hence

$$J_\nu(Z)=J_{2a\gamma}(Z) \tag{6}$$

and for real Z such Bessel functions can have zeroes.[9] These are the bound states.

Now let k be real and greater than zero. A proof by contradiction[10] leads us to $f(k)\neq 0$ for real $k\neq 0$. This is a special case of a well-known general result,[11] and it can be derived in various ways. The point $k=0$ is special and although $f(0)$ may be equal to zero for s waves, this is not a bound state.[11] It is also well known[11] that a Jost function has no zeroes, other than bound states, in the lower-half k plane. For our particular case of a s wave, attractive exponential potential, this result is quickly seen with Eq. (A2) (with $b=0$ and $C=1$) and Eq. (A3), or with Ref. 12.

We now enter the upper k plane. Let us first put $k=i\gamma$ with $\gamma>0$, and define $\nu\equiv 2iak=-2a\gamma$. We momentarily assume the potential is too weak to have a bound state.

As γ increases toward plus infinity, ν will pass through negative integers, say $-n$. So we have

$$J_{-n}(Z)=(-1)^nJ_n(Z),$$
$$J_{-n-1}(Z)=(-1)^{n+1}J_{n+1}(Z)=-(-1)^nJ_{n+1}(Z). \tag{7}$$

When n is a positive integer, $n=2iak$ leads to a k which is on the lower half of the imaginary axis. Since we have assumed that there is no bound state, $J_n(Z)$ has the same sign for all n. Hence Eqs. (7) show that $J_{-n}(Z)$ and $J_{-n-1}(Z)$ have the opposite signs. Since $J_\nu(Z)$ is finite for all ν when $Z\neq 0$, $J_{-2a\gamma}(Z)$ has a zero between $\nu=-n$ and $\nu=-n-1$. This means that $J_{-2a\gamma}(Z)$ has an infinity of zeroes and these are the virtual states. Thus, $J_{2iak}(Z)$, and hence $f(k)$, equals zero an infinite number of times for k on the upper imaginary axis. If bound states are present then $J_{2a\gamma}(Z)$ changes sign a finite number of times, but $J_{-2a\gamma}(Z)$ will still have an infinity of zeroes. We remark that Coulomb[13] has shown that the zeroes of $J_\nu(Z)$, for ν real, asymptotically approach the negative integers, which are the redundant poles' locations according to Eq. (3).

Finally, we consider $k=k_1+ik_2$ with $k_2>0$ and $k_1\neq 0$. For this case

$$\nu\equiv 2iak=-2a|k_2|+2iak_1, \tag{8}$$

so that Re $(\nu)<0$. We follow Coulomb[13] and use Eq. (A2) with $b=1$ and let $C\to\infty$. We assume ν and Z are such

 Copyright © 1974 American Institute of Physics

that $J_\nu(Z)=0$, and hence $J_{\nu*}(Z)=0$. This means that the right-hand side of Eq. (A2) contributes zero for $b=1$. To evaluate the contribution as $C\to\infty$, we need[14]

$$J_\nu(CZ)\to(2/\pi CZ)^{1/2}\cos(CZ-\nu\pi/2-\pi/4). \quad (9)$$

Equation (9) implies that the $J_\nu(CZ)J_{\nu*}(CZ)$ term goes as C^{-1} when $C\to\infty$. Thus, the leading term is the square bracket of Eq. (A2). A bit of algebra shows that

$$XZ\,[J_{\nu+1}(XZ)J_{\nu*}(XZ)-J_\nu(XZ)J_{\nu*+1}(XZ)]_{c=X}\underset{c\to\infty}{\rightarrow}$$

$$(1/\pi)\{\cos[\pi/2+(\nu-\nu^*)\pi/2]-\cos[\pi/2-(\nu-\nu^*)\pi/2]\}. \quad (10)$$

Hence

$$\int_1^\infty J_\nu(XZ)J_{\nu*}(XZ)\,dX/X=(2/\pi)\sin[(\nu-\nu^*)\pi/2]/(\nu^2-\nu^{*2}). \quad (11)$$

Now the integrand of Eq. (11) is positive-definite; while for $\mathrm{Re}(\nu)<0$ and $\nu^2\neq\nu^{*2}$ the right-hand side of Eq. (11) is always negative. We arrive at a contradiction, which means that

$$J_{2iak}(Z)\neq 0 \quad (12)$$

when k is in the upper-half plane and $k_1\neq 0$. By Eq. (3) we see that the Jost function has no zeroes in the same region.

We have noted that for an attractive exponential potential, the s-wave Jost function is zero at bound states. For k in the upper-half plane, $f(k)$ can be zero only when k is on the imaginary axis. Thus, there are no s-wave resonances for this potential. This is an amusing contrast to the case of finite-range potentials,[11] where there are an infinity of resonances; but is similar to the situation for a Hulthén potential,[15] where there are no resonances either. In addition, the exponential and the Hulthén potentials both have an infinite number of virtual states, while finite-range potentials have only a finite number.[11] These similarities lead us to believe the above results may have generalizations, but attempts to treat repulsive exponential potentials and Morse potentials[16] have not been successful yet.

ACKNOWLEDGMENTS

We thank Dr. M. Bawin, Dr. J. Humblet, and Dr. H. S. Picker for useful conversations.

APPENDIX

A consideration of the series definition[17] of $J_\alpha(Z)$ shows that

$$J_{\alpha*}(Z)=(J_\alpha(Z))^* \quad (\mathrm{A1})$$

for α complex and Z real. Watson[18] derives the following helpful integral relationship:

$$\int_b^C J_\nu(XZ)J_{\nu*}(XZ)\,dX/X=\{-[XZ/(\nu^2-\nu^{*2})]$$
$$\times[J_{\nu+1}(XZ)J_{\nu*}(XZ)-J_\nu(XZ)J_{\nu*+1}(XZ)]$$
$$+J_\nu(XZ)J_{\nu*}(XZ)/(\nu+\nu^*)\}_b^C. \quad (\mathrm{A2})$$

We also invoke the relation[19]

$$J_{\alpha+1}(Z)=(\alpha/Z)J_\alpha(Z)-\frac{\partial}{\partial Z}J_\alpha(Z) \quad (\mathrm{A3})$$

when we treat a particular case of Eq. (A2).

*Chercheur Institut Interuniversitaire des Sciences Nucléaires of Belgium.
†Work supported in part by the Atomic Energy Control Board of Canada. Present address: Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627.

[1]L.D. Faddeev, J. Math. Phys. 4, 72 (1963) and references therein.
[2]H.S. Picker and J.P. Lavine, Phys. Rev. C 6, 1542 (1972), Eq. (6′).
[3]R.G. Newton, J. Math. Phys. 1, 319 (1960).
[4]L. Sartori, J. Math. Phys. 4, 1408 (1963).
[5]H.A. Bethe and R. Bacher, Rev. Mod. Phys. 8, 82 (1936).
[6]M.G. Fuda, J. Math. Phys. 12, 1163 (1971); M.S. Stern and A.E.A. Warburton, J. Phys. A 5, 1233 (1972).
[7]Ref. 3, Eq. (10.6).
[8]D. Ter Haar, Physica **XII**, 501 (1946); R. Jost, Helv. Phys. Acta 20, 256 (1947).
[9]A. Gray and G.B. Mathews, *A Treatise on Bessel Functions and Their Applications to Physics*, 2nd edition prepared by A. Gray and T.M. MacRobert (Dover, New York, 1966).
[10]Ref. 9, Eq. III-(42).
[11]Ref. 3, Sec. 4.
[12]Ref. 9, Eq. VI-(36).
[13]J. Coulomb, Bull. Sci. Math. 60, (71 of the series), 297 (1936).
[14]Ref. 9, Eq. V-(53).
[15]R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 421.
[16]P.M. Morse, J.B. Fisk, and L.I. Schiff, Phys. Rev. 50, 748 (1936).
[17]Ref. 9, Eq. II-(16).
[18]G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1958), p. 135.
[19]Ref. 9, Eq. II-(20).

The renormalized projection operator technique for quadratic stochastic differential equations.* II

Vincent A. LoDato

Marist Research Institute, Poughkeepsie, New York 12601

(Received 14 March 1973; final revised manuscript received 14 May 1974)

An equation for the mean value of the contaminant $\langle\Psi(\mathbf{x},t)\rangle$ is derived for the case when $\| L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\| \gg \|K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle\|$. The class of projection operators which produce this inequality is dictated by the following nonlinear stochastic equation:

$$
\begin{aligned}
&L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle - \nabla_{\mathbf{x}}\int\int d\mathbf{x}'\,dt'\,G_0(\mathbf{x},t\,|\,\mathbf{x}',t')\langle\delta v(\mathbf{x},t)\,\delta v(\mathbf{x}',t')\rangle\nabla_{\mathbf{x}'}\langle\Psi(\mathbf{x}',t')\rangle \\
&\quad = K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + K(\mathbf{x},t)\int\int d\mathbf{x}'\,dt'\,G_0(\mathbf{x}',t'\,|\,\mathbf{x}',t')\nabla_{\mathbf{x}'}\langle\Psi(\mathbf{x}',t')\rangle : \\
&\qquad \times\int\int d\mathbf{x}''\,dt''\,G_0(\mathbf{x}',t'\,|\,\mathbf{x}'',t'')\langle\delta v(\mathbf{x}',t')\,\delta v(\mathbf{x}'',t'')\rangle\nabla_{\mathbf{x}''}\langle\Psi(\mathbf{x}'',t'')\rangle + S(\mathbf{x},t).
\end{aligned}
$$

This is a valid approach when the reaction time is much greater than the transport time, $T_{\text{trans}} \ll T_{\text{react}}$.

1. INTRODUCTION

The solution of the nonlinear stochastic equation[1-8]

$$
\begin{aligned}
&[L_0(\mathbf{x},t) + L_1(\mathbf{x},t,\omega)]\Psi(\mathbf{x},t,\omega) \\
&\quad = S(\mathbf{x},t) + K^{(1)}(\mathbf{x},t)\Psi(\mathbf{x},t,\omega) + K(\mathbf{x},t)\Psi(\mathbf{x},t,\omega) : \Psi(\mathbf{x},t,\omega)
\end{aligned}
\tag{1.1}
$$

is extremely important, especially in its applications to the transport of contaminants in which chemistry is occurring. In this formal notation, $S(\mathbf{x},t)$ are the source and/or sink terms which are independent of the random variations and $K^{(1)}(\mathbf{x},t)$ and $K(\mathbf{x},t)$ are the reaction coefficient matrices for the unimolecular and bimolecular processes. Since the concentrations $\Psi(\mathbf{x},t,\omega)$ are contaminants, their effect on the temperature field is assumed to be small and hence, the K's are not dependent on the random variations. The operator $L_0(\mathbf{x},t)$ corresponds to a deterministic operator, e.g., the streaming operator D/Dt, while $L_1(\mathbf{x},t,\omega)$ is a stochastic operator (e.g., L_1 might be $\delta v(\mathbf{x},t,\omega)\cdot\nabla$, which is the random part of the convection term and in a well-known averaging process[8] approaches the eddy diffusivity term). The variable ω is assumed to span the sample space Ω and associated with it is the normalized probability density function $P(\omega)$. Consequently, when one solves Eq. (1.1), the only meaningful physical observable is the moments of contaminants and, in particular, its average value

$$\langle\Psi(\mathbf{x},t)\rangle = \int_\Omega \Psi(\mathbf{x},t,\omega)P(\omega)\,d\omega.$$

Since the equation is nonlinear and, in particular, quadratically nonlinear, due to the possibility of bimolecular kinetic processes, normal perturbation theory breaks down due to secularity[9,10] and it is extremely inconvenient to use due to the nonlinear processes, coupled to the fact that averaging must be performed on each term of the series. Thus, three problems arise: (1) how to truncate the series so that the equation generated adequately represents physical reality, (2) what is the proper averaging process, (3) the equations which are derived should be computationally feasible to solve or demonstrate some use for plausibility arguments.

A particular type of nonlinear equation of this type is the well-known Navier–Stokes equation. The similarity is that both are quadratically nonlinear. (The literature on this subject abounds,[11-15] especially notable is the work of Kraichnan.[16-20]) The problems which one encounters are the three mentioned above.

In this paper we shall briefly review and discuss the normal perturbation and hierarchal approaches to this problem, and then apply the renormalized projection operator (RPO) technique to this class of equations. Previously, we applied such an approach to the linear stochastic equation[21]

$$[L_0(\mathbf{x},t) + L_1(\mathbf{x},t,\omega)]\Psi(\mathbf{x},t,\omega) = S(\mathbf{x},t) \tag{1.2}$$

and derived approximate solutions to such an equation and arrived at the nearest neighbor and Kraichnan equations via a diagram technique and RPO technique.

2. THE PERTURBATION AND HIERARCHICAL APPROACH

Returning to Eq. (1.1),

$$
\begin{aligned}
&[L_0(\mathbf{x},t) - K^{(1)}(\mathbf{x},t) + \epsilon L_1(\mathbf{x},t,\omega)]\Psi(\mathbf{x},t,\omega) \\
&\quad = S(\mathbf{x},t) + \lambda K(\mathbf{x},t)\Psi(\mathbf{x},t,\omega) : \Psi(\mathbf{x},t,\omega),
\end{aligned}
\tag{2.1}
$$

we have introduced the dimensionless parameters ϵ and λ, which are measures of deviations from the equation $L_0(\mathbf{x},t)\Psi(\mathbf{x},t) = S(\mathbf{x},t)$. Obviously, when ϵ and λ are much less than unity, the Neumann expansion is valid, and one need only retain the first few terms in the series.

For the sake of convenience, let us redefine $L_0(\mathbf{x},t) - K^{(1)}(\mathbf{x},t)$ to be $L_0(\mathbf{x},t)$. Now the series expansion becomes

$$
\begin{aligned}
\Psi(\mathbf{x},t,\omega)
&= \Psi_0(\mathbf{x},t,\omega) + [L_0(\mathbf{x},t) + \epsilon L_1(\mathbf{x},t,\omega)]^{-1}S(\mathbf{x},t) \\
&\quad + [L_0(\mathbf{x},t) + \epsilon L_1(\mathbf{x},t,\omega)]^{-1}K(\mathbf{x},t)\Psi(\mathbf{x},t,\omega) : \Psi(\mathbf{x},t,\omega).
\end{aligned}
\tag{2.2}
$$

The terms $\Psi_0(\mathbf{x},t,\omega) + [L_0(\mathbf{x},t) + \epsilon L_1(\mathbf{x},t,\omega)]^{-1}S(\mathbf{x},t)$ correspond to the solution of Eq. (2.1) without the bimolecular reaction term; namely, $[L_0(\mathbf{x},t) + L_1(\mathbf{x},t,\omega)] \times\Psi_T(\mathbf{x},t,\omega) = S(\mathbf{x},t)$. These terms are defined by $\Psi_T(\mathbf{x},t,\omega)$, which is basically the solution of the transport processes coupled to unimolecular kinetics. Hence,

$$
\begin{aligned}
\Psi(\mathbf{x},t,\omega) &= \Psi_T(\mathbf{x},t,\omega) + \lambda[L_0(\mathbf{x},t) + \epsilon L_1(\mathbf{x},t,\omega)]^{-1} \\
&\quad \times K(\mathbf{x},t)\Psi(\mathbf{x},t,\omega) : \Psi(\mathbf{x},t,\omega).
\end{aligned}
\tag{2.3}
$$

Copyright © 1974 American Institute of Physics

Now, expanding the terms in powers of λ and assuming the weak statistical dependence approximation[21] and letting $\mathcal{L}(\mathbf{x},t,\omega)\equiv L_0(\mathbf{x},t)+L_1(\mathbf{x},t,\omega)$, one arrives at

$$\langle\Psi(\mathbf{x},t)\rangle=\langle\Psi_T(\mathbf{x},t)\rangle+\lambda\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle \times K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle+\mathcal{O}(\lambda^2) \tag{2.4a}$$

and

$$\langle\Psi(\mathbf{x},t)\rangle=\langle\Psi_T(\mathbf{x},t)\rangle+\lambda\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle + 2\lambda^2\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\mathcal{L}^{-1}(\mathbf{x},t) \times[K(\mathbf{x},t)\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)]\rangle+\mathcal{O}(\lambda^3); \tag{2.4b}$$

continuing, we have

$$\begin{aligned}\langle\Psi(\mathbf{x},t)\rangle=\langle\Psi_T(\mathbf{x},t)\rangle&+\lambda\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle\\ &+2\lambda^2\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\mathcal{L}^{-1}(\mathbf{x},t)\\ &\times[K(\mathbf{x},t)\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)]\rangle\\ &+\lambda^3\{4\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\mathcal{L}^{-1}(\mathbf{x},t)\\ &\times[K(\mathbf{x},t)\Psi_T(\mathbf{x},t):(\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle\\ &\times K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle)]\rangle\\ &+\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\mathcal{L}^{-1}(\mathbf{x},t)[K(\mathbf{x},t)\\ &\times\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)]:\mathcal{L}^{-1}(\mathbf{x},t)\\ &\times[K(\mathbf{x},t)\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)]\rangle\}+\mathcal{O}(\lambda^4),\end{aligned} \tag{2.4c}$$

etc.

From the perturbation approach of the problem, not much information can be gained and as we progress to higher order terms the complexity increases enormously. Hence, it becomes near impossible to test the error term (or the next higher order term that one is neglecting). Also, if secularity is the problem, such an approach is useless. Hence, for most practical problems, the equations generated are not computationally feasible nor can any information be gained concerning a plausibility analysis. We also compounded the problem by the averaging process assumed. All in all, one can say that such an approach is highly impractical unless $\lambda\ll 1$ and Eq. (2.4a) is valid, i.e.,

$$\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle^{-1}\langle\Psi(\mathbf{x},t)\rangle = S(\mathbf{x},t)+\lambda K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle+\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle^{-1}\langle\psi_0(\mathbf{x},t)\rangle, \tag{2.5}$$

then the perturbation approach is valid and useful.

Pursuing these same lines, Eq. (2.1) may be rewritten as

$$\Psi(\mathbf{x},t,\omega)=\Psi^{(I)}(\mathbf{x},t)-\epsilon L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega) +\lambda L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\Psi(\mathbf{x},t,\omega):\Psi(\mathbf{x},t,\omega) \tag{2.6}$$

where $\Psi^{(1)}(\mathbf{x},t)$ is the solution of $L_0(\mathbf{x},t)\Psi^{(I)}(\mathbf{x},t)=S(\mathbf{x},t)$. If we operate on the left-hand sides by $\Psi(\mathbf{x},t,\omega)$: and average, one obtains

$$\begin{aligned}\langle\Psi(\mathbf{x},t):\Psi(\mathbf{x},t)\rangle &=\langle\Psi(\mathbf{x},t)\rangle:\Psi^{(I)}(\mathbf{x},t)\\ &-\epsilon\langle\Psi(\mathbf{x},t):L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\Psi(\mathbf{x},t)\rangle\\ &+\lambda\langle\Psi(\mathbf{x},t):[L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\Psi(\mathbf{x},t):\Psi(\mathbf{x},t)]\rangle+\mathcal{O}(\lambda^2).\end{aligned} \tag{2.7}$$

Now multiplying Eq. (2.6) by $L_1(\mathbf{x},t,\omega)$ and averaging and again assuming the weak statistical dependence (WSD) approximation with the definition $\langle L_1(\mathbf{x},t)\rangle=0$, one finds

$$\begin{aligned}\langle L_1(\mathbf{x},t)\Psi(\mathbf{x},t)\rangle &\approx\langle L_1(\mathbf{x},t)\rangle\Psi^{(I)}(\mathbf{x},t)-\epsilon\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\langle\Psi(\mathbf{x},t)\rangle\\ &\quad+\lambda\langle L_1(\mathbf{x},t)\rangle L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\Psi(\mathbf{x},t):\Psi(\mathbf{x},t)\rangle\\ &\approx-\epsilon\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\langle\Psi(\mathbf{x},t)\rangle.\end{aligned} \tag{2.8}$$

Again returning to Eq. (2.6) and averaging, we have

$$\langle\Psi(\mathbf{x},t)\rangle=\Psi^{(I)}(\mathbf{x},t)-\epsilon L_0^{-1}(\mathbf{x},t)\langle L_1(\mathbf{x},t)\Psi(\mathbf{x},t)\rangle +\lambda L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\Psi(\mathbf{x},t):\Psi(\mathbf{x},t)\rangle. \tag{2.9}$$

Neglecting terms of $\mathcal{O}(\lambda^2)$ and $\mathcal{O}(\epsilon\lambda)$, substituting Eqs. (2.7) and (2.8) into Eq. (2.11), and simplifying, gives the result

$$[L_0(\mathbf{x},t)-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle-\lambda K(\mathbf{x},t)\Psi^{(I)}(\mathbf{x},t):] \times\langle\Psi(\mathbf{x},t)\rangle=S(\mathbf{x},t) \tag{2.10}$$

or

$$\langle\Psi(\mathbf{x},t)\rangle=[L_0(\mathbf{x},t)-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle-\lambda K(\mathbf{x},t) \times\Psi^{(I)}(\mathbf{x},t):]^{-1}[\Phi_{\text{int}}(\mathbf{x},t)+S(\mathbf{x},t)], \tag{2.11}$$

where Φ_{int} is the initial condition of the homogeneous part and is assumed to be deterministic. If it was a stochastic function it must be included in the averaging process. It is interesting to note that the term $L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle$ is the smoothing operator derived by Keller,[22] and if the approximation to the operator $\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle^{-1}$ is $L_0(\mathbf{x},t)-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle$, then with a little rearranging in Eq. (2.11), one finds

$$\langle\Psi(\mathbf{x},t)\rangle=\{1-\lambda(L_0(\mathbf{x},t)-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1} \times K(\mathbf{x},t)\Psi^{(I)}(\mathbf{x},t):\}^{-1}\langle\Psi_T(\mathbf{x},t)\rangle \tag{2.12}$$

and since λ is small one finds

$$\langle\Psi(\mathbf{x},t)\rangle=\{1+\lambda(L_0-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1} \times K(\mathbf{x},t)\Psi^{(I)}(\mathbf{x},t):\}\langle\Psi_T(\mathbf{x},t)\rangle \tag{2.13}$$

or

$$[L_0(\mathbf{x},t)-\epsilon^2\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle]\langle\Psi(\mathbf{x},t)\rangle =S(\mathbf{x},t)+\lambda K(\mathbf{x},t)\Psi^{(I)}(\mathbf{x},t):\langle\Psi_T(\mathbf{x},t)\rangle. \tag{2.14}$$

In this result we have an inhomogeneous equation to solve and the original nonlinear part is approximated by inhomogeneous and averaged transport concentrations.

If one has the condition that $\Psi(\mathbf{x},t,\omega):\Psi_T(\mathbf{x},t,\omega) \gg\lambda\Psi(\mathbf{x},t,\omega):\mathcal{L}^{-1}(\mathbf{x},t,\omega)K(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega):\Psi(\mathbf{x},t,\omega)$, then following along in the same manner as before

$$\Psi(\mathbf{x},t,\omega):\Psi(\mathbf{x},t,\omega)\approx\Psi_T(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega); \tag{2.15}$$

then

$$[L_0(\mathbf{x},t)+\epsilon L_1(\mathbf{x},t,\omega)-\lambda K(\mathbf{x},t)\Psi_T(\mathbf{x},t,\omega):]\Psi(\mathbf{x},t,\omega)=S(\mathbf{x},t) \tag{2.16}$$

and solving for $\Psi(\mathbf{x},t,\omega)$ gives

$$\Psi(\mathbf{x},t,\omega) =\{1-\lambda[L_0(\mathbf{x},t)+\epsilon L_1(\mathbf{x},t,\omega)]^{-1}K(\mathbf{x},t)\Psi_T(\mathbf{x},t,\omega):\}^{-1}$$

$$\times\Psi_T(\mathbf{x},t,\omega). \tag{2.17}$$

Again, if λ is small and the weak statistical dependence approximation is valid, then

$$\langle\Psi(\mathbf{x},t)\rangle = \langle\Psi_T(\mathbf{x},t)\rangle + \lambda\langle \mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle. \tag{2.18}$$

We see that equation (2.18) is the same as (2.4a). However, if λ is not small, then equation (2.17) becomes

$$\begin{aligned}\Psi(\mathbf{x},t,\omega) &= \Psi_T(\mathbf{x},t,\omega) + \lambda(L_0(\mathbf{x},t)+\epsilon L_1(\mathbf{x},t,\omega))^{-1}K(\mathbf{x},t)\Psi_T(\mathbf{x},t,\omega): \\ &\times\Psi(\mathbf{x},t,\omega) + \sum_{k'=2}^{\infty}\lambda^{k'}\{[L_0(\mathbf{x},t)+\epsilon L_1(\mathbf{x},t,\omega)]^{1} \\ &\times K(\mathbf{x},t)\Psi_T(\mathbf{x},t,\omega):]\}^{k'}\Psi_T(\mathbf{x},t,\omega)\end{aligned} \tag{2.19}$$

and averaging in the WSD approximations gives

$$\begin{aligned}\langle\Psi(\mathbf{x},t)\rangle &= \langle\Psi_T(\mathbf{x},t)\rangle + \lambda\langle L_0(\mathbf{x},t)+\epsilon L_1(\mathbf{x},t)\rangle \\ &\times K(\mathbf{x},t)\langle\Psi_T(\mathbf{x},t):\Psi_T(\mathbf{x},t)\rangle \\ &+\sum_{k'=2}^{\infty}\lambda^{k'}\langle[\langle\mathcal{L}^{-1}(\mathbf{x},t)\rangle K(\mathbf{x},t)\Psi_T(\mathbf{x},t):]^{k'}\Psi_T(\mathbf{x},t)\rangle.\end{aligned} \tag{2.20}$$

3. THE RENORMALIZED PROJECTION OPERATOR

Returning to Eq. (1.1) (and for the sake of simplicity incorporate the unimolecular term in L_0), let us define the solution $\Psi(\mathbf{x},t,\omega)$ in terms of an averaged part plus a fluctuating component. Within the averaged solution a projection operator $\langle P(\mathbf{x},t)\rangle$ is defined. This has the effect of telescoping the mean value $\langle\Psi(\mathbf{x},t)\rangle$, and it will approach the first smoothing solution when $\langle P(\mathbf{x},t)\rangle \ll 1$.

Now, let the solution of $\Psi(\mathbf{x},t,\omega)$ be

$$\begin{aligned}\Psi(\mathbf{x},t,\omega) &= [1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle] \\ &\times\langle\Phi(\mathbf{x},t)\rangle + \delta\Psi(\mathbf{x},t,\omega).\end{aligned} \tag{3.1}$$

Substituting Eq. (3.1) into Eq. (1.1) and averaging, gives the result

$$\begin{aligned}&L_0(\mathbf{x},t)[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle] \\ &\times\langle\Phi(\mathbf{x},t)\rangle = -\langle L_1(\mathbf{x},t)\delta\Psi(\mathbf{x},t)\rangle + S(\mathbf{x},t) \\ &\qquad+\{K(\mathbf{x},t)[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t) \\ &\qquad\times L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle] \\ &\qquad\times\langle\Phi(\mathbf{x},t)\rangle\}:\{[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t) \\ &\qquad\times L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t)\rangle\} \\ &\qquad+K(\mathbf{x},t)\langle\delta\Psi(\mathbf{x},t):\delta\Psi(\mathbf{x},t)\rangle.\end{aligned} \tag{3.2}$$

By substracting Eq. (1.1) from Eq. (3.2), one finally obtains

$$\begin{aligned}&L_0(\mathbf{x},t)\delta\Psi(\mathbf{x},t,\omega)+L_1(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega) \\ &=K(\mathbf{x},t)\{\Psi(\mathbf{x},t,\omega):\Psi(\mathbf{x},t,\omega)-\langle\delta\Psi(\mathbf{x},t):\Psi(\mathbf{x},t)\rangle\} \\ &\quad+\langle L_1(\mathbf{x},t)\delta\Psi(\mathbf{x},t)\rangle - K(\mathbf{x},t)\{[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t) \\ &\quad\times L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t)\rangle\}: \\ &\quad\times\{[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1} \\ &\quad\times\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t)\rangle\}\end{aligned} \tag{3.3}$$

or

$$\begin{aligned}\delta\Psi(\mathbf{x},t,\omega) &= \beta^{-1}L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\{\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega) \\ &-\langle\delta\Psi(\mathbf{x},t):\delta\Psi(\mathbf{x},t)\rangle\} - \beta^{-1}L_0^{-1}(\mathbf{x},t) \\ &\times\{L_1(\mathbf{x},t,\omega)\delta\Psi(\mathbf{x},t,\omega)-\langle L_1(\mathbf{x},t)\delta\Psi(\mathbf{x},t)\rangle\} \\ &-\beta^{-1}L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t,\omega)[1-(L_0(\mathbf{x},t) \\ &\times\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t)\rangle,\end{aligned} \tag{3.4}$$

where

$$\begin{aligned}\beta \equiv \{&1-L_0(\mathbf{x},t)-2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)[1-L_0(\mathbf{x},t)\langle L_1(\mathbf{x},t) \\ &\times L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t)\rangle:\};\end{aligned} \tag{3.5}$$

furthermore, if we define $\alpha(\mathbf{x},t)\equiv L_0(\mathbf{x},t)\beta$ and $\hat{G}\equiv G-\langle G\rangle$, then

$$\begin{aligned}\delta\Psi(\mathbf{x},t,\omega) &= \alpha(\mathbf{x},t)^{-1}K(\mathbf{x},t)\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega) \\ &-\alpha(\mathbf{x},t)^{-1}L_1(\mathbf{x},t,\omega)\delta\Psi(\mathbf{x},t,\omega)-\alpha(\mathbf{x},t)^{-1} \\ &\times L_1(\mathbf{x},t,\omega)[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t) \\ &\times L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\langle\Phi(\mathbf{x},t,\omega)\rangle.\end{aligned} \tag{3.6}$$

If $\alpha(\mathbf{x},t)\equiv L_0(\mathbf{x},t)[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]$, then one may solve Eq. (1.1) for $\langle\Phi(\mathbf{x},t)\rangle$ and upon averaging, one arrives at

$$\begin{aligned}\langle\Phi(\mathbf{x},t)\rangle &= \Phi_0(\mathbf{x},t)+\alpha^{-1}(\mathbf{x},t)S(\mathbf{x},t)-\alpha^{-1}(\mathbf{x},t)\langle L_1(\mathbf{x},t)\delta\Psi(\mathbf{x},t)\rangle \\ &+\alpha^{-1}(\mathbf{x},t)K(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle : L_0^{-1}(\mathbf{x},t) \\ &\times\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle+\alpha^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\delta\Psi(\mathbf{x},t):\delta\Psi(\mathbf{x},t)\rangle.\end{aligned} \tag{3.7}$$

The problem now resolves itself to finding the solution of Eqs. (3.6) and (3.7). However, let us define the operator Ω, which when operating on $\delta\Psi(\mathbf{x},t,\omega)$ produces

$$\begin{aligned}\Omega_0\delta\Psi(\mathbf{x},t,\omega) &\equiv \alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t,\omega)\delta\Psi(\mathbf{x},t,\omega) \\ &= \alpha^{-1}(\mathbf{x},t)[L_1(\mathbf{x},t,\omega)\delta\Psi(\mathbf{x},t,\omega) \\ &\quad-\langle L_1(\mathbf{x},t)\delta\Psi(\mathbf{x},t)\rangle]\end{aligned} \tag{3.8}$$

and

$$\begin{aligned}\Omega_0\Psi(\mathbf{x},t,\omega) &\equiv \alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega) \\ &= \alpha^{-1}(\mathbf{x},t)[L_1(\mathbf{x},t,\omega)\Psi(\mathbf{x},t,\omega) \\ &\quad-\langle L_1(\mathbf{x},t)\Psi(\mathbf{x},t)\rangle].\end{aligned} \tag{3.9}$$

This operator Ω_0 has the property that $\langle\Omega_0 f\rangle = 0$, and also $\langle\Omega_0^{N+1}f\rangle = 0$, where f is a bounded random function. Also, let us define Ω_1, which is

$$\begin{aligned}&\Omega_1\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega) \\ &\equiv\alpha^{-1}(\mathbf{x},t)K(\mathbf{x},t)\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega) \\ &=\alpha^{-1}(\mathbf{x},t)K(\mathbf{x},t)\{\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega) \\ &\quad-\langle\delta\Psi(\mathbf{x},t):\delta\Psi(\mathbf{x},t)\rangle\}.\end{aligned} \tag{3.10}$$

This also has the property $\langle\Omega_1\delta\Psi(\mathbf{x},t):\delta\Psi(\mathbf{x},t)\rangle = 0$, and also $\langle\Omega_1^{N+1}f\rangle = 0$ for all N. Consequently, the solutions for Eqs. (3.6) and (3.7) are

$$\begin{aligned}\delta\Psi(\mathbf{x},t,\omega) &= \Omega_1\delta\Psi(\mathbf{x},t,\omega):\delta\Psi(\mathbf{x},t,\omega)-\Omega_0\delta\Psi(\mathbf{x},t,\omega) \\ &-\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t,\omega)L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}\langle\Phi(\mathbf{x},t)\rangle &= \Phi_0(\mathbf{x},t)+\alpha^{-1}(\mathbf{x},t)S(\mathbf{x},t)+\alpha^{-1}(\mathbf{x},t)K(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\\ &\times\langle\Phi(\mathbf{x},t)\rangle : L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)+\alpha^{-1}(\mathbf{x},t)\langle L_1(\mathbf{x},t)\\ &\times\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t) < +\alpha^{-1}(\mathbf{x},t)\\ &\times\{K(\mathbf{x},t)\langle\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle :\\ &\times\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle\}\rangle + O(\Omega_0,\Omega_1).\end{aligned} \quad (3.12)$$

Since $L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle=\alpha(\mathbf{x},t)\langle\Phi(\mathbf{x},t)\rangle$, then

$$\begin{aligned}L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle &= K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + S(\mathbf{x},t)\\ &+\langle L_1(\mathbf{x},t)\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\langle\Psi(\mathbf{x},t)\rangle\\ &+K(\mathbf{x},t)\langle\{\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \alpha^{-1}(\mathbf{x},t)\\ &\times L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\}\rangle\end{aligned} \quad (3.13)$$

or

$$\begin{aligned}&[L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)]\langle\Psi(\mathbf{x},t)\rangle\\ &=K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + S(\mathbf{x},t)+K(\mathbf{x},t)\\ &\times\langle\{\alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \alpha^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\}\rangle.\end{aligned} \quad (3.14)$$

Recalling that $\alpha(\mathbf{x},t)=L_0(\mathbf{x},t)\{1-2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\times[1-(L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle)^{-1}\langle P(\mathbf{x},t)\rangle]\times\langle\Phi(\mathbf{x},t)\rangle :\}$, or simply, $\alpha(\mathbf{x},t)=L_0(\mathbf{x},t)\{1-2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle :\}$, then Eq. (3.14) becomes

$$\begin{aligned}&[L_0(\mathbf{x},t)-\langle L_1(\mathbf{x},t)L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle]\langle\Psi(\mathbf{x},t)\rangle\\ &=K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + S(\mathbf{x},t)\\ &+K(\mathbf{x},t)\langle L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\\ &\times\langle\Psi(\mathbf{x},t)\rangle+\sum_{j=1}^{\infty}K(\mathbf{x},t)\langle L_0(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\\ &\times(2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle :)^j L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\\ &+K(\mathbf{x},t)\langle(2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle :)^j L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\\ &\times\langle\Psi(\mathbf{x},t)\rangle : L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle + K(\mathbf{x},t)\langle(2L_0^{-1}(\mathbf{x},t)K(\mathbf{x},t)\\ &\times\langle\Psi(\mathbf{x},t)\rangle :)^j L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \sum_{j'=1}^{\infty}(2L_0^{-1}(\mathbf{x},t)\\ &\times K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle)^{j'}L_0^{-1}(\mathbf{x},t)L_1(\mathbf{x},t)\rangle\}\langle\Psi(\mathbf{x},t)\rangle.\end{aligned} \quad (3.15)$$

If one has the further limit that $\|L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\| \gg \|K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle\|$, and if $L_1(\mathbf{x},t,\omega)\equiv\delta v(\mathbf{x},t,\omega)\cdot\nabla_{\mathbf{x}}$, then

$$\begin{aligned}&L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle-\nabla_{\mathbf{x}}\int\int d\mathbf{x}'\,dt' G_0(\mathbf{x},t\,|\,\mathbf{x}',t')\langle\delta v(\mathbf{x},t)\delta v(\mathbf{x}',t')\rangle\\ &\times\nabla_{\mathbf{x}'}\langle\Psi(\mathbf{x}',t')\rangle = K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + K(\mathbf{x},t)\\ &\times\int\int d\mathbf{x}'\,dt' G_0(\mathbf{x},t\,|\,\mathbf{x}',t')\nabla_{\mathbf{x}'}\langle\Psi(\mathbf{x}',t')\rangle :\\ &\times\int\int d\mathbf{x}''\,dt'' G_0(\mathbf{x}',t'\,|\,\mathbf{x}'',t'')\langle\delta v(\mathbf{x}',t')\\ &\times\delta v(\mathbf{x}'',t'')\rangle\nabla_{\mathbf{x}''}\langle\Psi(\mathbf{x}'',t'')\rangle + S(\mathbf{x},t).\end{aligned} \quad (3.16)$$

The validity of the equation occurs for the class of projection operators which produce the inequality $\|L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\| \gg \|K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle\|$, thus the choice of operator $\langle P(\mathbf{x},t)\rangle$ need not necessarily be unique. This is a necessary condition for the validity of Eq. (3.16). In most problems of contaminants in which the transport time is less than the time for chemical reactions, this inequality holds, and Eq. (3.16) is valid and in the case when $\|L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle\| \gg \|K(\mathbf{x},t)\times\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle\|$, then

$$\begin{aligned}L_0(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle &= \nabla_{\mathbf{x}}\int\int d\mathbf{x}'\,dt' G_0(\mathbf{x},t\,|\,\mathbf{x}',t')\\ &\times\langle\delta v(\mathbf{x},t)\delta v(\mathbf{x}',t')\rangle\nabla_{\mathbf{x}'}\langle\Psi(\mathbf{x}',t')\rangle\\ &+K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle + S(\mathbf{x},t).\end{aligned} \quad (3.17)$$

The coupling of the velocity correlation term to the kinetic processes is applicable when there is not a great inequality between the different correlation times. Still, under a further limit when the kinetic processes are not dominant, $\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle$ can be replaced by $\langle\Psi^{(I)}(\mathbf{x},t)\rangle : \langle\Psi^{(I)}(\mathbf{x},t)\rangle$, where the last term is the solution of the transport equation without kinetic processes. The equation (3.17) then becomes an inhomogeneous type and its validity is in the first Born approximation. From this formalism the next terms in the series [the last term in Eq. (3.16)] correspond to propagating the mean value $\langle\Psi(\mathbf{x},t)\rangle$ from points $(\mathbf{x}',t')$ and $(\mathbf{x}'',t'')$ in configuration space to $(\mathbf{x},t)$ and $(\mathbf{x}',t')$ in which there is coupling due to velocity fluctuations in the domains $[\mathbf{x}',\mathbf{x}'']$ and $[t',t'']$, and chemistry occurring at $(\mathbf{x},t)$.

CONCLUSION

The renormalized projection operator technique can produce a class of equations which are valid within the choice of the operator $\langle P(\mathbf{x},t)\rangle$. Given the operator $\langle P(\mathbf{x},t)\rangle$, it is possible to derive a Kraichnan type equation for the nonlinear case as in the linear case. The choice of the operator $\langle P(\mathbf{x},t)\rangle$ depends on the physical problem in question. However, from this approach, one is able to arrive at better limits for the convergence of the series which are manageable [such as $\|L_0(\mathbf{x},t)\times\langle\Psi(\mathbf{x},t)\rangle\| > \|K(\mathbf{x},t)\langle\Psi(\mathbf{x},t)\rangle : \langle\Psi(\mathbf{x},t)\rangle\|$] and, secondly, the equations derived are solvable and have the added feature of telescoping the normal renormalization and hierarchical approaches.

*Paper I is "The renormalized projection operator technique for linear stochastic differential equations," J. Math. Phys. 14, 340 (1973).

[1]A.S. Monin, and A.M. Yaglom, *Statistical Fluid Mechanics; Mechanics of Turbulence*, edited by J.L. Lumley (MIT, Cambridge, Mass., 1971), Vol. I.

[2]A. Yu Rozanov, *Stationary Random Processes*, translated by A. Feinstein (Holden-Day, San Francisco, 1967).

[3]Shihi Pai, *Fluid Dynamics of Jets* (Van Nostrand, New York, 1954).

[4]J.L. Lumley, and H.A. Panofsky, *The Structure of Atmospheric Turbulence* (Interscience, New York, 1964).

[5]R.H. Kraichnan, J. Math. Phys. 2, 124 (1961).

[6]V.A. LoDato, "Chemical Kinetics in a Turbulent Fluid" (to be published).

[7]V.A. LoDato, J. Math. Phys. 14, 340 (1973).

[8]V.A. LoDato, "Stochastic Processes in Heat and Mass Transport" in *Probabilistic Methods in Applied Mathematics*, edited by A.T. Bharucha-Reid (Academic, New York, 1973) Vol. III.

[9]V.I. Tatarski, Zh. Eksp. Teor. Fiz. 42, 1386 (1962).

[10]V.I. Tatarski, "Rasprostraneie voin v turbulentnoi atmosfere (Wave propagation in a turbulent atmosphere)," in *Izsatel'st vo Nauk* (Moscow, 1967).

[11]M.J. Beran, *Statistical Continuum Theories* (Wiley-Interscience, New York, 1968).

[12] G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge U. P., London, 1956).
[13] R. C. L. Bosworth, *Heat Transfer Phenomena* (Wiley, New York, 1952).
[14] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications* (McGraw-Hill, New York, 1960).
[15] U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), Vol I, pp. 75–198.
[16] R. H. Kraichnan, Phys. Fluids **1**, 358 (1958).
[17] R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[18] R. H. Kraichnan, Proc. Symp. Appl. Math.: Hydrodynamic Instability **13**, 199 (1962).
[19] R. H. Kraichnan, Phys. Fluids **7**, 1031 (1964).
[20] R. H. Kraichnan, **13**, 22 (1970).
[21] V. A. LoDato, *Simulated Environmental Physics* (Wiley-Interscience, New York) (in press).
[22] J. B. Keller, "Stochastic equations and the theory of turbulence," in *Geophysical Fluid Dynamics, Course Lectures and Seminars*, Woods Hole, Massachusetts, Woods Hole Oceanographic Institute, Massachusetts, 1966, pp. 131–136.
[23] In the result (3.16) and the definition of $L_1(\mathbf{x}, t, \omega)$ we are assuming that the class of problems have an incompressible velocity field.

A remark on the connection between stochastic mechanics and the heat equation

Sergio Albeverio

Istituto Nazionale Fisica Nucleare, Università di Napoli, Naples, Italy

Raphael Høegh-Krohn

Institute of Mathematics, University of Oslo, Blindern, Oslo, Norway

(Received 22 April 1974)

We prove that the solution of Nelson's stochastic mechanics equation associated with any stationary solution ψ of the Schrödinger equation is the homogeneous Markov process of the heat equation with Dirichlet boundary condition on the hypersurface $\psi=0$.

1. INTRODUCTION

In Ref. 1 Nelson introduced the concept of stochastic mechanics and discussed its relations with quantum mechanics. This opens the possibility of discussing quantum mechanical questions in terms of Markov processes as well as studying Markov processes by the techniques used in quantum mechanics. Nelson's stochastic mechanics has received a new momentum by the discussions of Guerra and Guerra and Ruggiero,[2] who gave its generalization to infinitely many degrees of freedom and showed its strong connection with Euclidean quantum field theory.

We shall sketch Nelson's argument for the case of a system with n degrees of freedom and a conservative force field $-\nabla V$. For simplicity we set all the masses equal to one.

Starting with the Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(x,t)=-\tfrac{1}{2}\Delta\psi(x,t)+V(x)\psi(x,t), \tag{1}$$

we write the solution in the form

$$\psi(x,t)=\rho^{1/2}(x,t)\exp[iS(x,t)] \tag{2}$$

where $\rho=|\psi|^2$. Set $v(x,t)=\nabla S(x,t)$, where ∇ is the gradient with respect to x. Then a Markov process $\xi(t)$ in R^n is completely described by taking $\rho(x,t)$ to be the distribution of $\xi(t)$ and $v(x,t)$ its current velocity, i.e.,

$$v(\xi(t),t)=\tfrac{1}{2}(D^+ + D^-)\,\xi(t), \tag{3}$$

where D^+ and D^- are the mean forward and mean backwards derivatives:

$$D^{\pm}F(\xi(t),t)=\lim_{\epsilon\to 0\pm}\epsilon^{-1}E_t[F(\xi(t+\epsilon),t+\epsilon)-F(\xi(t),t)],$$

where E_t is the conditional expectation with respect to $\xi(t)$.

If we define the displacement (drift) $\alpha(x,t)$ by

$$\alpha(x,t)=v(x,t)+\tfrac{1}{2}\nabla\ln\rho(x,t), \tag{4}$$

then it follows from the theory of stochastic differential equations that $\xi(t)$ satisfies the stochastic differential equation

$$d\xi(t)=\alpha(\xi(t),t)\,dt+dw(t), \tag{5}$$

where $w(t)$ is the standard Brownian motion in R^n given by

$$E(dw_i(t))=0,\quad E(dw_i(t)\cdot dw_j(t))=\delta_{ij}\,dt.$$

It follows now from (1) that $\xi(t)$ satisfies the Newton equation in the form that the mean acceleration is equal to the force $-\nabla V$, i.e.,

$$\tfrac{1}{2}(D^+D^- + D^-D^+)\,\xi(t)=-\nabla V(\xi(t)). \tag{6}$$

On the other hand, if we assume that we have a Markov process $\xi(t)$ which satisfies (5) and (6) for some function α of $\xi(t)$ and t, then we may define the function $\rho(x,t)$ as the distribution function for $\xi(t)$. We may assume that the displacement $\alpha(x,t)$ is a gradient, so, if we now define $v(x,t)$ by (4), v is also a gradient. This gives us a real function $S(x,t)$ by $v=\nabla S$, where S is determined up to a constant. Defining now $\psi(x,t)$ by (2) Nelson proved that, under regularity conditions on the displacement $\alpha(x,t)$, $\psi(x,t)$ satisfies the Schrödinger equation (1). In the next section we shall see that in the case where $\alpha(x,t)$ is not regular enough, then (5) and (6) have more solutions than those coming from (1). This was already noted by Nelson in Ref. 1.

Guerra and Ruggiero[2] have recently discussed the extension of Nelson's work to the case of the boson field. They make the very interesting observation that the corresponding Euclidean Markov field coincides with the lowest energy generalized stochastic process associated with classical field theory through the procedure of Nelson's stochastic mechanics. So that in this sense the underlying four-dimensional manifold on which the Markov field is defined can be considered as the physical space–time. This has lead us to the considerations in the next section, where we shall discuss the relation between Nelson's stochastic mechanics and the heat equation for a system with n degrees of freedom.

2. RELATIONS BETWEEN STOCHASTIC MECHANICS AND THE HEAT EQUATION

The connection between the Schrödinger equation (1) and the corresponding heat equation

$$-\frac{\partial}{\partial t}\psi(x,t)=-\tfrac{1}{2}\Delta\psi(x,t)+V(x)\psi(x,t) \tag{7}$$

in the sense that the solutions of (7) are analytic in Re$t>0$ and continuous for Re$t\geq 0$ and their values on the imaginary axis are solutions of (1) has long been known and utilized, and it is this relation that forms the basis for Euclidean field theory. In this correspondence the parameter t in (7) has the interpretation of an imaginary time. However, by the observation made by Guerra and Ruggiero in Ref. 2 we are lead to the interpretation of the parameter t in (7) as the real physical time for the

Copyright © 1974 American Institute of Physics

stationary process in stochastic mechanics corresponding to the ground state for the Schrödinger equation (1). Hence by Nelson's equivalence between stochastic mechanics and quantum mechanics we are lead to the interpretation of the parameter t in (7) as the real physical time of quantum mechanics, at least for the ground state. We shall see below that such an interpretation also holds, with some modifications, for any stationary solution of (1).

Let us now assume that the potential $V(x)$ is such that (1) admits a stationary solution, i.e., a solution of the form $\psi(x,t)=\exp(-i\lambda t)\,\varphi(x)$ with $\varphi(x)\in L_2(R^n)$, so that

$$(-\tfrac{1}{2}\Delta + V - \lambda)\varphi = 0. \tag{8}$$

Since this is a real equation $\varphi(x)$ can always be chosen as a real function. In this case we have in the notations of the previous section that $\rho=\varphi^2(x)$ and $S(x,t)=-\lambda t$. We remark that we have permitted the square root in (2) to be the positive or negative square root, depending on whether $\varphi(x)$ is positive or negative, i.e., $\rho^{1/2}(x)=\varphi(x)$. This is necessary in order to make $S(x,t)$ a continuous function of x. We now get that $v(x,t)=\nabla S=0$. Hence the current velocity (3) for the Markov process $\xi(t)$ is zero. The corresponding displacement (drift) is therefore given by $\alpha(x,t)=\frac{1}{2}\nabla\ln\rho$. Let us assume that $V(x)$ is a smooth function so that by (8) $\varphi(x)$ is also a smooth function. In this case $\alpha(x,t)$ is a smooth function outside the set $\varphi(x)=0$, where $\alpha(x,t)$ is given by

$$\alpha(x)=[1/\varphi(x)]\,\nabla\varphi(x). \tag{9}$$

Hence, for a stationary solution of the Schrödinger equation, the current velocity satisfies $v(x,t)=0$ and the displacement $\alpha(x)$ is equal to the osmotic velocity $\frac{1}{2}\nabla\ln\rho$, which is time independent and given in terms of the eigenfunction $\varphi(x)$ by (9). If $\varphi(x)>0$ for all x, then $\alpha(x)$ will satisfy regularity conditions sufficient to secure that the stochastic differential equation

$$d\xi(t)=\alpha(\xi(t))\,dt + dw(t) \tag{10}$$

has a solution and this solution is unique. See Refs. 3 and 1. So in this case we have a unique Markov process $\xi(t)$ which is homogeneous in time with $\rho(x)$ as the distribution for $\xi(t)$ or, if we want, $\rho(x)\,dx$ as the invariant measure for the homogeneous process $\xi(t)$. The situation when $\varphi(x)$ has zeros is more complex since this leads, by (9), to singularities for the displacement $\alpha(x)$, and hence the above mentioned existence and uniqueness theorem does not apply. We shall, however, show that even in this case there exist solutions of (10). The solution we construct is homogeneous with invariant distribution $\rho(x)$, which is related to $\alpha(x)$ by $\alpha(x)=\frac{1}{2}\nabla\ln\rho(x)$.

In fact let us look for solutions of

$$d\xi(t)=\alpha(\xi(t))\,dt + dw(t), \tag{11}$$

where $\xi(t)$ is required to be a homogeneous Markov process with given invariant distribution $\rho(x)$ related to the displacement $\alpha(x)$ by

$$\alpha(x)=\tfrac{1}{2}\nabla\ln\rho(x). \tag{12}$$

Since (12) implies that the displacement $\alpha(x)$ is equal to the osmotic velocity, we have that the current velocity $v(\xi(t),t)=\frac{1}{2}(D^+ + D^-)\,\xi(t)$ must be equal to zero because the displacement is $\alpha(\xi(t),t)=D^+\xi(t)$ and the osmotic velocity is $\frac{1}{2}\nabla\ln\rho(x,t)=\frac{1}{2}(D^+ - D^-)\,\xi(t)$. Let $\xi_*(t)$ be the time reversed process. We then have that $D^-\xi_*(t) = -D^+\xi(t)$ and by the fact that the current velocity of $\xi(t)$ is zero, we get

$$D^+\xi_*(t)=D^+\xi(t), \tag{13}$$

which then gives that also the reversed process satisfies (11). It is therefore natural to seek solutions of (11) which satisfy the condition that $\xi_*=\xi$ (in law).

Since $\xi(t)$ has an invariant distribution $\rho(x)$ we may define a semigroup P_t in $L_2(\rho\,dx)$ by $(P_t f)(x)=E_0[f(\xi(t))]$ for any $f\in L_2(\rho\,dx)$, where E_0 is the conditional expectation with respect to $\xi(0)$. The condition $\xi_*=\xi$ implies that $P_t^*=P_t$, where P_t^* is the adjoint semigroup. It follows from the proof of Theorem 3, II, Chap. 2, Sec. 9 of Ref. 3 that if $\xi(t)$ is a solution of (11) and f is a smooth function which is zero in a neighborhood of the singularities of $\alpha(x)$, then the strong limit of $-(1/t)(P_t f - f)$ exists as $t\to 0$ and is given by

$$Af=-(1/2\rho)\nabla\cdot(\rho\nabla f). \tag{14}$$

Hence we get that the infinitesimal generator A of P_t is a closed extension of the operator $-(1/2\rho)\nabla\cdot\rho\nabla$ defined on smooth functions which are zero near the zeros of ρ. The condition $\xi_*=\xi$ is equivalent to the condition $A=A^*$. Hence we know that if there is a solution of (11) with invariant distribution ρ and satisfying the condition $\xi=\xi_*$, then the semigroup P_t generated by the process is of the form $P_t=\exp(-tA)$, where A is a self-adjoint extension of $-(1/2\rho)\nabla\cdot\rho\nabla$ defined on smooth functions which are zero near the zeros of ρ. There is an obvious restriction for A, namely that we should have $\|P_t\|\leqslant 1$, E_0 being a conditional expectation, so that $A\geqslant 0$.

We may now identify $L_2(\rho\,dx)$ with $L_2(R^n)$ by the identification $f \longleftrightarrow \varphi f$ since $\varphi^2=\rho$. Then the operator $-(1/2\rho)\nabla\cdot\rho\nabla$ is identified with

$$-\tfrac{1}{2}\Delta + (V-\lambda) \tag{15}$$

since φ satisfies the equation $-\frac{1}{2}\Delta\varphi + (V-\lambda)\varphi=0$. Hence, in the $L_2(R^n)$ representation, $P_t=\exp(-tA)$, where A is a positive self-adjoint extension of (15) defined on smooth functions which are zero near the zeros of $\varphi(x)$.

In the case λ is the lowest eigenvalue λ_0 we know that $\varphi(x)$ is always different from zero. Hence (15) is essentially self-adjoint and for this special case we have then simply that $f(t,x)=(P_t f_0)(x)$ is the solution of the heat equation

$$-\frac{\partial}{\partial t}f=(-\tfrac{1}{2}\Delta + (V-\lambda_0))f, \tag{16}$$

with initial condition $f(0,x)=f_0(x)$. Hence in this case, which is also the case in which we have existence and uniqueness for (11), we get that the Markov process $\xi(t)$ of the stochastic mechanics is identical with the heat equation process described by (16).

In the case where λ is not the lowest eigenvalue, we have that $\varphi(x)$ has zeros. In this case we take A to be the Friedrichs extension, i.e., the minimal extension that conserves positivity of $-\frac{1}{2}\Delta + (V-\lambda)$ defined on smooth functions which are zero near the zeros of $\varphi(x)$. This is a self-adjoint operator $A_\varphi\geqslant 0$, and it is well known that it is $-\frac{1}{2}\Delta_\varphi+(V-\lambda)$, where Δ_φ is the Laplacian with Dirichlet boundary conditions on the hypersurface $\varphi(x)=0$. It follows then immediately that $\varphi(x)$ is in the domain of A_φ and that $A_\varphi\varphi(x)=0$ so that $\rho(x)$ is an invariant measure for the process $\xi(t)$ generated by the probability semigroup $P_t^\varphi=\exp(-tA_\varphi)$. Since the eigenfunction belonging to the lowest eigenvalue λ_0 for the eigenvalue problem (8) can be taken to be positive everywhere, we have that an eigenfunction φ belonging to any eigenvalue $\lambda>\lambda_0$ must take both positive and negative values. Hence the hypersurface $\varphi(x)=0$ divides the space R^n into at least two disjoint domains. Let Λ_α, $\alpha=1,\ldots,l$ be the domains into which R^n is divided by the hypersurface $\varphi(x)=0$. Since then Δ_φ is the direct sum of Δ_{Λ_α} operating in $L_2(\Lambda_\alpha)$, where Δ_{Λ_α} is the Laplacian with Dirichlet boundary conditions on the boundary of the domain Λ_α, we get that the process $\xi(t)$ given by P_t^φ does not have a unique invariant measure. We have namely that φ decomposes in a direct sum $\varphi=\sum_\alpha\varphi_\alpha$ with $\varphi_\alpha=\chi_{\Lambda_\alpha}\varphi$ such that $P_t^\varphi\varphi_\alpha=\varphi_\alpha$. That is to say, all the distributions $\rho_\alpha=\varphi_\alpha^2$ are invariant distributions for the process. In fact we see that the process $\xi(t)$ given by P_t^φ never crosses the hyperplane $\varphi(x)=0$, and if we start it in any of the connected domains Λ_α it will always remain there. However, if we restrict it to any of the domains Λ_α it has a unique invariant measure ρ_α. If we define $f(x,t)=(P_t^\varphi f_0)(x)$ we have

$$\frac{\partial}{\partial t}f=(-\tfrac{1}{2}\Delta_\varphi+(V-\lambda))f. \tag{17}$$

So also in the case where φ is an eigenfunction not belonging to the lowest eigenvalue λ_0 we get that the Markoff process $\xi_\varphi(t)$ in the stochastic mechanics is identical with a heat equation process, namely the one described by (17). We summarize the above results in the following theorem.

Theorem 1: Let $\psi(x,t)=\exp(-it\lambda)\varphi(x)$ be a stationary solution of the Schrödinger equation (1), with a smooth potential $V(x)$ which permits stationary solutions. Then the corresponding stochastic mechanics equation has a solution $\xi_\varphi(t)$ which is a homogeneous Markoff process that is invariant under time reversal and has $\rho(x)=\varphi(x)^2$ as invariant distribution. Moreover, the paths of $\xi_\varphi(t)$ are continuous and do not cross the hypersurface $\varphi(x)=0$, and the semigroup generated by $\xi_\varphi(t)$ is a heat equation semigroup with infinitesimal generator given by

$$-\tfrac{1}{2}\Delta_\varphi+(V-\lambda),$$

where Δ_φ is the Laplacian with Dirichlet boundary conditions on the hypersurface $\varphi(x)=0$.

Remark: It follows from this theorem that in the stationary case for higher eigenvalues the stochastic mechanics equation has several solutions, namely those obtained by starting the process in one or some of the domains given by the hypersurface $\varphi(x)=0$. One may ask the question whether the solution is unique in each of the connected domains Λ_α given by the hypersurface $\varphi(x)=0$. We are able to prove this only in the one-dimensional case. The proof goes by explicit examination of all self-adjoint extensions of (15) defined on all smooth functions which are zero near the end points of the interval Λ_α determined by two consecutive zeros of $\varphi(x)$. By using the fact that all such extensions are given by self-adjoint boundary conditions, it suffices then to show by a simple calculation that the Dirichlet boundary condition is the only one which has $\varphi(x)$ as an eigenfunction, recalling that V was assumed to be smooth.

ACKNOWLEDGMENTS

We are very indebted to Professor F. Guerra for illuminating discussions. The first named author would like to thank both the Istituto di Fisica Teorica, I. N. F. N., Università di Napoli, and the Matematisk Institutt, Universitetet i Oslo, for their hospitality.

[1]E. Nelson, Phys. Rev. 150, 1079 (1966); E. Nelson, *Dynamical Theories of Brownian Motion*, *Mathematical Notes* (Princeton U. P., Princeton, N.J., 1967).
[2]F. Guerra and P. Ruggiero, Phys. Rev. Lett. 31, 1022 (1973); F. Guerra, "On the connection between Euclidean-Markov field theory and stochastic quantization," Lecture given at Varenna Summer School, (1973); F. Guerra, "On stochastic field theory," Report at the Aix-en-Provence 2nd International Conference on Elementary Particles (1973).
[3]I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations* (Springer-Verlag, Berlin, 1972), Chap. I, Sec. 3.

Separation of tensor equations in a homogeneous space by group theoretical methods*

B. L. Hu

Institute of Theoretical Physics, Department of Physics, Stanford University, Stanford, California 94305
(Received 18 March 1974)

Details of the group-theoretical method for the separation of tensor equations in a homogeneous space are given. As illustrations, the vector and tensor harmonics in a $SO(3)$-homogeneous space are constructed, with applications to the study of electromagnetic and gravitational perturbations in the mixmaster universe.

I. INTRODUCTION

In studying problems concerning electromagnetic and gravitational radiation in a curved space, one usually has to deal with vector and tensor wave equations of some form. The general procedure for solving these equations involves three main steps: First, one chooses a coordinate system and reduces the covariant derivatives in the tensor equations to ordinary derivatives, thus introducing the Christoffel symbols and their derivatives with dependence on explicit space—time variables. Second, one tries to construct complete sets of orthonormal tensor harmonics and expands the tensor field on the manifold in terms of these. When substituted into the tensor wave equations, the tensor Fourier coefficients (the amplitude functions) obey a set of partial differential equations. Lastly, one has to ponder about the separability of the wave equations into ordinary differential equations for each spatial variable. If the space possesses a certain degree of symmetry and if one can incorporate that into the tensor harmonics, then the equations will probably be separable. But the construction of the tensor harmonics with the inclusion of arbitrary symmetry does not usually follow definite rules, and sometimes that is just as difficult as the attempt to achieve separability of the wave equations.

For spaces possessing a high degree of symmetry, the methods for the separation of tensor equations have been well thought out. Earlier works of Regge and Wheeler[1] on the perturbations in the Schwarzschild metric [of $SO(3)$ symmetry on the 2-sphere S^2] and of Lifshitz and Khalatnikov[2] on the Friedmann universe [of $SO(4)$ symmetry on the 3-sphere S^3] suggested one way of constructing the tensor harmonics. They are derivable by the action of invariant operators (e.g., the $\hat{r}$, $\hat{L}$ and $\hat{\nabla}$ operators in Ref. 1) on the scalar harmonics of the space. An alternative way, as exemplified by the work of Mathews[3] and Hu,[4] constructs tensor harmonics by coupling scalar functions with basis tensors belonging to definite representations of the symmetry group. The equivalence of these methods is only natural and was proved in the case of the Schwarzschild metric by Zerilli.[5] However, for spaces with lower symmetry, especially when the space manifold and the group manifold do not coincide (unlike the Schwarzschild and Friedmann spaces), the above methods become hard to apply because the invariant operators of the group do not comply with the symmetry of the space and the tensor basis is not derivable from simple representations.

In this paper, we shall focus our attention on the general class of *homogeneous spaces*,[6] which are generated by groups of motion that preserve the metric forms. For a homogeneous space possessing certain degrees of symmetry as characterized by the underlying group, one expects to find group theoretical methods to be of fundamental importance to the solution of the above problems. Can the basis harmonics be derived by some simple group operation, and in the solution of tensor wave equations, can one derive the equations governing the amplitude functions of each normal mode with the proper symmetry accounted for without going through the expansion of tensor harmonics and the separation of variables?

This was the essence of the method proposed recently by Hu and Regge.[7] It makes use of group symmetry properties of homogeneous spaces to construct the tensor harmonics and separate the tensor equations. By this method, one can derive the field equations governing the amplitude functions directly, without knowing the explicit forms of the harmonics. As an illustration the method was applied to separate the perturbation equations in a closed, anisotropic universe (type IX, the mixmaster universe). In this paper we shall elaborate on the method and supply more applications. In Sec. II the formalism is presented in detail and the useful geometric quantities in an $SO(3)$-homogeneous space are presented. In Sec. III we construct the vector and tensor harmonics from a solution of the wave equations. In Sec. IV the method is applied to the study of electromagnetic and gravitational waves in anisotropic homogeneous universes. The method presented here can easily be extended to study the separation of wave equations and the construction of tensor harmonics in other types of homogeneous spaces.

II. FORMALISM—GROUP THEORETICAL METHOD

In seeking a solution to a tensor wave equation, one usually has to first expand the covariant derivatives into ordinary derivatives and then look for an expansion of the tensor field $h_{\mu\nu}$ into basis harmonics and finally ponder about the separability of the wave equation. This procedure is rather involved and sometimes even inhibitive. However, for a homogeneous space, since every point is equivalent to every other point by a group translation, one can *choose* to perform all computations at one specific point in space. The general form of the equations are generated by simple group invariant operations on the manifold. The advantage over the traditional approach described above are many. Firstly, as the tensor equations are resolved at one point, the question of separation of variables does not arise. The time dependent differential equations for the amplitude functions obtained at one point are just as general as at any other

Copyright © 1974 American Institute of Physics

point in space. Secondly, in a homogeneous space, the set of tensor harmonics are composed of direct products of the basis invariant forms operating on the representation function of the group. In this case, one does not have to construct basis tensor harmonics as functions of the whole space [like the spherical tensor harmonics $Y_{lm}(\theta, \phi)$ in the Schwarzschild metric[1,3,5] and the hyperspherical tensor harmonics $Y^{\eta}_{lm}(\chi, \theta, \phi)$ in the Robertson–Walker metric[2,4]), but, rather, one can evaluate the product at one point and use the invariant operators to generate the complete set. Any tensor field in a homogeneous space can be expanded in terms of these tensor harmonics. In fact, to derive differential equations for the amplitude functions, one does not even have to know the tensor harmonics explicitly. All that enters are the transformation properties of the amplitude functions, which are carried by the tensor harmonics; and the action of tensor harmonics under invariant operators is simply derivable from the basic group structure. This is where the merit of the group theoretical formalism resides. Lastly, in the process of reducing the tensor wave equation with covariant derivatives to ordinary derivatives, one can avoid the complications in calculating the Christoffel symbols as spatial functions by suitably choosing a convenient point in a simple coordinate system and carry out all calculations there. Since all geometric quantities involve no derivatives of the metric tensor higher than the second order, one can expand the metric tensor or any tensor quantities only up to the second order in the coordinate variables—provided that a point like the origin in the Euclidean coordinate is chosen. These considerations yield tremendous simplifications in the computations. In the following, we shall take the $SO(3)$-homogeneous space as a model and illustrate the above method of approach. In the first part, we calculate the Christoffel symbols and their derivatives by performing a power series expansion of the general metric tensor. In the second part, we explain the action of the invariant operators. The tensor harmonics will be constructed in Sec. III following this method.

A. Expansion of the metric in E^4 coordinates

The metric of a homogeneous space is given by

$$dl^2 = \bar{\gamma}_{ab}\sigma^a\sigma^b \quad (a, b = 1, 2, 3), \tag{1}$$

where $\bar{\gamma}_{ab}$ is a constant symmetric tensor and the σ^a are the invariant basis differential forms of the space. They obey the relations

$$d\sigma^a = \tfrac{1}{2} C^a_{bc}\sigma^b \wedge \sigma^c, \tag{2}$$

where C^a_{bc} is the structure constant of the underlying symmetry group. For $SO(3)$-homogeneous space (Bianchi Type IX)[8] the C^a_{bc} is equal to ϵ_{abc}, the total antisymmetric tensor. The basis forms σ^a are expressible in terms of coordinate differentials. One example is the Euler angle parametrization (θ, ϕ, ψ) [8]

$$\begin{aligned}
\sigma^1 &= -\sin\psi\, d\theta + \cos\psi \sin\theta\, d\phi,\\
\sigma^2 &= \cos\psi\, d\theta + \sin\psi \sin\theta\, d\phi,\\
\sigma^3 &= d\psi + \cos\theta\, d\phi.
\end{aligned} \tag{3}$$

For simplicity of computation, the Cartesian coordinates x_i $(i = 1-4)$ are preferred here. The invariant basis forms σ^a $(a = 1-4)$ on the 3-sphere S^3 are given in terms of the coordinate differentials dx^i of the Euclidean space E^4 by[9]

$$\begin{aligned}
\sigma^1 &= 2(-x_4dx_1 - x_3dx_2 + x_2dx_3 + x_1dx_4),\\
\sigma^2 &= 2(x_3dx_1 - x_4dx_2 - x_1dx_3 + x_2dx_4),\\
\sigma^3 &= 2(-x_2dx_1 + x_1dx_2 - x_4dx_3 + x_3dx_4),\\
\sigma^4 &= 2(x_1dx_1 + x_2dx_2 + x_3dx_3 + x_4dx_4).
\end{aligned} \tag{4}$$

By introducing the transformation matrices S and Ω_a, the above relations can be reexpressed as

$$\sigma^a = 2S_{ai}(x)\, dx^i = 2\Omega_{aji}x^j dx^i. \tag{5}$$

(Throughout this paper, summation is extended over repeated indices unless otherwise stated.) From the orthogonality conditions of S, i.e.,

$$S_{aj}S_{bj} = \delta_{ab}, \quad S_{ak}S_{aj} = \delta_{kj}, \tag{6a}$$

one can easily deduce the following relations for Ω_a

$$\Omega_{aij}\Omega_{aik} = \delta_{jk}, \quad \Omega_{aij}\Omega_{alj} = \delta_{il}. \tag{6b}$$

(In the above summations only, the indices a, b, j, k run from 1 to 4.) Furthermore, from the explicit form of σ^a it is clear that Ω_{aji} is antisymmetric with respect to the interchange of i and j for $a = 1, 2, 3$.

The coordinate differentials of E^4 are expressible in terms of σ^a by

$$dx^i = \tfrac{1}{2}S_{ai}\sigma^a. \tag{7a}$$

By introducing the invariant vectors $\mathbf{e}_b$ on S^3 dual to the basis forms σ^a, $\sigma^a(\mathbf{e}_b) = \delta^a_b$, and obeying commutation relations

$$[\mathbf{e}_a, \mathbf{e}_b] = -\epsilon_{abc}\mathbf{e}_c \tag{8}$$

the coordinate derivatives (vector fields) of E^4 are given by

$$\frac{\partial}{\partial x_i} = 2S_{ai}\mathbf{e}_a. \tag{7b}$$

When expressed in terms of coordinate differentials dx^i, the spatial metric can be written as

$$dl^2 = g_{ij}(x)dx^i dx^j,$$

where

$$g_{ij}(x) = \gamma_{ab}S_{ai}(x)S_{bj}(x), \quad \gamma_{ab} \equiv 4\bar{\gamma}_{ab}. \tag{9}$$

We now proceed to find an explicit expression for the metric tensor $g_{ij}(x)$ in terms of the spatial variables. The calculation can be greatly simplified if we take into consideration the observations mentioned above , i.e.:

(1) Since the space is homogeneous, one can *choose* to evaluate all geometric field quantities at any arbitrary point in space. In E^4 coordinates with restrictions on the three sphere, a convenient point that renders greatest simplicity is the pole $(x_1 = x_2 = x_3 = 0,\ x_4 = 1)$.

(2) Since the curvature tensors are related to the second derivatives of the metric tensor, at the pole it would suffice to retain terms up to the second order in x_i in the expansion of $g_{ij}(x)$.

Hence, from (4) and (9), writing $x_4^2 = 1 - \sum_{i=1}^3 x_i^2$, dx_4

$= -\sum_{i=1}^{3} x_i dx_i$, also setting $x_4 = 1$, the square of the basic forms are seen to be given by

$$\begin{aligned}\tfrac{1}{4}(\sigma^1)^2 &= (1 + x_1^2 - x_2^2 - x_3^2)dx_1^2 + x_3^2 dx_2^2 + x_2^2 dx_3^2 - 2x_2x_3dx_2dx_3 \\ &\quad + 2(x_3 + x_1x_2)dx_1dx_2 + 2(-x_2 + x_1x_3)dx_1dx_3, \\ \tfrac{1}{4}(\sigma^1\sigma^2) &= (x_1x_2 - x_3)dx_1^2 + (x_3 + x_1x_2)dx_2^2 - x_1x_2dx_3^2 \\ &\quad + (1 - 2x_3^2)dx_1\,dx_2 + (x_1 + 2x_2x_3)dx_1dx_3 \\ &\quad + (-x_2 + 2x_1x_3)dx_2dx_3.\end{aligned}$$

From this, it is easy to relate the metric coefficients $g_{ij}(x)$ to γ_{ab}:

$$\begin{aligned}g_{11} &= \gamma_{11}(1 + x_1^2 - x_2^2 - x_3^2) + 2\gamma_{12}(-x_3 + x_1x_2) + 2\gamma_{13}(x_2 + x_1x_3) \\ &\quad + \gamma_{22}(x_3^2) - 2\gamma_{23}(x_2x_3) + \gamma_{33}(x_2^2), \qquad (10) \\ g_{12} &= \gamma_{11}(x_3 + x_1x_2) + \gamma_{22}(-x_3 + x_1x_2) - \gamma_{33}x_1x_2 + \gamma_{12}(1 - 2x_3^2) \\ &\quad + \gamma_{23}(x_2 + x_1x_3) + \gamma_{13}(-x_1 + 2x_2x_3).\end{aligned}$$

Other components are obtained from the above expressions by cyclically permutating the indices 1, 2, 3. After close observation of these formulas, we deduce the following algebraic expression for the general metric in $SO(3)$-homogeneous spaces (Binachi Type IX) in E^4 coordinates on the 3-sphere (expanded here to quadratic order in x_i):

$$\begin{aligned}g_{ij}(x) &= \gamma_{ij} + \epsilon_{kjl}\gamma_{ik}x_l + \epsilon_{kil}\gamma_{kj}x_l + (\gamma_{mm})x_ix_j \\ &\quad + 2\gamma_{kl}\epsilon_{kim}\epsilon_{ljn}x_mx_n + [(\gamma_{mn}x_mx_n) - (\gamma_{mm}x_nx_n)]\delta_{ij}.\end{aligned} \qquad (11)$$

For the contravariant metric g^{ij}, an expansion to the first order in x_i will suffice for our purpose. This is due to the fact that, in calculating the first derivatives of the Christoffel symbols, we need to know the expansion up to the first order in x_i only. Thus,

$$\begin{aligned}g^{11} &= \gamma^{11} + 2(\gamma^{13}x_2 - \gamma^{12}x_3), \\ g^{12} &= \gamma^{12} - \gamma^{13}x_1 + \gamma^{23}x_2 + (\gamma^{11} - \gamma^{22})x_3,\end{aligned} \qquad (12)$$

or, in closed form, $g^{ij} = \gamma^{ij} + \epsilon_{kjl}\gamma^{ik}x_l + \epsilon_{kil}\gamma^{kj}x_l$.

We give the explicit expressions for $\Gamma_{i,kl}$ for $i = 1$ below; the other components are obtainable by cyclically permutating the three indices (i, k, l):

$$\begin{aligned}\Gamma_{1,11} &= \gamma_{11}x_1 + \gamma_{12}x_2 + \gamma_{13}x_3, \\ \Gamma_{1,12} &= \gamma_{13} + \gamma_{12}x_1 + (\gamma_{33} - \gamma_{11})x_2 - \gamma_{23}x_3, \\ \Gamma_{1,13} &= -\gamma_{12} + \gamma_{13}x_1 - \gamma_{23}x_2 + (\gamma_{22} - \gamma_{11})x_3, \\ \Gamma_{1,22} &= 2\gamma_{23} + (\gamma_{11} + 2\gamma_{22} - 2\gamma_{33})x_1 - \gamma_{12}x_2 + 3\gamma_{13}x_3, \\ \Gamma_{1,23} &= (\gamma_{33} - \gamma_{22}) + 4\gamma_{23}x_1 - 2\gamma_{13}x_2 - 2\gamma_{12}x_3, \\ \Gamma_{1,33} &= -2\gamma_{23} + (\gamma_{11} - 2\gamma_{22} + 2\gamma_{33})x_1 + 3\gamma_{12}x_2 - \gamma_{13}x_3.\end{aligned} \qquad (13)$$

With these formulas in hand we can easily calculate the Christoffel symbols and their derivatives evaluated at the pole. All the nonzero components are given as follows (here, for completeness, we allow the metric coefficients γ_{ab} to be time-dependent, a dot denoting derivative with respect to t):

$$\Gamma^0_{ij} = \tfrac{1}{2}\dot{\gamma}_{ij}, \quad \frac{\partial \Gamma^0_{ij}}{\partial x_l} = \tfrac{1}{2}(\epsilon_{kjl}\dot{\gamma}_{ik} + \epsilon_{kil}\dot{\gamma}_{kj}),$$

$$\Gamma^i_{0j} = \tfrac{1}{2}\gamma^{in}\dot{\gamma}_{nj}, \quad \frac{\partial \Gamma^i_{0j}}{\partial x_l} = \tfrac{1}{2}(\epsilon_{kjl}\gamma^{in}\dot{\gamma}_{nk} + \epsilon_{kil}\gamma^{kn}\dot{\gamma}_{nj}), \qquad (14)$$

$$\Gamma^m_{ij} = -\gamma^{ml}(\epsilon_{kjl}\gamma_{ik} + \epsilon_{kil}\gamma_{kj}),$$

$$\begin{aligned}\frac{\partial \Gamma^i_{jk}}{\partial x_l} &= 1 \quad \text{for } i = j = k = l \\ &= \gamma^{ii}(\gamma_{pp} - \gamma_{kk}) + \gamma^{pp}(\gamma_{kk} - \gamma_{ii}) + (\gamma^{ik}\gamma_{ik} - \gamma^{kp}\gamma_{kp}) \\ &\qquad \text{for } i = j \neq k = l\ (p \neq i \neq k) \\ &= \gamma^{ii}(\gamma_{ii} + 2\gamma_{jj} - 2\gamma_{pp}) + 3\gamma^{ip}\gamma_{ip} - \gamma^{ij}\gamma_{ij} \\ &\qquad \text{for } j = k \neq i = l \\ &= 2[\gamma^{il}(\gamma_{ii} - \gamma_{pp}) + \gamma_{il}(\gamma^{pp} - \gamma^{ii}) + (\gamma^{ip}\gamma_{lp} - \gamma^{lp}\gamma_{ip})] \\ &\qquad \text{for } i = j = k \neq l \\ &= \gamma^{ik}(\gamma_{pp} - \gamma_{ii}) + \gamma^{ii}\gamma_{ik} - \gamma^{ip}\gamma_{kp} \quad \text{for } i = j = l \neq k \qquad (15) \\ &= \gamma^{kl}(\gamma_{ii} - \gamma_{kk}) + (\gamma^{kk} - 2\gamma^{ii})\gamma_{kl} - 3\gamma^{ik}\gamma_{il} + 4\gamma_{ik}\gamma^{il} \\ &\qquad \text{for } i = j \neq k \neq l \\ &= 2(\gamma^{ip}\gamma_{jp} - \gamma_{ij}\gamma^{pp}) \quad \text{for } i \neq j = k = l \\ &= \gamma^{il}(\gamma_{ll} - 2\gamma_{ii} + 2\gamma_{jj}) + 3\gamma^{ii}\gamma_{il} - 3\gamma^{ij}\gamma_{jl} + 2\gamma^{jl}\gamma_{ij} \\ &\qquad \text{for } i \neq j = k \neq l \\ &= 4\gamma^{ii}\gamma_{jk} - 2\gamma^{ij}\gamma_{ik} - 2\gamma^{ik}\gamma_{ij} \quad \text{for } i = l \neq j \neq k \\ &= \gamma^{ik}(\gamma_{ii} + \gamma_{kk} - 2\gamma_{jj}) - \gamma_{ik}(\gamma^{ii} + \gamma^{kk}) + \gamma^{ij}\gamma_{jk} + \gamma^{jk}\gamma_{ij} \\ &\qquad \text{for } i \neq j = l \neq k.\end{aligned}$$

In Eq. (15), no summation is assumed over repeated indices, and $p \neq i \neq k$. For the diagonal metric, $\gamma_{ij} = l_i^2\delta_{ij}$, the above formulas simplify a great deal. In particular, only three subcases for $\partial\Gamma^i_{jk}/\partial x^l$ remain nonzero. Defining $\gamma_i \equiv l_i^2$ and $\kappa_i \equiv \dot{l}_i/l_i$, we reduce Eqs. (14) and (15) for the diagonal case as follows:

$$\begin{aligned}&\Gamma^0_{ij} = \kappa_i\gamma_i\delta_{ij}, \quad \frac{\partial \Gamma^0_{ij}}{\partial x_k} = \epsilon_{ijk}(\kappa_i\gamma_i - \kappa_j\gamma_j), \\ &\Gamma^i_{0j} = \kappa_i\delta_{ij}, \quad \frac{\partial \Gamma^i_{0j}}{\partial x_k} = \epsilon_{ijk}(\kappa_i - \kappa_j), \\ &\Gamma^i_{jk} = \epsilon_{ijk}\left(\frac{\gamma_k - \gamma_j}{\gamma_i}\right), \\ &\frac{\partial \Gamma^i_{jk}}{\partial x_l} = 1 \quad \text{for } i = j = k = l \\ &\qquad = \frac{(\gamma_p - \gamma_i)(\gamma_p + \gamma_i - \gamma_k)}{\gamma_p\gamma_i} \quad \text{for } i = j \neq k = l\ (p \neq i \neq k) \\ &\qquad = \frac{(\gamma_i + 2\gamma_j - 2\gamma_p)}{\gamma_i} \quad \text{for } j = k \neq i = l.\end{aligned} \qquad (16)$$

B. Invariant operators and representation functions

Vector and tensor harmonics are generated by the action of invariant operators of the space on scalar harmonics, which are representation functions of the underlying symmetry group. For the $SO(3)$-homogeneous space, the invariant vectors obey the commutation relations (8). In terms of the Euler angle variables (3) they are given by[8,10]

$$\mathbf{e}_1 = -\sin\psi\frac{\partial}{\partial\theta} + \frac{\cos\psi}{\sin\theta}\frac{\partial}{\partial\phi} - \cot\theta\cos\psi\frac{\partial}{\partial\psi},$$

$$\mathbf{e}_2 = \cos\psi \frac{\partial}{\partial\theta} + \frac{\sin\psi}{\sin\theta}\frac{\partial}{\partial\phi} - \cot\theta\sin\psi\frac{\partial}{\partial\psi}, \tag{17}$$
$$\mathbf{e}_3 = \frac{\partial}{\partial\psi}.$$

The invariant vectors $\mathbf{e}_i$ are simply related to the angular momentum operators of the three-dimensional rotation group in quantum mechanics by

$$\hat{L}_i = i\mathbf{e}_i \quad (i=1,2,3),$$

where $\hat{L}_i$ are the intrinsic angular momentum operators of a rigid body. The Killing vectors ξ_a that generate the symmetry transformation of the space satisfy the commutation relations:

$$[\xi_b, \xi_c] = C^a_{bc}\,\xi_a \tag{18}$$

and are related to the spatial angular momentum operators by $\hat{L}_{x_i} = -i\xi_i$ $(x_i = x, y, z)$. The Casimir operator $C = \sum_{a=1}^3 \mathbf{e}_a^2$ is an invariant of the group just as the total angular momentum operator $\hat{L}^2 = \hat{L}_1^2 + \hat{L}_2^2 + \hat{L}_3^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2$ is a constant of motion in the quantum mechanics of rigid rotators.

The representation function f of the group satisfies the differential equation $\hat{L}^2 f = \lambda^2 f$, which in Euler angle variables is given by

$$\left(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\phi^2} - 2\cos\theta\frac{\partial^2}{\partial\phi\partial\psi} + \frac{\partial^2}{\partial\psi^2}\right) - \lambda^2\right] f = 0. \tag{19}$$

The solution is the well-known Wigner function[11]

$$f \equiv \mathcal{D}^J_{KM}(\theta,\phi,\psi) = \exp(iM\phi)\, d^J_{KM}(\theta)\exp(iK\psi) \tag{20}$$

with eigenvalues $\lambda^2 = J(J+1)$. They are simultaneous eigenfunctions of $\hat{L}^2$, $\hat{L}_3$, $\hat{L}_z$ characterized by the quantum numbers J, K, M respectively:

$$\begin{aligned} \hat{L}^2\mathcal{D}^J_{KM} &= J(J+1)\mathcal{D}^J_{KM},\\ \hat{L}_3\mathcal{D}^J_{KM} &= K\mathcal{D}^J_{KM},\\ \hat{L}_z\mathcal{D}^J_{KM} &= M\mathcal{D}^J_{KM}. \end{aligned} \tag{21}$$

In all spaces that have $SO(3)$ symmetry, any scalar harmonic function can be constructed from linear combinations of the Wigner function. The general function possesses definite (J,M) states, but the K states will be mixed. More discussions of the scalar harmonics can be found in Ref. 10.

In terms of the E^4 coordinates, the invariant operators are related to the spatial derivatives by (7b). At the pole $(x_4 = 1,\ x_1 = x_2 = x_3 = 0)$, $S_{ai} = -\delta_{ai}$ and hence the differential operators are given by

$$\left.\frac{\partial}{\partial x_i}\right|_0 = -2\mathbf{e}_i = 2i\hat{L}_i. \tag{22a}$$

Repeated operations of $\hat{L}_i$ yield expressions for the second differential operators

$$\left.\frac{\partial^2}{\partial x_i\partial x_j}\right|_0 = \left.\frac{\partial^2}{\partial x_j\partial x_i}\right|_0 = -2(\hat{L}_i\hat{L}_j + \hat{L}_j\hat{L}_i). \tag{22b}$$

From the simple relations

$$\begin{aligned} &\hat{L}_+\mathcal{D}_K = i\epsilon_K\mathcal{D}_{K-1}, \quad \hat{L}_-\mathcal{D}_K = -i\epsilon_{K+1}\mathcal{D}_{K+1},\\ &\hat{L}_3\mathcal{D}_K = K\mathcal{D}_K, \end{aligned} \tag{23}$$

where $\hat{L}_\pm \equiv \hat{L}_1 \pm i\hat{L}_2$ and $\epsilon_K = [(J+K)(J-K+1)]^{1/2}$, one deduces the following relations connecting the action of the invariant operators on the representation functions:

$$\begin{aligned} \partial_3 &= 2iK\mathcal{D}_K,\\ \partial_1 &= -\epsilon_K\mathcal{D}_{K-1} + \epsilon_{K+1}\mathcal{D}_{K+1},\\ \partial_2 &= i\epsilon_K\mathcal{D}_{K-1} + i\epsilon_{K+1}\mathcal{D}_{K+1}; \end{aligned} \tag{24}$$

$$\begin{aligned} \partial_1\partial_2 = \partial_2\partial_1 &= (-4\hat{L}_1\hat{L}_2 - 2i\hat{L}_3)\mathcal{D}_K\\ &= i(-\epsilon_K\epsilon_{K-1}\mathcal{D}_{K-2} + \epsilon_{K+1}\epsilon_{K+2}\mathcal{D}_{K+2}),\\ \partial_2\partial_3 = \partial_3\partial_2 &= (-4\hat{L}_2\hat{L}_3 - 2i\hat{L}_1)\mathcal{D}_K\\ &= (1-2K)\epsilon_K\mathcal{D}_{K-1} - (1+2K)\epsilon_{K+1}\mathcal{D}_{K+1},\\ \partial_3\partial_1 = \partial_1\partial_3 &= (-4\hat{L}_1\hat{L}_3 + 2i\hat{L}_2)\mathcal{D}_K\\ &= i(1-2K)\epsilon_K\mathcal{D}_{K-1} + i(1+2K)\epsilon_{K+1}\mathcal{D}_{K+1}; \end{aligned} \tag{25}$$

$$\begin{aligned} \partial_1^2 &= -4\hat{L}_1^2\mathcal{D}_K\\ &= \epsilon_K\epsilon_{K-1}\mathcal{D}_{K-2} - 2[J(J+1) - K^2]\mathcal{D}_K + \epsilon_{K+1}\epsilon_{K+2}\mathcal{D}_{K+2},\\ \partial_2^2 &= -4\hat{L}_2^2\mathcal{D}_K\\ &= -\epsilon_K\epsilon_{K-1}\mathcal{D}_{K-2} - 2[J(J+1) - K^2]\mathcal{D}_K - \epsilon_{K+1}\epsilon_{K+2}\mathcal{D}_{K+2},\\ \partial_3^2 &= -4\hat{L}_3^2\mathcal{D}_K = -4K^2\mathcal{D}_K. \end{aligned} \tag{26}$$

Here ∂_i denotes the operation of the spatial derivative $\partial/\partial x_i$ on the $\mathcal{D}_K$ functions evaluated at the pole. These formulas express the transformation properties of the representation functions, and are used for the derivation of recursive relations for the amplitude coefficients.

At this point, we have completed all the necessary steps for the reduction of the covariant tensor equations. We shall demonstrate in the next section how one constructs tensor harmonics by this method.

III. VECTOR AND TENSOR HARMONICS

Scalar, vector, and tensor harmonics in a homogeneous space can in general be expressed in terms of the basis invariant forms of the space with the expansion coefficients coupling to the representation function of the particular underlying symmetry group. For $SO(3)$-homogeneous space, the general form of the scalar, vector and tensor harmonics can be expressed as

$$\Phi^{JM}(x) = \sum_{K=-J}^{J} \phi^K \mathcal{D}^J_{KM}(x), \tag{27}$$

$$A_i^{JM}(x) = \sum_{K=-J}^{J} \bar{A}^K_a\, \sigma^a_{(i)}\mathcal{D}^J_{KM}(x), \tag{28}$$

$$h_{ij}^{JM}(x) = \sum_{K=-J}^{J} \bar{h}^K_{ab}\, \sigma^a_{(i)}\,\sigma^b_{(j)}\mathcal{D}^J_{KM}(x). \tag{29}$$

The general harmonics belong to definite angular momentum states (J,M). For each definite value of J, there are $(2J+1)$ components of the amplitude coefficients coupled to the representation function in the intrinsic magnetic quantum number K. These harmonics satisfy the respective scalar, vector, and tensor wave equations (here i,j,m are space indices that run from 1 to 3)

$$\Phi_{;m}{}^{;m} = 0, \tag{30}$$

$$A_{i;m}{}^{;m} = 0, \tag{31}$$

$$h_{ij;m}{}^{;m} = 0. \tag{32}$$

In the rest of our discussion, to keep the algebra within reasonable reach, we shall limit our calculations to the diagonal metric $\bar{\gamma}_{ab} = \mathrm{diag}(\gamma_1, \gamma_2, \gamma_3)$. The method remains fully general. As a quick illustration of the group theoretical method let us first deduce the scalar harmonics of the space.

A. The scalar harmonics

The scalar harmonics are a set of scalar functions in the space that satisfy the Laplace equation [the wave equation (30)]:

$$^{(3)}\Delta\Phi \equiv \Phi_{;m}{}^{;m} = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_i}\left(\sqrt{-g}\, g^{ij}\frac{\partial}{\partial x_j}\right)\Phi = 0. \qquad (33)$$

With the scalar functions expressed in the form (27), one wants to find a set of recursive relations for the Fourier coefficients ϕ^K. The relations dictated by the Laplace equation define the set of scalar harmonics. The usual way is to express the Laplacian operator in coordinate variables and seek for a separation of variables in the equation. In the present approach we simply evaluate the Laplacian at one point, say, the pole, and (33) becomes

$$^{(3)}\Delta = \frac{1}{\gamma_1}\frac{\partial^2}{\partial x_1^2} + \frac{1}{\gamma_2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{\gamma_3}\frac{\partial^2}{\partial x_3^2}. \qquad (34)$$

Then making use of (26), we obtain almost immediately the recursion relations

$$(1/\gamma_1 - 1/\gamma_2)(\epsilon_{K+2}\epsilon_{K+1}\phi^{K+2} + \epsilon_K\epsilon_{K-1}\phi^{K-2})$$
$$- \{2[J(J+1) - K^2](1/\gamma_1 + 1/\gamma_2) + 4K^2/\gamma_3\}\phi^K = 0. \qquad (35)$$

(cf. Appendix A of Ref. 10). (In the above, the indices K in ϕ^K have been shifted by the action of the invariant operators.) The set of coefficients ϕ^K satisfying the relations (35) defines the scalar harmonics. If the expansion coefficients are made time-dependent in (30), the derived equations from the four-dimensional wave equations will describe scalar waves in a homogeneous universe.[10]

B. The vector harmonics

For the case of the vector harmonics, the amplitude functions are expanded in terms of the basis invariant forms with a coupling in the representation functions, as in (28). The coefficients $\bar{A}_a^K$ obey certain recursion relations that arise from the solution of the vector wave equation (31).[12] We first expand the covariant derivatives in (31) into ordinary derivatives and the Christoffel symbols

$$A_{i;m}{}^{;m} = g^{mn}\{A_{i,m,n} - (\Gamma^k_{im,n}A_k + \Gamma^k_{im}A_{k,n})$$
$$- (\Gamma^k_{in}A_{k,m} + \Gamma^k_{mn}A_{i,k}) + (\Gamma^k_{in}\Gamma^l_{km} + \Gamma^k_{mn}\Gamma^l_{ik})A_l\} = 0. \qquad (36)$$

Here, a comma denotes ordinary derivative. For a diagonal metric $\gamma^{mn} = (1/\gamma_m)\delta_{mn}$, we take the expressions for Γ and Γ' as given in (16), to reduce (36) to

$$(i=1)\quad \Delta A_1 - \frac{2}{\gamma_2\gamma_3}\left((\gamma_2 - \gamma_1)\frac{\partial A_3}{\partial x_2} + (\gamma_1 - \gamma_3)\frac{\partial A_2}{\partial x_3}\right)$$
$$+ \left[\frac{2}{\gamma_1\gamma_2\gamma_3}[\gamma_1^2 - (\gamma_2 - \gamma_3)^2] - \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} + \frac{1}{\gamma_3}\right)\right]A_1 = 0, \qquad (37)$$

where $\Delta \equiv (1/\gamma_m)\,\partial^2/\partial x_m^2$ is the Laplace operator. The other two equations are obtained by cyclically permuting the indices 1, 2, and 3. We then proceed to evaluate the derivative terms. Rewrite

$$A_i(x) = \sum_K A_i^K(x) \mathcal{D}_K(x) \qquad (38a)$$

and

$$A_i^K(x) = \sum_{a=1}^{3} \bar{A}_a^K \sigma^a_{(i)}(x) = -\sum_{a=1}^{3} \alpha_a^K S_{ai}(x), \qquad (38b)$$

where $\alpha_a^K \equiv -2\bar{A}_a^K$ is defined in such a way that it is equal to A_i^K at the pole. It is also understood that we are dealing with states of fixed (J, M) here. The terms that involve derivatives of Γ in (36) have been calculated before. As $A_i(x)$ are given by a product of space dependent functions, the derivatives of them contain two terms, i.e.,

$$\frac{\partial A_i}{\partial x_j} = \sum_K \left(\frac{\partial A_i^K(x)}{\partial x_j}\mathcal{D}_K + A_i^K(x)\frac{\partial \mathcal{D}_K}{\partial x_j}\right).$$

The derivatives of the $\mathcal{D}_K$ functions can be related to the action of the invariant operators and were given in (24)–(26). The derivatives of $A_i^K(x)$ can be obtained by expanding (38b) into a series in powers of x_i, in exactly the same way as was done for the metric tensor. To the second order, they are given by (at the pole):

$$A_i^K(x) = \alpha_i^K - \epsilon_{ijk}\alpha_j^K x_k + (\alpha_m^K x_m)x_i. \qquad (39)$$

From this, the first and second derivatives of $A_i^K(x)$ at the pole can be read off easily, e.g.,

$$\frac{\partial A_1^K}{\partial x_2} = \alpha_3^K, \quad \frac{\partial^2 A_2}{\partial x_1 \partial x_2} = \alpha_1^K, \quad \text{etc.}$$

Substituting the above relations for the derivative terms in (37), we get

$$\left\{\alpha_1^K\Delta + 2\left(\frac{\gamma_1}{\gamma_2\gamma_3} + \frac{1}{\gamma_2} - \frac{1}{\gamma_3}\right)\alpha_3^K\partial_2 + 2\left(-\frac{\gamma_1}{\gamma_2\gamma_3} + \frac{1}{\gamma_2} - \frac{1}{\gamma_3}\right)\alpha_2^K\partial_3\right.$$
$$\left. + \left[\left(\frac{1}{\gamma_2} + \frac{1}{\gamma_3}\right) - \frac{2}{\gamma_1\gamma_2\gamma_3}[\gamma_1^2 + (\gamma_2 - \gamma_3)^2]\right]\alpha_1^K\right\}\mathcal{D}_K(0) = 0. \qquad (40)$$

Now, all that remains is to write out by (24)–(26) the action of the invariant operators on $\mathcal{D}_K$. The shifting of indices on $\mathcal{D}_K$ is then transferred to that on the amplitude coefficients α^K, leaving a common spatial dependence in $\mathcal{D}_K$. But since $\mathcal{D}_K$ is completely general, one arrives at an equation relating the coefficients of neighboring K states valid throughout all space:

$$\left(\frac{1}{\gamma_1} - \frac{1}{\gamma_2}\right)(\epsilon_{K+2}\epsilon_{K+1}\alpha_1^{K+2} + \epsilon_{K-1}\epsilon_K\alpha_1^{K-2})$$
$$+ \left[\left(\frac{1}{\gamma_2} + \frac{1}{\gamma_3}\right) - \frac{2}{\gamma_1\gamma_2\gamma_3}[\gamma_1^2 + (\gamma_2 - \gamma_3)^2]\right.$$
$$\left. - 2\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)[J(J+1) - K^2] - \frac{4K^2}{\gamma^3}\right]\alpha_1^K$$
$$+ 4iK\left(-\frac{\gamma_1}{\gamma_2\gamma_3} + \frac{1}{\gamma_2} - \frac{1}{\gamma_3}\right)\alpha_2^K + 2i\left(\frac{\gamma_1}{\gamma_2\gamma_3} + \frac{1}{\gamma_2} - \frac{1}{\gamma_3}\right)$$
$$\times(\epsilon_{K+1}\alpha_3^{K+1} + \epsilon_K\alpha_3^{K-1}) = 0. \qquad (41)$$

For any J, there are $(2J+1)$ equations in (41). Adding two other equations resulting from permuting the indices in Eq. (40), there are altogether $3(2J+1)$ equations for an equal number of unknowns α_i^K ($i=1$ to 3, $K=-J$ to J). These sets of recursion relations for the amplitude coefficients define the vector harmonics [in the form (28)] on the diagonal $SO(3)$-homogeneous space.

C. The tensor harmonics

The tensor harmonics are constructed in much the same way as the vector harmonics. Here, the direct product of $\sigma^a\sigma^b$ form a basis for any tensor field on the space. The expansion coefficients $\bar{h}_{ab}^K$ are coupled to the representation functions $\mathcal{D}_{KM}^J$ in the form (29). Rewrite

$$h_{ij}(x)=\sum_K h_{ij}^K(x)\mathcal{D}_K(x),$$

where

$$\begin{aligned}h_{ij}^K(x)&=\sum_{a,b}\bar{h}_{ab}^K\sigma_{(i)}^a(x)\sigma_{(j)}^b(x)\\&=\sum_{a,b}h_{ab}^K S_{ai}(x)S_{bj}(x).\end{aligned}\tag{42}$$

We want to find a set of relations for the coefficients h_{ab}^K from the wave equation (32). Let us first expand all covariant derivatives in terms of ordinary derivatives and the Christoffel symbols

$$\begin{aligned}h_{ij;m}{}^{;m}=g^{mn}[&h_{ij,m,n}-(\Gamma_{im,n}^k h_{kj}+\Gamma_{jm,n}^k h_{ki})\\&-(\Gamma_{im}^k h_{kj,n}+\Gamma_{in}^k h_{kj,m}+\Gamma_{jn}^k h_{ki,m}\\&+\Gamma_{jm}^k h_{ki,n}+\Gamma_{mn}^k h_{ij,k})+(\Gamma_{in}^k\Gamma_{km}^l+\Gamma_{mn}^k\Gamma_{ki}^l)h_{lj}\\&+(\Gamma_{jn}^k\Gamma_{km}^l+\Gamma_{mn}^k\Gamma_{kj}^l)h_{il}\\&+(\Gamma_{im}^k\Gamma_{jn}^l+\Gamma_{in}^k\Gamma_{jm}^l)h_{kl}].\end{aligned}\tag{43}$$

In simplifying the wave equation one needs to evaluate the first and second derivatives of $h_{ij}^K(x)$ and $\mathcal{D}_{KM}^J$. The spatial derivatives of $\mathcal{D}_{KM}^J$ have been related to the action of the invariant operators by (24)–(26). The derivatives of $h_{ij}^K(x)$ evaluated at the pole can either be obtained by using the Killing condition or by performing a metric expansion. The Killing equation (cf. Taub, Sec. 2, in Ref. 8) $\xi_{i;j}+\xi_{j;i}=0$ yields

$$\frac{\partial h_{ij}^K}{\partial x^l}=h_{ik}^K G_{jl}^k+h_{kj}^K G_{il}^k,\tag{44}$$

where

$$G_{jl}^k\equiv-\frac{\partial\xi_a^k}{\partial x^j}(\xi^{-1})_l^a.$$

From the explicit forms of ξ_a, one finds that at the pole

$$G_{jl}^k|_0=\epsilon_{kjl}.$$

Alternatively, by making an expansion of $h_{ij}^K(x)=\bar{h}_{ab}^K\sigma^a(x)\sigma^b(x)$ in powers of x_i in exactly the same way as was done for the metric tensor, one can deduce the derivatives just by reading off the coefficients of the first and second order terms in (11) with γ_{ab} replaced by h_{ab}.

With the expressions (16) for the Christoffel symbols and their derivatives and the relations that govern the derivatives of h_{ij}^K at hand, we are now ready to simplify Eq. (43). To give some idea of how one proceeds, let us work out the $(i,j=1,1)$ component of the second derivative term in (43) as an example: First, write out the derivatives of the product $h_{11}(x)=h_{11}^K(x)\mathcal{D}_K(x)$, then,

$$\frac{1}{\gamma_m}h_{11,m,m}=\left(\frac{1}{\gamma_m}\frac{\partial^2 h_{11}^K}{\partial x_m^2}+2\frac{1}{\gamma_m}\frac{\partial h_{11}^K}{\partial x_m}\partial_m+h_{11}^K\Delta\right)\mathcal{D}_K.$$

To relate the derivatives of h_{ij}^K to h_{ab}^K, use the explicit formulas (10), which gives, e.g.,

$$\frac{\partial^2 h_{11}^K}{\partial x_2^2}=h_{33}^K-h_{11}^K,\qquad\frac{\partial h_{11}^K}{\partial x_2}=2h_{13}^K,\quad\text{etc.}$$

And, finally,

$$\begin{aligned}\frac{1}{\gamma_m}h_{11,m,m}=\Big[&\frac{h_{11}^K}{\gamma_1}+\frac{h_{33}^K-h_{11}^K}{\gamma_2}+\frac{h_{22}^K-h_{11}^K}{\gamma_3}\\&+4\left(\frac{h_{13}^K}{\gamma_2}\partial_2-\frac{h_{12}^K}{\gamma_3}\partial_3\right)+h_{11}^K\Delta\Big]\mathcal{D}_K.\end{aligned}$$

Proceeding in the same way for the other terms, after some algebra, one arrives at the following equations for the coefficients h_{ab}^K:

$$\begin{aligned}h_{11;m}{}^{;m}=\Big\{&h_{11}^K\Delta+4\left(\frac{h_{13}^K}{\gamma_2}\partial_2-\frac{h_{12}^K}{\gamma_3}\partial_3\right)\\&+h_{11}^K\left[\left(-\frac{1}{\gamma_1}+\frac{1}{\gamma_2}+\frac{1}{\gamma_3}\right)-\frac{4}{\gamma_1\gamma_2\gamma_3}(\gamma_1^2-(\gamma_2-\gamma_3)^2\right]\\&-\left(\frac{h_{22}^K}{\gamma_3}+\frac{h_{33}^K}{\gamma_2}\right)+\frac{2}{\gamma_3\gamma_2}\Big[(\gamma_1+\gamma_2-\gamma_3)^2\left(\frac{h_{22}^K}{\gamma_2}\right)\\&+(\gamma_1-\gamma_2+\gamma_3)^2\left(\frac{h_{33}^K}{\gamma_3}\right)\Big]\Big\}\mathcal{D}_K(0)=0,\end{aligned}\tag{45}$$

$$\begin{aligned}h_{12;m}{}^{;m}=\Big\{&h_{12}^K\Delta+2\left(-\frac{h_{13}^K}{\gamma_1}\partial_1+\frac{h_{23}^K}{\gamma_2}\partial_2+\frac{(h_{11}^K-h_{22}^K)}{\gamma_3}\partial_3\right)\\&+2h_{12}^K\left[4\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)-\frac{1}{\gamma_3}-3\left(\frac{\gamma_1}{\gamma_2\gamma_3}+\frac{\gamma_2}{\gamma_3\gamma_1}+\frac{\gamma_3}{\gamma_1\gamma_2}\right)\right]\\&\times\mathcal{D}_K(0)=0.\end{aligned}$$

By means of Eqs. (24)–(26), one obtains a set of recursion relations relating the coefficients h_{ab}^K of neighboring K components. These relations define the tensor harmonics in the diagonal $SO(3)$-homogeneous space.

IV. APPLICATIONS: ELECTROMAGNETIC AND GRAVITATIONAL PERTURBATIONS IN THE MIXMASTER UNIVERSE

The group theoretical method introduced here can be used to study any kind of tensor equations in a spatially homogeneous universe. One does not have to calculate the explicit forms of the tensor harmonics as defined by the set of recursion relations on the amplitude coefficients, but can proceed directly in the same way as was illustrated in the previous section. Thus, by allowing the spatial metric coefficients to be time-dependent,

$$dl^2=\bar{\gamma}_{ab}(t)\sigma^a\sigma^b,\tag{46}$$

the metric $ds^2=-dt^2+dl^2$ describes a spatially homogeneous universe. The particular type is characterized by the classification of the structure constants.[8] For $SO(3)$ symmetry, the space with diagonal metric is called the mixmaster universe.[13] To describe electromagnetic perturbations in an empty mixmaster universe, one seeks a solution to the time-dependent wave equation[14]

$$A_{\mu;\alpha}{}^{;\alpha}+R_{\mu\nu}^{(0)}A^\nu=0,\tag{47}$$

where $R^{(0)}_{\mu\nu}$, the Ricci tensor, is equal to zero for an empty background. The $A_\mu(x,t)$ are the four-dimensional vector potentials related to the electromagnetic fields $F_{\mu\nu}$ by $F_{\mu\nu} \equiv A_{\nu;\mu} - A_{\mu;\nu}$ (Greek indices run from 0 to 3). Allowing time dependence for the vector functions $\alpha^K_a(t)$ $(a=1,2,3)$ in (38) and assuming

$$A_0^{JM}(x,t) = \sum_K \alpha_0^K(t)\, D^J_{KM}(x), \tag{48}$$

one derives the differential equations for the spatial components by an extension of (41).

The additional terms in the $\mu=1$ equation (40) read as follows:

$$\{-\ddot{\alpha}_1^K + (\kappa_1 - \kappa_2 - \kappa_3)\dot{\alpha}_1^K + [\dot{\kappa}_1 + \kappa_1(3\kappa_1 + \kappa_2 + \kappa_3)]\alpha_1^K - 2\kappa_1\alpha_0^K\partial_1\}D_K(0). \tag{49a}$$

The $\mu=0$ equation is new:

$$[-\ddot{\alpha}_0^K + \alpha_0^K\Delta - (2\kappa_m/\gamma_m)\alpha_m^K\partial_m - (\textstyle\sum_m\kappa_m)\dot{\alpha}_0^K + (\textstyle\sum_m\kappa_m^2)\alpha_0^K]D_K(0) = 0. \tag{49b}$$

In addition, the vector potentials satisfy the divergence conditions

$$A_\mu{}^{;\mu} = 0 = g^{\mu\nu}(A_{\mu,\nu} - \Gamma^\lambda_{\mu\nu}A_\lambda),$$

which is resolved to

$$[\alpha_0^K + (\textstyle\sum_m\kappa_m)\alpha_0^K - (\alpha_m^K/\gamma_m)\partial_m]D_K(0) = 0. \tag{50}$$

These equations are further reduced by (24)—(26) to a set of coupled differential equations for the potential functions $\alpha^K_\mu(t)$. Equation (50) acts as a constraint equation on the variables α_μ and $\dot{\alpha}_\mu$ and the dynamic equations (49) describe the evolution of electromagnetic perturbations in the mixmaster universe.

As another example, the equations describing small first order tensor perturbations in an empty background metric are given by[15–17]

$$\delta R_{\mu\nu} = \tfrac{1}{2}(h_{\mu\nu;\alpha}{}^{;\alpha} - h_{\mu\alpha;\nu}{}^{;\alpha} - h_{\nu\alpha;\mu}{}^{;\alpha} + h^\alpha_{\alpha;\mu;\nu}) = 0. \tag{51}$$

For the diagonal mixmaster universe, this problem has been studied by Hu and Regge.[7] There, Eqs. (51) are first expanded in terms of the ordinary derivatives and the Christoffel symbols

$$\begin{aligned}2\delta R_{\mu\nu} = g^{\alpha\beta}\{&(h_{\mu\nu,\alpha,\beta} - h_{\mu\alpha,\nu,\beta} - h_{\nu\alpha,\mu,\beta}) \\ &+ 2(\Gamma^\rho_{\mu\nu,\beta}h_{\rho\alpha} + \Gamma^\rho_{\mu\nu}h_{\rho\alpha,\beta}) \\ &+ \Gamma^\rho_{\alpha\beta}(h_{\mu\rho,\nu} + h_{\nu\rho,\mu} - h_{\mu\nu,\rho}) + \Gamma^\rho_{\mu\beta}(h_{\rho\alpha,\nu} + h_{\nu\alpha,\rho} - h_{\rho\nu,\alpha}) \\ &- 2(\Gamma^\rho_{\mu\beta}\Gamma^\sigma_{\rho\nu} + \Gamma^\rho_{\nu\beta}\Gamma^\sigma_{\mu\rho})h_{\sigma\alpha} - 2\Gamma^\rho_{\alpha\beta}\Gamma^\sigma_{\mu\nu}h_{\sigma\rho} \\ &+ \Gamma^\rho_{\nu\beta}(h_{\rho\alpha,\mu} + h_{\mu\alpha,\rho} - h_{\rho\mu,\alpha})\} + h_{,\mu,\nu} - \Gamma^\rho_{\mu\nu}h_\rho\end{aligned} \tag{52}$$

where $h \equiv h^\alpha_\alpha$ is the trace of the perturbation. Then, expressing $h_{ij}(x,t)$ in terms of the tensor harmonics as in (42), one makes use of Eqs. (11), (16), and (24)—(26) to simplify (52). By following the same procedure as illustrated in Sec. III, one can derive the perturbation equations with relative ease. The final equations were given in Ref. 7 and shall not be listed here. There, the synchronous conditions

$$h_{00} = h_{0i} = 0 \tag{53}$$

were imposed. For each K component, one arrives at four constraint equations $\delta G_{00} = \delta R_{0i} = 0$ (first order differential in time) and six dynamic equations $\delta R_{ij} = 0$ (second order) for 12 unknown functions h_{ij} and $\dot{h}_{ij}$, which are all coupled. Since there are $(2J+1)$ components for each fixed J, one needs to specify $8(2J+1)$ variables as initial conditions.

Parallel to the linearized theory, there is an equivalent way to describe tensor perturbations in a curved background. In terms of a new quantity

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \tfrac{1}{2}hg_{\mu\nu} \tag{54}$$

and by specializing to the Lorentz gauge condition

$$\bar{h}_{\mu\alpha}{}^{;\alpha} = 0, \tag{55}$$

Eq. (51) can be rewritten as

$$\bar{h}_{\mu\nu;\alpha}{}^{;\alpha} + 2R^{(0)}_{\alpha\mu\beta\nu}\bar{h}^{\alpha\beta} = 0. \tag{56}$$

This equation describes the propagation of gravitational waves in a curved background. $R^{(0)}_{\alpha\mu\beta\nu}$ are the background Riemann tensor components. In addition, the trace condition[18]

$$\bar{h} = 0 \tag{57}$$

can also be imposed globally along with (55).

It is a simple extension of our treatment in Sec. III on tensor harmonics to derive the tensor wave equation (56) (we shall drop the bar over $\bar{h}_{\mu\nu}$ here and simply call them $h_{\mu\nu}$). The components h_{00}, h_{0i}, and h_{ij} are expanded as scalar, vector, and tensor harmonics, as in (27), (28), and (29) respectively. Condition (55) then reads

$$h_{\mu\alpha}{}^{;\alpha} = g^{\alpha\beta}(h_{\mu\alpha,\beta} - \Gamma^\gamma_{\mu\beta}h_{\alpha\gamma} - \Gamma^\gamma_{\alpha\beta}h_{\mu\gamma}) = 0,$$

$$(\mu=0):\quad \left[-\dot{h}^K_{00} + \frac{1}{\gamma_m}h^K_{0m}\partial_m - \left(\kappa_m\frac{h^K_{mm}}{\gamma_m}\right) - \left(\sum_m\kappa_m\right)h^K_{00}\right]D_K(0) = 0, \tag{58a}$$

$$(\mu=1):\quad \left[-\dot{h}^K_{01} + \frac{1}{\gamma_m}h^K_{1m}\partial_m + 2\left(\frac{1}{\gamma_2} - \frac{1}{\gamma_3}\right)h^K_{23} - \left(\sum_m\kappa_m\right)h^K_{01}\right]D_K(0) = 0. \tag{58b}$$

The other ten equations (56) are obtained following the method introduced above. [The Riemann tensor coefficients $R^{(0)}_{\alpha\mu\beta\nu}$ in (56) can be found in, e.g., Appendix A of Ref. 19.] Notice that under the Lorentz gauge one can no longer impose the synchronous condition (53) globally. Since the procedure is now quite familiar, we shall not proceed with the details.

The group theoretical method introduced here for the study of tensor equations in the mixmaster universe is equally applicable for homogeneous spaces of all symmetry types. The procedure can be carried over in exactly the same manner as outlined in Sec. III.

ACKNOWLEDGMENTS

The author is indebted to Professor T. Regge for his advice and guidance and to the Institute for Advanced Study, Princeton, New Jersey, for hospitality where part of this work was performed.

*Supported in part by the National Science Foundation Grant GP-39178.

[1]T. Regge and J.A. Wheeler, Phys. Rev. 108, 1063 (1957).
[2]E.M. Lifshitz and I.M. Khalatnikov, Advan. Phys. 12, 185 (1963).
[3]J. Mathews, J. Soc. Indust. Appl. Math. 10, 768 (1962).
[4]B. L. Hu, "Hyperspherical Tensor Harmonics," unpublished report, Princeton University, 1971.
[5]F.J. Zerilli, J. Math. Phys. 11, 2203 (1970).
[6]For a simple introduction to homogeneous spaces with reference to cosmology, see Landau and Lifshitz, *Classical Theory of Fields* (Pergamon, Oxford, 1971), 3rd rev. English ed., Sec. 112.
[7]B. L. Hu and T. Regge, Phys. Rev. Lett. 29, 1616 (1972).
[8]For a discussion of the Bianchi classification scheme of homogeneous space, see, e.g., A.H. Taub, Ann. Math. 53, 472 (1951); G.F.R. Ellis and M.A.H. MacCallum, Comm. Math. Phys. 12, 108 (1969).
[9]C.W. Misner, J. Math. Phys. 4, 924 (1963), Appendix B.
[10]B.L. Hu, Phys. Rev. D 8, 1048 (1973).
[11]E.g., A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).
[12]As an example of solving vector wave equations in the usual treatment, see I.M. Gel'fand, R.A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation Group and Lorentz Groups and Their Applications* (in English trans., Pergamon, Oxford, 1963), p. 110.
[13]C.W. Misner, Phys. Rev. Lett. 22, 1071 (1969).
[14]The problem of homogeneous electromagnetic fields in anisotropic universes had earlier been treated by, e.g., D.R. Brill, Phys. Rev. 133, B845 (1964); L.P. Hughston and K. Jacobs, Astrophys. J. 160, 147 (1970); N. Batakis, J.M. Cohen, Ann. Phys. 73, 578 (1972). The present problem concerns *in*homogeneous electromagnetic fields and contains the above topics as special cases (lowest mode perturbation).
[15]The derivation of this equation can be found in Ref. 1 or P.C. Peters, Phys. Rev. 146, 938 (1966) or F.J. Zerilli, Phys. Rev. D 2, 2141 (1970), Appendix C. All are on perturbation of the Schwarzschild metric.
[16]Tensor perturbations to homogeneous universes had earlier been studied by E.M. Lifshitz and I.M. Khalatnikov (Ref. 2) for the Friedmann Universe; S. Bonanos [Comm. Math. Phys. 22, 190 (1971); 26, 259 (1972)] for the Taub and other symmetric universes, and T.E. Perko, R.A. Matzner and L.C. Shepley [Phys. Rev. D 6, 969 (1972)] for the Kasner universe. See also G.E. Tauber, Tensor 14, 18 (1963).
[17]For studies on perturbations of higher order and the effect of gravitational non-linear interaction, see K. Tomita, Prog. Theor. Phys. 37, 831 (1967); R.A. Issacson, Phys. Rev. 166, 1263, 1272 (1968).
[18]See, e.g., C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 969.
[19]B.L. Hu, Ph.D. Dissertation, Princeton University, 1972 (unpublished).

Generalized static electromagnetic fields in relativity

A. F. da F. Teixeira, Idel Wolk, and M. M. Som

Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro ZC-82, Brazil

(Received 22 April 1974)

A general class of static cylindrically symmetric solutions of the Einstein–Maxwell equations coupled with a zero-rest-mass scalar field is obtained under the assumption $-r^2 g_{rr} = g_{00} g_{\phi\phi} g_{zz}$. These solutions reduce to Marder's well-known exterior solution in the absence of electromagnetic and scalar fields.

1. INTRODUCTION

In the literature there are some interesting static solutions of the Einstein—Maxwell equations. In a space—time region symmetric under rotations about a spatial axis as well as under translations about the same axis, these solutions correspond to three types of fields: azimuthal fields (Mukherji[1]), radial fields (Mukherji[1], Bonnor[2]), and longitudinal fields (Bonnor[2], Melvin[3], Ghosh and Sengupta[4]). Mukherji obtained a class of solutions of the field equations corresponding to an infinite straight wire carrying current, using pseudocylindrical coordinates. Bonnor[5] observed that with this solution for the azimuthal electromagnetic field it is difficult to interpret the constants in the way Mukherji interpreted them, as parameters representing mass, current, and radius. Indeed, one can eliminate the constant representing mass in his solution by a suitable coordinate transformation. Further, for vanishing electromagnetic field his solution goes over to Marder's[6] solution only for a particular value of the parameter associated with the gravitational mass in Marder's solution.

In the case of the already-known solutions of the field equation corresponding to an infinite line-charge, Som[7] showed that if there are no singularities in the field, then one must allow negative values of R_0^0 in the source region; negative values of R_0^0 would then require $T_0^0 - T/2 < 0$, which demands a very unusual property of matter. Thus one is forced to infer that no solution seems to exist for a line-charge with positive mass.

Similar is the situation with the solution corresponding to longitudinal fields. The only solution known so far which does not give rise to this situation is that of Melvin. However, Melvin's solution corresponds to a magnetic universe free of any source.

In this paper we have studied all these fields coupled with zero-rest mass scalar fields. Though inclusion of a scalar field does not remove the previously mentioned difficulties, some interesting results are obtained. Furthermore, we have obtained a new class of solutions of the field equations corresponding to an infinite wire carrying current. For vanishing electromagnetic and scalar fields all our solutions go over immediately to Marder's solution.

2. STATIC FIELDS

A. Basic equations

The field equations of space—time containing electromagnetic fields and a zero-rest mass scalar field, but no matter are

$$R_{\mu\nu} - g_{\mu\nu} R/2 = -\kappa(E_{\mu\nu} + S_{\mu\nu}), \qquad (2.1)$$

with

$$E_{\mu\nu} = \epsilon_0 (F_\mu{}^\alpha F_{\alpha\nu} - g_{\mu\nu} F_\alpha{}^\beta F_\beta{}^\alpha/4) \qquad (2.2)$$

and

$$S_{\mu\nu} = S_{,\mu} S_{,\nu} - g_{\mu\nu} g^{\rho\sigma} S_{,\rho} S_{,\sigma}/2, \qquad (2.3)$$

where $F_{\mu\nu}$ is the skew-symmetric electromagnetic field tensor which satisfies Maxwell's equations for empty space,

$$F_{[\mu\nu;\alpha]} = 0 \qquad (2.4)$$

and

$$F^{\mu\nu}{}_{;\nu} = 0, \qquad (2.5)$$

the semicolon denoting covariant differentiation; and S is the zero-rest mass scalar field, which satisfies

$$S_{;\mu}^{;\mu} = 0. \qquad (2.6)$$

B. Surviving components of electromagnetic fields

We shall first obtain the surviving components of the electromagnetic fields in space—time region where the metric tensor $g_{\mu\nu}$, the Ricci tensor $R_{\mu\nu}$, and the tensor $S_{\mu\nu}$ are all diagonal; then from Eq. (2.1) one has $E_{\mu\nu}$ diagonal. The vanishing of E_{01} and E_{23} implies, respectively,

$$g^{22} F_{20} F_{21} + g^{33} F_{30} F_{31} = 0 \qquad (2.7)$$

and

$$g^{00} F_{02} F_{03} + g^{11} F_{12} F_{13} = 0. \qquad (2.8)$$

One then has

$$F_{20} F_{21} (F_{30} F_{31})^{-1} < 0 \quad \text{except when } F_{20} F_{21} = F_{30} F_{31} = 0 \qquad (2.9)$$

and

$$F_{02} F_{03} (F_{12} F_{13})^{-1} > 0 \quad \text{except when } F_{02} F_{03} = F_{12} F_{13} = 0. \qquad (2.10)$$

Equations (2.7) and (2.8) are then only compatible when either the pair of components (F_{02}, F_{31}) or the pair of components (F_{03}, F_{12}) vanishes. Similar considerations about the vanishing of E_{02} and E_{13} show that at least one of the two pairs of components (F_{03}, F_{12}) and (F_{01}, F_{23}) must vanish; and the vanishing of E_{03} and E_{12} imply vanishing of at least one of the two pairs (F_{01}, F_{23}) and (F_{02}, F_{31}). Consequently, of these three pairs of components (F_{01}, F_{23}), (F_{02}, F_{31}), and (F_{03}, F_{12}), always only one pair of components survives, for nonvanishing electromagnetic fields.

 Copyright © 1974 American Institute of Physics

C. Field equations in static metrics with cylindrical symmetry

The line element for a static system with cylindrical symmetry is

$$ds^2 = e^{2\eta}\,dt^2 - e^{2\lambda}dr^2 - r^2 e^{2\beta}d\phi^2 - e^{2\gamma}dz^2, \tag{2.11}$$

with η, λ, β, and γ functions of r only. It is known (Ref. 8, Chap. 8) that for any static cylindrically symmetric system, the metric components can be described by only three functions of the radial coordinate. Usually one encounters in the literature $g_{rr} = g_{zz}$. However, one finds that with this assumption no solution exists corresponding to an infinite straight wire carrying a steady current. Another possibility is $r^2 g_{rr} = g_{\phi\phi}$, considered by Mukherji,[1] but one has difficulties in interpreting the constants (Bonnor[5]). So we try the relation $r^2 g_{rr} = -g_{00}g_{\phi\phi}g_{zz}$, that is

$$\lambda = \eta + \beta + \gamma. \tag{2.12}$$

Then the surviving components of Ricci tensor are (a subscript 1 means d/dr)

$$R^0_0 = -\exp[-2(\eta+\beta+\gamma)](\eta_{11} + \eta_1/r), \tag{2.13}$$

$$R^1_1 = -\exp[-2(\eta+\beta+\gamma)][\eta_{11} + \beta_{11} + \gamma_{11} + (\beta_1 - \eta_1 - \gamma_1)/r - 2(\eta_1\beta_1 + \eta_1\gamma_1 + \beta_1\gamma_1)], \tag{2.14}$$

$$R^2_2 = -\exp[-2(\eta+\beta+\gamma)](\beta_{11} + \beta_1/r), \tag{2.15}$$

and

$$R^3_3 = -\exp[-2(\eta+\beta+\gamma)](\gamma_{11} + \gamma_1/r). \tag{2.16}$$

Since the electromagnetic field $F_{\mu\nu}$ depends only on r, we have from (2.4)

$$F_{02} = \mathcal{E}_\phi, \tag{2.17}$$

$$F_{03} = \mathcal{E}_z, \tag{2.18}$$

and

$$F_{23} = c\mathcal{B}_r, \tag{2.19}$$

where $\mathcal{E}_\phi$, $\mathcal{E}_z$, and $\mathcal{B}_r$ are constants; and from (2.5) one has

$$F_{01} = \mathcal{E}_r r^{-1} e^{2\eta}, \tag{2.20}$$

$$F_{12} = c\mathcal{B}_z r e^{2\beta}, \tag{2.21}$$

and

$$F_{13} = c\mathcal{B}_\phi r^{-1} e^{2\gamma}, \tag{2.22}$$

where $\mathcal{E}_r$, $\mathcal{B}_z$, and $\mathcal{B}_\phi$ are constants.

From the invariant $F^\mu{}_\nu F^\nu{}_\mu$, when only one component of $F_{[\mu\nu]}$ survives, we see that in the weak field approximation we may associate the constants $\mathcal{E}_r$, $\mathcal{B}_\phi$, and $\mathcal{B}_z$ with uniform charge per unit length along the z axis, steady current along the z axis, and steady solenoidal current around the z axis, respectively; and we may associate $\mathcal{B}_r$, $\mathcal{E}_\phi$, and $\mathcal{E}_z$ with the corresponding magnetic analogues.

Also the massless scalar field depends only on r, so from (2.6)

$$S_1 = \mathcal{S} r^{-1} \tag{2.23}$$

where $\mathcal{S}$ is a constant.

If we now define the constants

$$\mathcal{C}_r = \kappa\epsilon_0(\mathcal{E}_r^2 + c^2\mathcal{B}_r^2)/2, \tag{2.24}$$

$$\mathcal{C}_\phi = \kappa\epsilon_0(\mathcal{E}_\phi^2 + c^2\mathcal{B}_\phi^2)/2, \tag{2.25}$$

and

$$\mathcal{C}_z = \kappa\epsilon_0(\mathcal{E}_z^2 + c^2\mathcal{B}_z^2)/2, \tag{2.26}$$

then the simplest form of the Einstein equations is

$$r^2\eta_{11} + r\eta_1 = \{\mathcal{C}_r e^{2\eta},\ \mathcal{C}_\phi e^{2\gamma},\ \mathcal{C}_z r^2 e^{2\beta}\}, \tag{2.27}$$

$$r^2\beta_{11} + r\beta_1 = \{-\mathcal{C}_r e^{2\eta},\ \mathcal{C}_\phi e^{2\gamma},\ -\mathcal{C}_z r^2 e^{2\beta}\}, \tag{2.28}$$

$$r^2\gamma_{11} + r\gamma_1 = \{-\mathcal{C}_r e^{2\eta},\ -\mathcal{C}_\phi e^{2\gamma},\ \mathcal{C}_z r^2 e^{2\beta}\}, \tag{2.29}$$

and

$$r^2[\eta_1\gamma_1 + (\eta_1 + \gamma_1)(\beta_1 + 1/r)] = \kappa\mathcal{S}^2/2 + \{-\mathcal{C}_r e^{2\eta},\ \mathcal{C}_\phi e^{2\gamma},\ \mathcal{C}_z r^2 e^{2\beta}\}, \tag{2.30}$$

where in each bracket $\{\}$ only one of the terms is to be considered, since in each problem only one of the constants $\mathcal{C}_r$, $\mathcal{C}_\phi$, $\mathcal{C}_z$ can be different from zero.

3. SOLUTIONS OF THE FIELD EQUATIONS

In all three problems (radial, azimuthal, and longitudinal) the method for obtaining the solution is identical; we use the first three equations for obtaining the expressions of η, β, and γ [λ is then got from (2.12)], with a total of six constants of integration; then the last equation (2.30) gives a relation which reduces these six constants to five. After that, one can easily reduce these five constants to only three essential ones, by suitable coordinate transformations. We have now three cases:

A. Azimuthal fields

In this case $\mathcal{C}_r = \mathcal{C}_z = 0$; then (2.29) gives

$$\gamma = -\log[(r/a)^b + \mathcal{C}_\phi(2b)^{-2}(r/a)^{-b}] = -\log\mathcal{B}, \tag{3.1}$$

with a and b constants of integration; the sum of (2.27) and (2.29) gives

$$\eta = -\gamma + h\log(r/d), \tag{3.2}$$

with h and d constants of integration; and the sum of (2.28) and (2.29) gives

$$\beta = -\gamma + p\log(r/f), \tag{3.3}$$

with p and f constants of integration; finally, substitution of (3.1) to (3.3) into (2.30) gives

$$ph - b^2 - \kappa\mathcal{S}^2/2 = 0. \tag{3.4}$$

One thus obtains

$$g_{00} = (r/d)^{2h}\mathcal{B}^2, \tag{3.5}$$

$$g_{11} = -(r/d)^{2h}(r/f)^{(2b^2+\kappa\mathcal{S}^2/2)/h}\mathcal{B}^2, \tag{3.6}$$

$$g_{22} = -r^2(r/f)^{(2b^2+\kappa\mathcal{S}^2/2)/h}\mathcal{B}^2, \tag{3.7}$$

and

$$g_{33} = -\mathcal{B}^{-2}. \tag{3.8}$$

With a suitable coordinate transformation, one obtains

$$g_{00} = \mathcal{J}_\phi(r/r_0)^{2(b+h)}, \tag{3.9}$$

$$g_{11} = -\mathcal{J}_\phi(r/r_0)^{-2+2(b+h)+2b^2/h+\kappa\mathcal{S}^2/h}, \tag{3.10}$$

$$g_{22} = -r^2\mathcal{J}_\phi(r/r_0)^{-2+2b^2/h+2b+\kappa\mathcal{S}^2/h}, \tag{3.11}$$

and

$$g_{33} = -\mathcal{J}_\phi^{-1}(r/r_0)^{-2b}, \quad (3.12)$$

where

$$r_0 = a(a/d)^h (a/f)^{(b^2+\kappa \mathcal{S}^2/2)/h} \quad (3.13)$$

and

$$\mathcal{J}_\phi = [1 + \mathcal{C}_\phi (2b)^{-2}(r/r_0)^{-2b}]^2. \quad (3.14)$$

Mukherji (Ref. 1, Sec. 2) gives a special case of this solution, for vanishing scalar field.

B. Radial fields

Following similar steps, one obtains for the radial case

$(\mathcal{C}_\phi = \mathcal{C}_z = 0)$

$$g_{00} = \mathcal{J}_r^{-1}(r/r_0)^{2(b+h)}, \quad (3.15)$$

$$g_{11} = -\mathcal{J}_r(r/r_0)^{-2+2(b+h)+2b^2/h+\kappa \mathcal{S}^2/h}, \quad (3.16)$$

$$g_{22} = -r^2\mathcal{J}_r(r/r_0)^{-2+2b+2b^2/h+\kappa \mathcal{S}^2/h}, \quad (3.17)$$

and

$$g_{33} = -\mathcal{J}_r(r/r_0)^{-2b}, \quad (3.18)$$

where

$$\mathcal{J}_r = \{1 - \mathcal{C}_r[2(b+h)]^{-2}(r/r_0)^{2(b+h)}\}^2. \quad (3.19)$$

Particular cases of this solution are given by Mukherji (Ref. 1, Sec. 3) and Bonnor (Ref. 2, Sec. 2.b).

C. Longitudinal fields

For the longitudinal case $(\mathcal{C}_r = \mathcal{C}_\phi = 0)$, one obtains

$$g_{00} = \mathcal{J}_z(r/r_0)^{2(b+h)}, \quad (3.20)$$

$$g_{11} = -\mathcal{J}_z(r/r_0)^{-2+b(b+h)+2b^2/h+\kappa \mathcal{S}^2/h}, \quad (3.21)$$

$$g_{22} = -r^2\mathcal{J}_z^{-1}(r/r_0)^{-2+2b+2b^2/h+\kappa \mathcal{S}^2/h}, \quad (3.22)$$

and

$$g_{33} = -\mathcal{J}_z(r/r_0)^{-2b}, \quad (3.23)$$

where

$$\mathcal{J}_z = \{1 + \mathcal{C}_z r_0^2[-2(b+h) + 2(b+h)^2/h + \kappa \mathcal{S}^2/h]^{-2} \times (r/r_0)^{-2(b+h)+2(b+h)^2/h+\kappa \mathcal{S}^2/h}\}^2. \quad (3.24)$$

Bonnor (Ref. 2, Sec. 2.a, and Ref. 5, Sec. 3) and Ghosh and Sengupta[4] give special cases of this solution.

In the absence of electromagnetic fields $(\mathcal{J}_r = \mathcal{J}_\phi = \mathcal{J}_z = 1)$, solutions for radial, azimuthal, and longitudinal problems become identical.

We next impose the condition that our coordinates be Weyl canonical ones, in the absence of electromagnetic fields; this means $g_{00}g_{22} = -r^2$, and g_{11} and g_{33} become identical as a consequence of (2.12). This imposition relates the three constants $b, h, \mathcal{S}$ by

$$h - (b+h)^2 - \kappa \mathcal{S}^2/2 = 0. \quad (3.25)$$

If we further relabel the combination $b + h$,

$$b + h = 2m, \quad (3.26)$$

then we get, in the absence of electromagnetic fields,

$$g_{00} = (r/r_0)^{4m}, \quad (3.27)$$

$$g_{11} = g_{33} = -(r/r_0)^{-4m(1-2m)+\kappa \mathcal{S}^2}, \quad (3.28)$$

and

$$g_{22} = -r^2(r/r_0)^{-4m}. \quad (3.29)$$

With electromagnetic fields the components of the metric are (3.27)–(3.29) multiplied by the corresponding factors $\mathcal{J}$ or their inverses $\mathcal{J}^{-1}$, which now take the form

$$\mathcal{J}_r = [1 - \mathcal{C}_r(4m)^{-2}(r/r_0)^{4m}]^2, \quad (3.30)$$

$$\mathcal{J}_\phi = [1 + \mathcal{C}_\phi(-4m + 8m^2 + \mathcal{S}^2)^{-2} \times (r/r_0)^{-4m+8m^2+\kappa \mathcal{S}^2}]^2, \quad (3.31)$$

and

$$\mathcal{J}_z = [1 + \mathcal{C}_z r_0^2(2 - 4m)^{-2}(r/r_0)^{2-4m}]^2. \quad (3.32)$$

We have thus obtained a set of solutions which go over to the solution given by Marder[6] in the absence of electromagnetic and massless scalar fields. In the absence of a scalar field, making parameter m vanish in the longitudinal solution gives the electromagnetic geon of Melvin.[3] It should be noticed that our radial and longitudinal solutions are expressed in Weyl canonical coordinates, while our azimuthal solution is not.

4. DISCUSSION OF THE RESULTS

In Sec. 2B we showed that static cylindrically symmetric systems may contain exclusively radial, or azimuthal or longitudinal electromagnetic field, not combinations of these fields. Choice of coordinates $r^2 g_{rr} = -g_{00}g_{\phi\phi}g_{zz}$ proved to give equations easier to solve than others more frequently used in literature, such as $g_{rr} = g_{zz}$ or $g_{\phi\phi} = r^2 g_{rr}$; our choice of coordinates and an appropriate labeling of constant of integration allowed solutions of the three independent systems to become identical in the absence of electromagnetic fields; and further these tend to Marder's solution in the case of a vanishing massless scalar field, when one of the two constants b and h is fixed by the condition (3.25).

It is evident from Eqs. (3.30) to (3.32) that one cannot have $m = 0$ for a radial field and $m = 1/2$ for a longitudinal field, independently of the scalar field. With an azimuthal field the situation is different: only in the absence of scalar field does the solution corresponding to azimuthal field not allow both $m = 0$ and $m = 1/2$. For vanishing electromagnetic field the expression for Kretschman scalar takes the form

$$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = [64m^2(1-2m)^2(1-2m+4m^2) - 16m(1-2m)(1-4m+8m^2)\kappa \mathcal{S}^2 + (3 - 8m + 16m^2)\kappa^2 \mathcal{S}^4] r_0^{-4} \times (r/r_0)^{-4+8m-16m^2-2\kappa \mathcal{S}^2}; \quad (4.1)$$

when $\mathcal{S} \neq 0$ this is always positive and tends to zero at infinity. However, when $\mathcal{S} = 0$ it tends to zero everywhere as m tends to either the value zero or to one-half.

Janis *et al.*[9] prescribed a method of obtaining some generalized electromagnetic fields from the vacuum field solutions irrespective of any symmetry, but their prescriptions do not admit combination of electric and magnetic fields. However, our solutions allow combinations of electric and magnetic fields.

ACKNOWLEDGMENTS

One of the authors (M.M.S.) acknowledges financial support from Conselho Nacional de Pesquisas (Brasil).

[1]B.C. Mukherji, Bull. Calcutta Math. Soc. **30**, 95 (1938).
[2]W.B. Bonnor, Proc. Phys. Soc. **66** A, 145 (1953).
[3]M.A. Melvin, Phys. Lett. **8**, 65 (1964).
[4]R. Ghosh and R. Sengupta, Nuovo Cimento **38**, 1579 (1965).
[5]W.B. Bonnor, Proc. Phys. Soc. **67** A, 225 (1954).
[6]L. Marder, Proc. Roy. Soc. A **244**, 524 (1958).
[7]M.M. Som, Proc. Phys. Soc. **83**, 328 (1964).
[8]J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).
[9]A.L. Janis, D.C. Robinson, and J. Winicour, Phys. Rev. **186**, 1729 (1969).

Gauge theories and Galilean symmetry

P. Roman and J. P. Leveille

Department of Physics, Boston University, Boston, Massachusetts 02215

(Received 30 May 1974)

Relations between a nonrelativistic local phase symmetry and the Galilean structure of the operator algebra are studied. The latter is derived from a few simple assumptions. For interacting systems, the assumption of phase independent localization leads to a unique Hamiltonian. Superselection rules for mass, time, and charge appear in intimate interrelationships.

1. INTRODUCTION

Ever since Weyl[1] so forcefully pointed out the role of gauge transformations in quantum theory, this topic held a fascination for many physicists. The idea, originally formulated for electromagnetic interactions, was put into a new perspective (local phase invariance) and extended to non-Abelian symmetries by Yang and Mills and by Utiyama.[2] In recent years, combining gauge invariance with spontaneous symmetry breaking, impressive progress was made toward the understanding and the unifying of elementary particle interactions.[3]

In his penetrating analyses of symmetry principles, it was repeatedly pointed out by Wigner[4] that gauge symmetry is radically different from other, classical (or geometrical) symmetries. Gauge transformations do not affect observations and do not correlate events. They appear to apply to *specific interactions* and are formulated in terms of the laws of nature. For this reason, Wigner uses the term "dynamical invariance" when speaking of gauge symmetry.

The dynamical nature of gauge invariances is further illustrated if we consider the (nonrelativistic and nonquantum mechanical) Lagrangian formalism. The Lagrangian of a classical point particle (with unit charge) in interaction with an electromagnetic field is

$$L=\tfrac{1}{2}M\dot{\mathbf{q}}^2-V+\dot{\mathbf{q}}\mathbf{A}. \qquad (1.1)$$

This is obviously not invariant under an electromagnetic gauge transformation

$$V\to V+\dot{\omega},\quad \mathbf{A}\to\mathbf{A}-\operatorname{grad}\omega. \qquad (1.2)$$

However, the corresponding action and the equation of motion resulting from (1.1), i.e.,

$$M\ddot{\mathbf{q}}=-\operatorname{grad}V-\dot{\mathbf{A}}+\dot{\mathbf{q}}\times\operatorname{curl}\mathbf{A}, \qquad (1.3)$$

are gauge invariant.

The situation is rather different in quantum mechanics. The Lagrangian density

$$\mathcal{L}=i\psi^*(\partial_t+iV)\psi-(1/2M)(\partial_k+iA_k)\psi^*(\partial_k-iA_k)\psi \qquad (1.4)$$

leads to the Schrödinger equation

$$i\partial_t\psi+(1/2M)(\partial_k-iA_k)(\partial_k-iA_k)\psi-V\psi=0. \qquad (1.5)$$

Now neither (1.4) nor (1.5) is invariant under (1.2). This is an untenable situation because, unlike the field strengths, the potentials are not measurable and determine the former only up to the gauge transformation (1.2). We therefore are led to postulate that every gauge transformation (1.2) of the fields must be accompanied by a gauge transformation

$$\psi\to e^{-i\omega(\mathbf{x},t)}\psi \qquad (1.6)$$

of the matter wavefunction. Then both the Lagrangian (1.4) and the dynamical equation (1.5) are invariant. In fact, the latter describes correctly the nonrelativistic quantum theoretical behavior of a (spinless) particle in an external electromagnetic field.

At this point we make an important stipulation: Galilean invariance of nonrelativistic physics is, very much like gauge invariance, essentially a dynamical invariance. This may sound surprising, because, after all, proper Galilei transformations are inertial transformations and, thus, they affect observations. To bring out our point more clearly, let us take the Lagrangian of a free point particle,

$$L=\tfrac{1}{2}M\dot{\mathbf{q}}^2. \qquad (1.7)$$

Whereas L is invariant only under Euclidean transformations and time displacements, the corresponding action and the equation of motion,

$$\ddot{\mathbf{q}}=0, \qquad (1.8)$$

are also invariant with respect to Galilei transformations. In fact, as Houtappel, van Dam, and Wigner[5] pointed out, even if we add to (1.7) a *velocity independent* potential V (which is Euclidean and time-displacement invariant), the resulting equation of motion will be Galilei invariant, even though the Lagrangian is not. Thus, Galilean invariance of the *dynamics* emerges as a consequence of a *smaller kinetic* invariance *provided the forces are of a special kind.*[6] This situation is in complete analogy with the emergence of the gauge invariant Eq. (1.3) from the nongauge invariant (1.1), where the invariance arose again from the special form of the interaction. In contrast, we observe that in relativistic mechanics the Lagrangian $L=\frac{1}{2}mu_\nu u^\nu$, which leads to the Lorentz invariant equation of motion $d(mu_\nu)/d\tau=0$, is already invariant under Lorentz transformations. Thus, inertial transformations in relativity are completely kinematical in nature, in contrast to nonrelativistic physics.

Let us now consider again nonrelativistic quantum mechanics. For a free particle we set

$$\mathcal{L}=i\psi^*\dot{\psi}-(1/2M)\partial_k\psi^*\partial_k\psi, \qquad (1.9)$$

and obtain the Schrödinger equation

$$i\partial_t\psi+(1/2M)\partial_k\partial_k\psi=0. \qquad (1.10)$$

If we do not transform ψ while performing a Galilei transformation on the coordinates, neither (1.9) nor (1.10) is invariant. To achieve invariance, we have to postulate that every Galilei transformation must be accompanied by a transformation

 Copyright © 1974 American Institute of Physics

$$\psi \to e^{if(\mathbf{x},t)}\psi, \quad \text{with } f = \tfrac{1}{2}Mv^2t + M\mathbf{v}\mathbf{x}, \tag{1.11}$$

of the wavefunction.[7] Comparing these observations with the paragraph preceding Eq. (1.6), we again see the close resemblance between the character of gauge and Galilei symmetry.

A much deeper relation between gauge and Galilei invariance is observed if one considers the introduction of an *external force*. It is easily seen that, if one adds a term $V(x)\psi^*\psi$ to (1.9) [with $V(x)$ not constant], then Galilean invariance of the resulting Schröndinger equation cannot be achieved, not even if V is velocity independent, space- and time-translation invariant.[8] However, if one couples the matter field ψ to an external electromagnetic field, as given by (1.4), then Galilei invariance of the corresponding Schrödinger equation (1.5) can be restored by taking advantage of the available gauge freedom, i.e., by performing also a suitable gauge transformation[9] on V, $\mathbf{A}$ (and ψ). Such Galilean invariant wave equations of particles (with arbitrary spin) in an external electromagnetic field have been studied by Lévy-Leblond.[10]

The surprising relation between these two seemingly disparate symmetries (which has no counterpart in standard relativistic quantum theory) was first noticed and studied, from a more general viewpoint, by Jauch.[11] He defined a *kinematical symmetry transformation* as a permutation of the set of all observables of a system which can be globally implemented by a unitary operator on the Hilbert space.[12] He then defined a physical system to be Galilei invariant if the transformation $Q_k \to Q_k$, $\dot{Q}_k \to \dot{Q}_k + v_k$ of the position and velocity[13] is a kinematical symmetry transformation. Within this framework (assuming that the Q_k form a complete set of commuting operators) he showed that the most general form for the Hamiltonian of a Galilei invariant system is given by

$$H = (1/2M)(\mathbf{P} - \mathbf{A})^2 + V, \tag{1.12}$$

where $\mathbf{A}$ and V are arbitrary functions of the coordinates (and possibly of time). Note that this result is different (and more profound) than our above observation concerning Galilei invariance (in the standard sense) of the interacting wave equation. But Jauch also showed that a unitarily implementable local phase transformation $\psi \to \exp(-i\omega)\psi$ with an arbitrary differentiable function ω is equivalent to the replacements (1.2) of V and $\mathbf{A}$ in the Schrödinger equation corresponding to (1.12). This ties up the two observations.

Recently Piron[14] rederived, in a slightly different framework, Jauch's interesting result. Lévy-Leblond[10,15] was also fascinated by the connections between Galilei and gauge invariance and called for a detailed analysis.

In this paper we study the problem from essentially the opposite direction than was done in Jauch's work. Adopting a locality postulate, we shall arrive in a natural manner at the Galilei group of a free particle. We then consider an interacting system, and adding the requirement that localization be phase independent, we obtain the unique form (1.12) for the Hamiltonian, with $\mathbf{A}$ being subject to $\mathbf{A} \to \mathbf{A} - \partial_k\omega$. Then we extend the locality postulate to hold also for time dependent local phases. We find that consistency then requires the transformation law $V \to V + \partial_t\omega$. Finally, we study various superselection rules that arise in the theory and point out their remarkable interrelationships.

2. THE KINEMATICAL GROUP

We commence with adopting the usual geometry for the space of nonrelativistic physics:

Assumption 1: The space of events is the homogeneous and isotropic Euclidean space E_3.

From this follows the existence of the symmetry group $E(3)$ of Euclidean transformations, with the Lie algebra

$$[P_k, P_l] = 0, \quad [J_k, P_l] = i\epsilon_{klj}P_j,$$
$$[J_k, J_l] = i\epsilon_{klj}J_j. \tag{2.1}$$

This algebra can be realized on the Hilbert space of square integrable wave functions $\psi(\mathbf{x})$ by setting

$$P_k \sim -i\partial_k, \quad J_k \sim -i\epsilon_{klj}x_l\partial_j. \tag{2.2}$$

The next, and crucial, step is the adoption of a *locality postulate*. Following the familiar argument[2] we stipulate that the phase of a wavefunction is a matter of convention, not only at a given point but also when we compare phases at different points. In other words, *we demand that a local phase transformation be an automorphism of Hilbert space*. In view of Wigner's theorem,[16] we can formalize this requirement as follows.

Assumption 2: To every transformation

$$\psi(\mathbf{x}) \to e^{i\omega(\mathbf{x})}\psi(\mathbf{x}) \tag{2.3}$$

with a differentiable $\omega(\mathbf{x})$, there corresponds in Hilbert space a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\psi)(\mathbf{x}) = e^{i\omega(\mathbf{x})}\psi(\mathbf{x}). \tag{2.4}$$

Using the realization (2.2), we now calculate

$$(\mathcal{U}P_k\mathcal{U}^{-1}\psi)(\mathbf{x}) = e^{i\omega}[-i\partial_k e^{-i\omega}\psi(\mathbf{x})] = (-i\partial_k - \partial_k\omega)\psi(\mathbf{x}),$$

i.e., under a local phase transformation (2.3)

$$P_k \to P_k - \partial_k\omega. \tag{2.5}$$

Similarly we find that

$$J_k \to J_k - \epsilon_{klm}x_l\partial_m\omega. \tag{2.6}$$

These results show that, unless we enlarge the algebra of observables, arbitrary local phase transformations cannot be kinematical transformations in Jauch's sense.[11] Indeed, by setting $\mathcal{U} = \exp(iF)$, Eq. (2.5) would imply that

$$P_k - \partial_k\omega = \mathcal{U}P_k\mathcal{U}^{-1} = P_k + i[F, P_k] + \ldots, \tag{2.7}$$

and since at this stage F is necessarily a function of $\mathbf{P}$ and $\mathbf{J}$ while $\partial_k\omega$ is a c-number multiple of the identity operator, (2.1) tells us that this equation cannot be satisfied (unless $\omega = \text{const.}$)

Suppose we postulate

Assumption 3: The algebra of observables is large enough to guarantee that arbitrary local phase transformations with a differentiable $\omega(\mathbf{x})$ are kinematical transformations.

How is the set $\{\mathbf{P}, \mathbf{J}\}$ of fundamental observables to be enlarged so as to satisfy Assumption 3? This question is answered by

Theorem 1: In order to satisfy Assumption 3, it is sufficient to adjoin to the set $\{\mathbf{P},\mathbf{J}\}$ the identity operator I and the generators Q_l $(l=1,2,3)$ of *linear* local phase transformations [corresponding to[17] $\omega(\mathbf{x})=c_l x_l$].

Proof: If $\omega(\mathbf{x})=c_l x_l$ and if we write[18] $F=M^{-1}c_l Q_l$, then Eq. (2.7) is satisfied provided we have

$$[P_k, Q_l]=-iM\delta_{kl}. \tag{2.8}$$

Since the c_l are linearly independent, apart from (2.8) we also have

$$[Q_k, Q_l]=0. \tag{2.9}$$

Equations (2.8) and (2.9) tell us that Q_k can be realized by setting

$$Q_k \sim Mx_k. \tag{2.10}$$

Furthermore, with $U=\exp(iM^{-1}c_l Q_l)$ we find that

$$UJ_kU^{-1}=J_k+iM^{-1}c_l[Q_l, J_k]+\ldots,$$

and, comparing this with (2.6), using the realization (2.10), and noting that now $\partial_m\omega=c_m$, we see that consistency requires

$$[J_k, Q_l]=i\epsilon_{klj}Q_j. \tag{2.11}$$

Let now ω be an arbitrary differentiable function, i.e., $\omega(\mathbf{x})=\sum_{n=0}^{\infty}c(\mathbf{a}_n\cdot\mathbf{x})^n$. The effect of the corresponding unitary transformation (whose existence is guaranteed by Assumption 2) on the operator algebra is characterized by (2.5), (2.6), and

$$Q_k\to Q_k. \tag{2.12}$$

Since $\partial_k\omega$ and $x_l\partial_m\omega$ in (2.5) and (2.6) is a power series in x, and since the realization (2.10) holds, the rhs of (2.5), (2.6), (2.12) is simply a permutation of the operator algebra, so that we have a kinematical symmetry transformation. This concludes the proof.

Remarks: (a) With $\omega=c_l x_l$, Eqs. (2.12), (2.5), (2.6) give

$$\begin{aligned}&Q_k\to Q_k,\\ &P_k\to P_k-c_k,\\ &J_k\to J_k-M^{-1}\epsilon_{klm}c_m Q_l\end{aligned} \tag{2.13}$$

as the effect of the corresponding local phase transformation. Thus, *Galilean boosts arise as particular local phase transformations.*

(b) The Heisenberg commutation relations (2.8) as well as the other two relations (2.9) and (2.11) involving Q have the role of consistency requirements.

(c) The algebra of observables is characterized by the Lie relations (2.1), (2.8), (2.9), and (2.11). The structure of the corresponding simply connected Lie group is[19,20]

$$K=SU(2)^J\otimes[T_3^P\otimes(T_3^Q\times T_1^M)]. \tag{2.14}$$

(d) Since we passed to the covering group $SU(2)^J$, the wavefunctions will be vector valued representations and should be labeled as $\psi^s_{s_3}(\mathbf{x})$, where s, s_3 are $SU(2)^J$ labels. Correspondingly, the realization (2.2) of J_k must be changed to

$$J_k\sim -i\epsilon_{klj}x_l\partial_j+\Sigma_k, \tag{2.15}$$

where Σ_k is an $SU(2)$ matrix. We then define spin T as the difference between total and "orbital" angular momentum,

$$\mathbf{T}\equiv\mathbf{J}-M^{-1}\mathbf{Q}\times\mathbf{P}=\Sigma. \tag{2.16}$$

The Casimir invariants of K are

$$C_1=MI, \tag{2.17a}$$

$$C_2=\mathbf{T}^2. \tag{2.17b}$$

We can interpret C_1 (which arose from linear phase transformations and indicates a superselection rule) as Galilean mass. Because of (2.16), the spectrum of C_2 is $s(s+1)$ with $s=0, 1/2, 1,\ldots$. Since the irreducible unitary representations of K are characterized by specifying the mass M and the spin s (which are the *kinematical* labels of a particle), we shall call K the *kinematical group*.

3. THE DYNAMICAL GROUP

In order to introduce dynamics, we first make the following definition.

Definition 1: A development transformation of an isolated system is a kinematical symmetry characterized by

$$\mathbf{P}\to\mathbf{P},\quad \mathbf{J}\to\mathbf{J},\quad \mathbf{Q}\to f(\mathbf{Q},\mathbf{P},\mathbf{J}).$$

The motivation of this form is that the intrinsic development must be compatible with the geometry of space, i.e., the corresponding generator should be invariant under space translations and rotations. We further desire that development transformations be continuously composable in an associative manner, be invertible, and independent of order. This means that the set of all development transformations must form an Abelian group. The simplest possibility is that we have a one-parameter group. Thus, we make

Assumption 4: Development transformations form a one-parameter Lie group T_1^H.

Any development transformation τ will then be represented by a unitary operator $U_\tau=\exp(i\tau H)$. Concerning the generator H we stipulate

Assumption 5: H is contained in the algebra of observables generated by P, Q, J.

This rather obvious assumption is weaker than demanding that, for example, the P_k form a complete set of commuting observables. Nevertheless, when combined with the P and J invariance requirement of Definition 1, it is powerful enough to tell us that[21]

$$H=H(\mathbf{P}^2,\mathbf{TP}, I). \tag{3.1}$$

Next we observe that development transformations of T_1^H give rise to an equivalence relation on the algebra of observables generated by the kinematic Lie group K. Indeed the relation $A\sim B$ iff $B=U_\tau AU_\tau^{-1}$ for some τ is easily seen to be an equivalence relation.[22] It is therefore reasonable to define a *dynamical group* G by the following

Assumption 6: The kinematical group K is isomorphic to the quotient group modulo T_1^H of some group G.

Thus, $K\approx G/T_1^H$, which implies that H and the generators of K must form a closed Lie algebra. This, then, makes the choice of the rhs in Eq. (3.1) unique,[23] and we have

$$H = P^2/2M - \mathcal{C}_1. \tag{3.2}$$

Here $\mathcal{C}_1$ is an arbitrary constant and the scale factor $2M$ was chosen for convenience.

With this form of H and the already known Lie relations of the kinematical group we find that

$$[H, Q_k] = -iP_k \tag{3.3}$$

and, of course,

$$[H, \mathrm{P}] = 0, \quad [H, \mathrm{J}] = 0. \tag{3.4}$$

The relations (2.1), (2.8), (2.9), (2.11) together with (3.3), (3.4) form the Lie algebra of the dynamical group G, and we observe that *it is precisely the abstract quantum mechanical Galilei group algebra*. Its structure is given as

$$G = T_1^H \otimes K = T_1^H \otimes \{SU(2)^J \otimes [T_3^P \otimes (T_3^Q \times T_1^M)]\}. \tag{3.5}$$

We can write $G = \mathcal{G} \otimes T_1^M$, where $\mathcal{G}$ is the abstract geometrical Galilei group, so that G is its scalar central extension.[24] By denoting the parameters associated with T_1^H, T_3^P, T_3^Q, $SO(3)^J$ by τ, a, v, R, respectively, exponentiation of the Lie algebra leads to the familiar composition law

$$(\tau, \mathrm{a}, \mathrm{v}, R)(\bar{\tau}, \bar{\mathrm{a}}, \bar{\mathrm{v}}, \bar{R}) = (\tau + \bar{\tau}, \mathrm{a} + R\bar{\mathrm{a}} + \bar{\tau}\mathrm{v}, \mathrm{v} + R\bar{\mathrm{v}}, R\bar{R}). \tag{3.6}$$

It is convenient to represent this abstract group on some homogeneous space. The simplest choice is to take the left coset space $\mathcal{G}/SO(3)^J \otimes T_3^Q$, whose elements[25] can be characterized by the pair $(\bar{\tau}, \bar{\mathrm{a}})$. Using (3.6), we find that the left action of $\mathcal{G}$ on the coset space is given by

$$(\bar{\tau}, \bar{\mathrm{a}}) \to (\bar{\tau} + \tau, R\bar{\mathrm{a}} + \mathrm{a} + \bar{\tau}\mathrm{v}). \tag{3.7}$$

We can identify our homogeneous space with $E_3(\mathrm{x}) \times E_1(t)$ by the map $(\bar{\tau}, \bar{\mathrm{a}}) \to (t, \mathrm{x})$ so that (3.7) gives

$$t \to t + \tau, \quad \mathrm{x} \to R\mathrm{x} + \mathrm{a} + t\mathrm{v}. \tag{3.8}$$

Thus, the familiar active viewpoint of the Galilei group consists in considering it as a set of endomorphisms of $E_3(\mathrm{x}) \times E_1(t)$. The relation between (3.7) and (3.8) is an isomorphism.

Even though the above procedure is hardly new, it permits us to interpret "nonrelativistic time" in a purely group theoretic manner. The one-dimensional space $E_1(t)$ was introduced, not at the start of kinematical considerations, but rather simply as a convenience permitting a simple active characterization of the dynamical group. It is possible to use, as a homogeneous space, not $\mathcal{G}/SO(3)^J \otimes T_3^Q$ but, for example $\mathcal{G}/SO(3)^J$. Then one is led to a representation of $\mathcal{G}$ on "phase space" $E_3(\mathrm{x}) \times E_3(\mathrm{p})$ and no explicit concept of "time variable" arises[26] (cf. Appendix A).

Once, however, the choice has been made to use the homogeneous space $\mathcal{G}/SO(3)^J \times T_3^Q$, we are led, in a natural manner, to a *sequence* of incoherent Hilbert spaces. We define, for each t, a Hilbert space $\mathcal{H}_t$ of square integrable functions by setting

$$\psi(\mathrm{x}; t) = \exp(-itH)\psi(\mathrm{x}) \tag{3.9}$$

and the total Hilbert space $\mathcal{H}$ is then a suitable direct integral of the "slices" $\mathcal{H}_t$. Whereas so far our observables P, Q, J, H were realized on $\mathcal{H}_{t=0}$, we can now search for their realization by differential operators on all of $\mathcal{H}$. A glance at (2.1), (2.8), (2.9), (2.11), (3.3), (3.4) shows that we can set

$$\begin{aligned} P_k &\sim -i\partial_k, \\ Q_k &\sim Mx_k + it\partial_k, \\ J_k &\sim -i\epsilon_{klj}x_l\partial_j + \Sigma_k, \\ H &\sim i\partial_t. \end{aligned} \tag{3.10}$$

In particular, H assumed a double role: On each slice $\mathcal{H}_t$ it has the realization $H \sim -(2M)^{-1}\partial_k\partial_k + \mathcal{C}_1$, whereas on $\mathcal{H}$ it is given by $i\partial_t$. This is emphasized by the usual Schrödinger equation (1.10), which arises when one applies[27] the Casimir invariant $\mathcal{C}_1$ of G, given by (3.2), onto the function space $\psi(\mathrm{x}; t)$. From this viewpoint, the emergence of the Schrödinger equation as a consistency condition is related to having selected the "homogeneous Galilei group" $SO(3)^J \otimes T_3^Q$ as the subgroup which defines a homogeneous G-space.

4. INTERACTING PARTICLES

The transformations of the basic observables acting on $\mathcal{H}$, which they undergo when a local phase transformation $\psi(\mathrm{x}; t) \to \exp[i\omega(\mathrm{x})]\psi(\mathrm{x}; t)$ is performed,[28] can be easily obtained if one uses the realizations (3.10). We get

$$P_k \to P_k - \partial_k\omega, \tag{4.1a}$$

$$Q_k \to Q_k + t\partial_k\omega, \tag{4.1b}$$

$$J_k \to J_k - M^{-1}\epsilon_{klm}Q_l\partial_m\omega, \tag{4.1c}$$

$$H \to H. \tag{4.1d}$$

Naturally, this permutation of observables (represented by a unitary operator on $\mathcal{H}$) is a kinematical symmetry transformation.[29]

Equation (4.1b) tells us that, *except on the slice* $t = 0$, the position operator Q is not invariant under local phase transformations. There is no reason why localization should be independent of the choice of phase $\omega(\mathrm{x})$ on slice $\mathcal{H}_{t=0}$, but depend on it on other slices. We therefore stipulate

Assumption 7: Localization does not depend on the choice of a phase $\omega(\mathrm{x})$.

In other words, we assume that arbitrary local phase transformations with a differentiable $\omega(\mathrm{x})$ commute with the particular local phase transformations[30] with $\omega(\mathrm{x}) = c_l x_l$. We shall call systems for which Q is invariant under transformations corresponding to $\psi(\mathrm{x}; t) \to \exp[i\omega(\mathrm{x})]\psi(\mathrm{x}; t)$, *covariantly interacting systems*. In this terminology, Assumption 7 may be paraphrased as stipulating that all physical systems are covariantly interacting. The question now arises: What characterizes a covariantly interacting system? This is answered by

Theorem 2: The Hamiltonian of a covariantly interacting spinless system has the form

$$H = (1/2M)(\mathrm{P} - \mathrm{A})^2 + V, \tag{4.2}$$

where A and V depend on $\mathrm{Q}(t)$ and where, under a local phase transformation,

$$A_k \to A_k - \partial_k\omega. \tag{4.3}$$

Proof: In order to satisfy the requirement that $\mathrm{Q} \to \mathrm{Q}$ under an arbitrary phase transformation, we must obviously modify the realization (3.10) of Q. We set[31]

$$Q_k \sim Mx_k + it\partial_k + tA_k(\mathbf{x}). \tag{4.4}$$

We now calculate, with $\mathcal{U}$ being the representative of the arbitrary phase transformation,

$$(\mathcal{U} Q_k \mathcal{U}^{-1}\psi)(\mathbf{x};t)$$
$$= e^{i\omega}[(Mx_k + it\partial_k)e^{-i\omega}\psi(\mathbf{x};t)] + t(\mathcal{U} A_k \mathcal{U}^{-1}\psi)(\mathbf{x};t)$$
$$= (Mx_k + it\partial_k + t\partial_k\omega)\psi(\mathbf{x};t) + t(\mathcal{U} A_k \mathcal{U}^{-1}\psi)(\mathbf{x};t).$$

Thus, $\mathbf{Q} \to \mathbf{Q}$ provided that

$$A_k \to \mathcal{U} A_k \mathcal{U}^{-1} = A_k - \partial_k\omega. \tag{4.5}$$

Next, we use (4.4) and $H \sim i\partial_t$, $P_k \sim -i\partial_k$ and compute that

$$[H, Q_k] = -i(P_k - A_k). \tag{4.6}$$

From this we can find H as a function of the operator algebra. We first observe that (4.6) refers to operators defined on $\mathcal{H}$. By transforming this equation with $\exp(it\mathcal{H})$ we obtain, in view of (3.9), the corresponding equation for the slice $\mathcal{H}_{t=0}$. Distinguishing operators on this slice by putting a bar over them, we have $[\bar{H}, \bar{Q}_k] = -i(\bar{P}_k - \bar{A}_k)$. Since $\bar{A}_k$ is a power series in $\bar{\mathbf{Q}}$ and since $[\bar{P}_k, \bar{Q}_l] = -iM\delta_{kl}$, we easily find that

$$\bar{H} = (1/2M)\bar{\mathbf{P}}^2 - (1/2M)\bar{P}_l\bar{A}_l - (1/2M)\bar{A}_l\bar{P}_l + \bar{N},$$

where $\bar{N}$ is an arbitrary function of $\bar{Q}$. However, this can be trivially rewritten as

$$\bar{H} = (1/2M)(\bar{\mathbf{P}} - \bar{\mathbf{A}})^2 + \bar{V}, \tag{4.7}$$

where again $\bar{V}$ is an arbitrary function of $\bar{\mathbf{Q}}$. Transforming this equation with $\exp(-itH)$, we obtain H on the slice $\mathcal{H}_t$, which is precisely the form given by (4.2).

QED

Remarks: (a) The meaning of Theorem 2 is that the essentially *kinematical* requirement "localization should be invariant under a local phase transformation with a phase $\omega(\mathbf{x})$ throughout all of $\mathcal{H}$" leads to the *necessity of an interaction* with fields $\mathbf{A}$ and V, and this interaction has a uniquely prescribed form.

(b) In the presence of interactions the behavior of P, J, H under arbitrary local phase transformations with $\omega(\mathbf{x})$ still persists as given by (4.1a, c, d) but (4.1b) is, of course, replaced by $Q_k \to Q_k$.

(c) When Eq. (4.2) is realized by differential operators, we obtain the Schrödinger equation (1.5). This is now invariant under "gauge transformations with a *time independent* $\omega(\mathbf{x})$", i.e., under the simultaneous replacements

$$\psi(\mathbf{x};t) \to e^{-i\omega(\mathbf{x})}\psi(\mathbf{x};t), \quad A_k \to A_k - \partial_k\omega(\mathbf{x}).$$

It may be worth while to point out that, since we still have the realization $P_k \sim -i\partial_k$, the operator P_k retains its meaning as a translation operator even in case of interactions. However, it cannot be identified with momentum. Since momentum is defined as the time derivative of position and since time derivative of any observable Ω is given by[32]

$$\frac{d\Omega}{dt} \equiv \dot{\Omega} = i[H, \Omega] + \partial_t\Omega, \tag{4.8}$$

Eq. (4.6) tells us that the momentum is now given by

$$\mathcal{P} = \mathbf{P} - \mathbf{A}. \tag{4.9}$$

Correspondingly we have

$$H = (1/2M)\mathcal{P}^2 + V. \tag{4.10}$$

Observe that both $\mathcal{P}$ and H are gauge invariant under time independent gauge transformations (but P is not).

We now observe the following. When, in Sec. 3, we decided to represent the abstract dynamical group on the homogeneous space $\mathcal{G}/SO(3)^J \otimes T_3^Q$ and thus were led to a sequence $\mathcal{H}_t$ of incoherent Hilbert spaces, we effectively introduced a superselection rule. It is reasonable to require that this superselection rule be made explicit. This can be easily done if we extend our basic locality postulate by stipulating that the phase of a wavefunction is a matter of convention, not only in each $\mathcal{H}_t$ but also when we compare phases at different slices $\mathcal{H}_t$ of $\mathcal{H}$. In other words, we demand that a local phase transformation with arbitrary, time dependent $\omega(\mathbf{x}, t)$ be an automorphism of $\mathcal{H}$. Thus, we replace Assumption 2 by the more general one:

Assumption 8: To every transformation

$$\psi(\mathbf{x};t) \to e^{i\omega(\mathbf{x},t)}\psi(\mathbf{x};t) \tag{4.11}$$

with a differentiable $\omega(\mathbf{x}, t)$ there corresponds in the Hilbert space $\mathcal{H}$ a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\psi)(\mathbf{x};t) = e^{i\omega(\mathbf{x},t)}\psi(\mathbf{x};t). \tag{4.12}$$

Since now states in the different $\mathcal{H}_t$ slices are given independent phases, it is clear that one cannot linearly superimpose such states. More about this superselection rule will be said in Sec. 5.

Using the realization $H \sim i\partial_t$ we now see that, under a *time dependent* phase transformation, H is no longer invariant; we rather have

$$H \to H + \partial_t\omega. \tag{4.13}$$

(The behavior of P, Q, J is not affected by making ω time dependent.) It then follows that Eq. (4.2) is no longer consistent: Under a time dependent gauge transformation the lhs transforms according to (4.13) but the rhs is unchanged. Consistency between the overall realization $H \sim i\partial_t$ in $\mathcal{H}$ and its realization on any slice $\mathcal{H}_t$ can be restored if we restrict the so far arbitrary V to fields which, under a general local phase transformation transform as

$$V \to V + \partial_t\omega. \tag{4.14}$$

This completes the characterization of covariantly interacting systems in our framework.

Remarks: (a) The Schrödinger equation, i.e., the realization of (4.2), is now invariant under *general* gauge transformations, i.e., under the simultaneous replacements

$$\psi(\mathbf{x};t) \to e^{-i\omega(\mathbf{x},t)}\psi(\mathbf{x};t), \quad A_k \to A_k - \partial_k\omega, \quad V \to V + \partial_t\omega.$$

(b) If we calculate the force $\dot{\mathcal{P}}$ with (4.8), (4.9), (4.10), we find that[33]

$$\dot{\mathcal{P}} = M^{-1}\mathcal{P} \times \mathrm{curl}\mathbf{A} - \mathrm{grad}V - \partial_t\mathbf{A}, \tag{4.15}$$

which is the familiar Lorentz force. It is invariant under general gauge transformations.

5. SUPERSELECTION RULES AND GAUGE INVARIANCE

When, in Sec. 2, we demanded that local phase transformations with $\omega(\mathbf{x})$ be (unitarily implementable) kinematic symmetries, we were led to the well-known mass-superselection rule [cf. Eq. (2.8)]. Another superselection rule arises when we demand that $A_k \to A_k - \partial_k\omega(\mathbf{x})$ be a kinematical symmetry. To see this, take the particular phase $\omega(\mathbf{x}) = c_l x_l$. Denote that part[34] of the corresponding unitary operator which acts on functions $\mathcal{R}$ of $\mathbf{A}$ (and V), by $\mathcal{U} = \exp(iN)$ and write $N = M^{-1}c_l\mathcal{K}_l$ ($\mathcal{K}_l$ is dimensionless). Then one has

$$A_k - c_k = \mathcal{U}A_k\mathcal{U}^{-1} = A_k + iM^{-1}c_l[\mathcal{K}_l, A_k] + \cdots.$$

This satisfied if

$$[\mathcal{K}_l, A_k] = iM\delta_{lk} \tag{5.1}$$

and we also have $[\mathcal{K}_l, \mathcal{K}_k] = 0$. Since the rhs of (5.1) is in the center of the algebra, we indeed have a new superselection rule for the complete system generated by $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{J}$, H, $\mathbf{A}$, V. On the space of functionals $\mathcal{R}$, the generator $\mathcal{K}_l$ can be realized as

$$\mathcal{K}_l \sim iM\frac{\delta}{\delta A_l}. \tag{5.2}$$

It is easy to see that this new superselection rule corresponds to electric charge. If, for the moment, we use conventional dimensionate units, then in Eq. (4.2) $\mathbf{A}$ must be replaced by $e\hat{\mathbf{A}}$ (and V by $e\hat{V}$). In these units, (5.1) gives

$$[\mathcal{K}_l, \hat{A}_k] = i(M/e)\delta_{lk}, \tag{5.3}$$

which shows that the "supersymmetry" observable is the (reciprocal of) specific charge.

It is rather remarkable that the mass and charge superselection rules follow from one and the same principle, viz., that local phase transformations with at least[35] a linear $\omega = c_l x_l$ be kinematical symmetries for the interacting system. Actually, something more holds. If we define the "gauge operator" $\mathbf{R}$ to be the generator of gauge transformations with $\omega(\mathbf{x}) = c_l x_l$ for the entire system, i.e., if we set

$$\mathbf{R} = \mathbf{Q} + \mathcal{K}, \tag{5.4}$$

then, from (4.2), (4.6), (5.1) we see that, as expected, $\mathbf{R}$ is a constant of motion:

$$[H, \mathbf{R}] = 0. \tag{5.5}$$

We now turn to another topic. When, in Sec. 4, we stipulated (via Assumption 8) that time dependent phase transformations be also kinematical symmetries, we introduced an explicit superselection rule. Let us consider the particular linear phase $\omega(\mathbf{x}, t) = kt$ (k const); denote the corresponding unitary operator which acts on H (and $\mathbf{P}, \mathbf{Q}, \mathbf{J}$) by $\mathcal{U} = \exp(iD)$ with $D = M^{-1}kW$ (W dimensionless). Then we have, because of (4.13),

$$H + k = \mathcal{U}H\mathcal{U}^{-1} = H + iM^{-1}k[W, H] + \cdots.$$

This implies that

$$[W, H] = -iM, \tag{5.6}$$

so that the presence of W among observables gives rise to a superselection rule. If we revert to conventional, dimensionate units, we get

$$[W, \hat{H}] = -i\hbar\hat{M}/\theta, \tag{5.7}$$

where the constant θ has the dimension of time. Thus, we are led to a superselection rule for the "time operator" (or rather, the reciprocal of "specific time"). From considerations concerning the measuring process, Piron[36] also arrived at the conclusion that, in the standard Schrödinger representation, one has a superselection rule for time. But his assumption was that time (as measured by clocks which are isolated from the physical system) is a physical observable, not a parameter. It is interesting that, without such an assumption, gauge theory leads to the same conclusion.

It may be worth while to point out that, as (5.6) shows, the observable canonically conjugate to H is the generator W of phase transformations with $\omega = kt$. Since H is realized on $\mathcal{H}$ by $i\partial_t$, one may realize W by Mt. It would be, however, incorrect to say that W is the "time operator": The supersymmetry operator is on the rhs of (5.6).

Since, whenever a time dependent phase transformation is performed, we must also transform $V \to V + \partial_t\omega$, it is necessary to include in the system of observables an operator which acts on the functionals $\mathcal{R}$ of V (and $\mathbf{A}$) and which generates the change $V \to V + k$ when $\omega = kt$. We write $\mathcal{U} = \exp(iZ)$ with $Z = M^{-1}k\mathcal{G}$, and find that

$$[\mathcal{G}, V] = -iM. \tag{5.8}$$

Thus, we have a superselection rule. The observable $\mathcal{G}$ which gives rise to it may be realized as

$$\mathcal{G} \sim -iM\frac{\delta}{\delta V}. \tag{5.9}$$

Clearly, $\mathcal{G}$ is the couterpart of W, just as $\mathcal{K}$ was the counterpart of $\mathbf{Q}$. However, we did not now obtain a new supersymmetry. Indeed, using conventional units, (5.8) becomes

$$[\mathcal{G}, \hat{V}] = -iM/e, \tag{5.10}$$

so that the superselection rule is for the (reciprocal of the) specific charge, i.e., the same supersymmetry that was the result of having included the observable $\mathcal{K}$ into the algebra.

We may summarize as follows: All superselection rules are brought about by the pair $\mathbf{Q}$, $\mathcal{K}$ (giving mass and specific charge superselection) and by the pair W, $\mathcal{G}$ (giving specific time and specific charge superselection). The operator $\mathbf{Q} + \mathcal{K}$ is the generator for gauge transformations with $\omega = c_l x_l$, and the operator $W + \mathcal{G}$ is the generator for gauge transformations with $\omega = kt$.

It is possible to slightly modify (and perhaps simplify) the description of the system. Let us perform a gauge transformation with

$$\omega(\mathbf{x}, t) = -\int_0^t V\,dt.$$

Then

$$H \to H - V \equiv \tilde{H}, \quad \mathcal{P} \to \mathcal{P} \equiv \tilde{\mathcal{P}}, \tag{5.11}$$

and Eq. (4.10) becomes, in this particular gauge,

$$\tilde{H} = (1/2M)\tilde{\mathcal{P}}^2. \tag{5.12}$$

Thus, $\tilde{H}$ is the energy, but as seen from (5.11), it is no longer the temporal time displacement operator when realized on $\mathcal{H}$. Once we chose this particular gauge, we are no longer permitted to perform gauge transformations with time dependent ω. Therefore, there is no need to have W and $\mathcal{G}$ in the algebra of observables. Hence, the "time superselection rule" disappears.[37] Furthermore, in this gauge, gauge transformations are an invariance transformation of the system, since $\tilde{H} \to \tilde{H}$.

6. CONCLUDING REMARKS

The major result of this study is the demonstration of the power of the locality principle. Combined with a few, generally accepted and rather weak requirements, it leads to an algebraic structure which can be identified with the Galilei group. In particular, Galilean boosts appear as the simplest, nontrivial gauge transformations. Adding the requirement that localization be gauge independent, one is led to a unique interaction structure. Light is shed on the peculiar relation between various superselection rules.

One may consider the algebraic structure that emerges if the locality principle is generalized to non-Abelian transformations. However, there does not seem to be much point in endowing nonrelativistic particles with internal symmetries [such as $SU(3)$].

It is obvious that if one starts, instead of the Euclidean space, with the Minkowski space of events, one will be led to a relativistic Galilean structure. The relativistic generalization of the Galilei group has been introduced by one of us a few years ago,[38] starting from very different arguments. We plan to pursue the present approach in the relativistic framework in the future.

APPENDIX A: REPRESENTATION ON THE TANGENT SPACE

The elements of the left coset space $\mathcal{G}/SO(3)^J$ can be represented by triples $(\bar{\tau}, \bar{\mathbf{a}}, \bar{\mathbf{v}})$. The left action of $\mathcal{G}$ gives

$$(\bar{\tau}, \bar{\mathbf{a}}, \bar{\mathbf{v}}) \to (\bar{\tau}+\tau,\ R\bar{\mathbf{a}}+\mathbf{a}+\bar{\tau}\mathbf{v},\ R\bar{\mathbf{v}}+\mathbf{v}). \tag{A1}$$

We introduce a new parameter ϵ by setting

$$-\bar{\tau}\mathbf{v}=\epsilon\bar{\mathbf{v}}.$$

At $\bar{\tau}' \equiv \bar{\tau}+\tau=0$, this is an invertible transformation, $\tau\mathbf{v}=\epsilon\bar{\mathbf{v}}$, and then (A1) becomes

$$(\bar{\tau}, \bar{\mathbf{a}}, \bar{\mathbf{v}})_{\bar{\tau}=-\tau} \to (\bar{\tau}+\tau, R\bar{\mathbf{a}}+\mathbf{a}-\epsilon\bar{\mathbf{v}}, R\bar{\mathbf{v}}+\mathbf{v})_{\bar{\tau}=-\tau}. \tag{A2}$$

If we identify the space $(\bar{\tau}, \bar{\mathbf{a}}, \bar{\mathbf{v}})_{\bar{\tau}=-\tau}$ with the phase space (or rather tangent space) $E_3(\mathbf{x})\times E_3(\mathbf{p})$ by the map $(\bar{\mathbf{a}}, \bar{\mathbf{v}}) \to (\mathbf{x}, \mathbf{p})$, then (A2) gives

$$\mathbf{x} \to R\mathbf{x}+\mathbf{a}-\epsilon\mathbf{p}, \quad \mathbf{p} \to R\mathbf{p}+\mathbf{v}. \tag{A3}$$

In Ref. 26 it has been explicitly shown that this transformation group is isomorphic to the Galilei group as defined on $E_3(\mathbf{x})\times E_1(t)$.

APPENDIX B: INTERACTING SYSTEM WITH SPIN

The matrix realization $\mathbf{T}=\Sigma$ of spin is valid on every slice $\mathcal{H}_t$ as long as no interactions are present. In the opposite case, however, this cannot hold, because, while we still have $[H, \mathbf{J}]=0$, we cannot have $[H, \mathbf{Q}\times\mathbf{P}]=0$, and, hence, $[H, \mathbf{T}]\neq 0$. Thus, the general form of $\mathbf{T}$ will be

$$T_k=\Sigma_k+tf_k(\mathbf{A}, V). \tag{B1}$$

To determine f_k, we make the following assumptions:

(i) T must still be an axial vector,[39]
(ii) No arbitrary constant shall be introduced,
(iii) T be linear in the fields,[40]
(iv) Spin does not depend on the choice of a phase $\omega(\mathbf{x})$.

The last requirement is on the same footing as Assumption 7 in Sec. 4 which was imposed on localization, and it in effect assures us that spin is a gauge invariant concept.

A moment's consideration shows that the above requirements determine f_k to be

$$f_k=(1/2M)\epsilon_{klj}\Sigma_l B_j,$$

where $B_j \equiv \epsilon_{jab}\partial_a A_b$ is the magnetic field.[41] Thus,

$$T_k=\Sigma_k+(1/2M)t\epsilon_{klj}\Sigma_l B_j, \tag{B2}$$

and, using the realization $H \sim i\partial_t$, we see that

$$[H, T_k]=(i/2M)\epsilon_{klj}\Sigma_l B_j. \tag{B3}$$

In order to find what term H' we must add to our $H=(1/2M)p^2+V$, we go to the slice $\mathcal{H}_{t=0}$, where (B3) gives

$$[\bar{H}, \Sigma_k]=(1/2M)i\epsilon_{klj}\Sigma_l\bar{B}_j.$$

Since $\bar{p}^2$ and $\bar{V}$ commute with Σ_k, this equation can be satisfied if we put $\bar{H}'=(2M)^{-1}\Sigma_n\bar{B}_n$. Indeed,

$$[\Sigma_n\bar{B}_n, \Sigma_k]=i\epsilon_{klj}\Sigma_l\bar{B}_j.$$

Transformforming back to slice $\mathcal{H}_t$, we thus finally have

$$H=(1/2M)p^2+(1/2M)\Sigma_l B_l+V. \tag{B4}$$

This interesting form of interaction was also found by Lévy-Leblond,[10] who derived it from multicomponent Galilean covariant wave equations with arbitrary spin. The major feature is the correct gyromagnetic ratio and the absence of electric moments. It thus appears that locality arguments to some extent at least incorporate significant predictions of detailed wave equations.

In conclusion we note that the presence of the magnetic dipole moment interaction term in (B4) does not affect the equation of motion $[H, Q_k]=-ip_k$.

[1] H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, Leipzig, 1931) [English transl. by H.P. Robertson (Dover, New York, 1950)].
[2] C.N. Yang and R. Mills, Phys. Rev. **96**, 191 (1954); R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[3] A comprehensive survey, covering developments up to the beginning of 1973, was given by E.S. Abers and B.W. Lee, Phys. Rep. **9**, 1 (1973). A review of some subsequent developments can be found, for example, in "Recent Progress in Gauge Theories", S. Weinberg, Harvard Univ. preprint, 1974.
[4] E.P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **51**, 956 (1964); in *Proceedings of the International School of Physics Enrico Fermi*, Course 29 (Academic, New York, 1964); Nobel Prize Lecture 1964. [All these articles are reprinted in *Symmetries and Reflections* (Indiana U.P., Bloomington, Indiana, 1967]).
[5] R.M.F. Houtappel, H. van Dam, and E.P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).
[6] Of course, Galilei invariant equations of motion can be obtained also for velocity dependent forces provided these

forces *are* Galilei invariant, cf. Eq. (3.2) of Ref. 5. However, in this case Galilei invariance was put in "by hand".

[7]Cf. J.-M. Lévy-Leblond, in *Group Theory and Its Applications*, Vol. II, edited by E.M. Loebl (Academic, New York, 1971). See also J. Voisin, J. Math Phys. **6**, 1519 (1965). Voisin gives another but equivalent f.

[8]On the other hand, if one considers *two interacting particles* acting on each other through a potential V which depends only on the length of the relative instantaneous coordinate, the c.m. motion can be separated off and the Schrödinger equation for the relative motion is still Galilei invariant as expected. But for our present discussion, this is not relevant.

[9]It suffices to choose a gauge function subject to the equations $\omega=\int_0^t[V(\mathbf{x}')-V(\mathbf{x})]\,dt+s(\mathbf{x})$ with $\partial_k s=A_k(\mathbf{x})-A_k(\mathbf{x}')-\partial_k\int_0^t[V(\mathbf{x}')-V(\mathbf{x})]\,dt$, where $\mathbf{x}'=\mathbf{x}+\mathbf{v}t$.

[10]J.-M. Levy-Leblond, Commun. Math. Phys. **6**, 286 (1967). See also *ibid.*, **4**, 157 (1967), and Ref. 7. A more complete study was done later by C.R. Hagen, Commun. Math. Phys. **18**, 97 (1970). See also C.R. Hagen and W.J. Hurley, Phys. Rev. Lett. **24**, 1381 (1970), and W.J. Hurley, Phys. Rev. D **3**, 2339 (1971).

[11]J.M. Jauch, Helv. Phys. Acta **37**, 284 (1964).

[12]Kinematical symmetries are a subset of canonical transformations. In turn, invariances (represented by unitary operators that commute with the Hamiltonian) are a subset of kinematical symmetries.

[13]The velocity is *defined* by $\dot{Q}_k=i[H,Q_k]$.

[14]C. Piron, Found. Phys. **2**, 287 (1972).

[15]J.-M. Lévy-Leblond, Thèse de Doctorat, Orsay, 1965. In (Ann.) Phys. (N.Y.) **57**, 481 (1970) he also considered a possible generalization of Jauch's work for relativistic systems. See also S.K. Wong, Nuovo Cim. **4 B**, 300 (1971).

[16]E.P. Wigner, *Gruppentheorie und ihre Anwendung* (Vieweg, Braunschweig, 1931) [English transl. by J.J. Griffin (Academic, New York, 1959)].

[17]Summation over l from 1 to 3 is understood.

[18]The arbitrary constant M^{-1} is inserted for dimensional reasons. Since c_l has dimension of $(\text{length})^{-1}$, and F is dimensionless, we can make the Q_l dimensionless if M^{-1} has the dimension of length. Since we use units in which $\hbar=c=1$, we can say that M has the dimension of mass.

[19]In this paper, $\times$ stands for direct product and $\otimes$ for semidirect product.

[20]$SU(2)$ arises as the universal convering of $SO(3)$. T_3 and T_1 denote three- and one-dimensional Abelian groups.

[21]Since in every irreducible representation $\mathbf{T}^2$ is a multiple of the identity, we can disregard it in (3.1).

[22]The existence of this equivalence relation is essentially tantamount to Assumption 4 and could be used to replace it.

[23]H cannot depend on **TP** because, for example, $[Q_k,\mathbf{TP}]=iMT_k$ which is not in the Lie albegra. Dependence on higher powers of $\mathbf{P}^2$ is excluded similarly.

[24]In $\mathcal{G}$ the $SU(2)^J$ part of G should be thought of as $SO(3)^J$. **P** and **Q** commute.

[25]I.e., the cosets of the subgroup $SO(3)^J\otimes T_3^Q$.

[26]This representation of $\mathcal{G}$ was introduced, from another point of view, by P. Roman, J.J. Aghassi, and P.L. Huddleston, J. Math. Phys. **13**, 1852 (1972).

[27]Because representations with different $\mathcal{C}_1$ are ray-equivalent, one might take $\mathcal{C}_1=0$.

[28]Presently we still confine ourselves to phase transformations with a t-independent ω, that is, we apply the locality postulate on each "slice" $\mathcal{H}_t$ simultaneoulsy. Assumption 2 has not yet been extended.

[29]Its generator is a function of $Q(t), P(t)\equiv\mathbf{P}$, $\mathbf{J}(t)$, H and of derivatives of $\mathbf{Q}(t)$. The latter are, of course, defined by $i[H,\mathbf{Q}(t)]$

[30]Recall that the boosts **Q** are precisely the generators of particular, linear phase transformations.

[31]In writing (4.4), two tacit assumption entered: (a) When the "interaction is switched off", we must recover (3.10); (b) when considering **Q** on the slice $t=0$, we must recover (2.10). For simplicity, we now also restrict ourselves to the spinless case. Systems with spin are discussed in Appendix B.

[32]This well-known definition follows from the fact that, because of (3.9), $\Omega=e^{itH}\overline{\Omega}e^{-itH}$, where $\overline{\Omega}$ denotes the observable on the slice $\mathcal{H}_{t=0}$.

[33]One must use the relations $[P_k,\Omega(\mathbf{Q})]=-iM\partial_k\Omega$ which are valid for any power series Ω of **Q**.

[34]The part which acts on functions of **P**, **Q**, **J**, H is, of course, $\mathcal{U}=\exp(iF)$ with $F=M^{-1}c_lQ_l$, cf. Theorem 1.

[35]This is the place to point out that once we include the generator $\mathcal{K}$ among our observables, we can accomodate the kinematic symmetry $A_k\to A_k-\partial_k\omega$ for arbitrary $\omega(\mathbf{x})$, just as was the case for the part which was a kinematic symmetry on the **P**, **Q**, **J**, H variables. As we saw in Sec. II, this is the consequence of the fact that $\omega(\mathbf{x})$ has a power series expansion. Similar remarks hold for the kinematic symmetries to be discussed in the rest of this section.

[36]C. Piron, Helv. Phys. Acta **42**, 330 (1969).

[37]Naturally, the slices $\mathcal{H}_t$ are still incoherent, by their very definition, but no supersymmetry operator is explicitly present. $\mathcal{H}_t$ is now merely a parameter.

[38]J.J. Aghassi, P. Roman, and R.M. Santilli, Phys. Rev. D **1**, 2753 (1970); J. Math. Phys. **11**, 2297 (1970); Nuovo Cimento **5 A**, 551 (1971); R.M. Santilli, Particles Nuclei **1**, 81 (1970). See also P.L. Huddleston, M. Lorente, and P. Roman, Preprint BU-PNS-5, Boston University, 1973.

[39]The algebra of noninteracting observables admits reflections as an outer automorphism, which justifies the requirement that this feature persist when interactions are on. We do not require, however, parity conservation.

[40]This is a simplicity assumption which we cannot further justify.

[41]The factor $(2M)^{-1}$ arises because of dimensional reasons, and because in H the field **A** occurs only with a factor $(2M)^{-1}$, so that, if no arbitrary constant is to be used, we must consider $(2M)^{-1}\mathbf{A}$ as the basic entity.

Proper particle mechanics

David Hestenes

Physics Department, Arizona State University, Tempe, Arizona 85281
(Received 1 April 1974)

Spacetime algebra is employed to formulate classical relativistic mechanics without coordinates. Observers are treated on the same footing as other physical systems. The kinematics of a rigid body are expressed in spinor form and the Thomas precession is derived.

INTRODUCTION

This paper shows how to formulate conventional relativistic mechanics without refering to observers or coordinates. To emphasize the distinctive features of this formulation, it will be called "*proper* mechanics." The common expression "relativistic mechanics" will be avoided here because, by the most straightforward interpretation of the adjective "relativistic," Einstein's mechanics is less rather than more relativistic than the socalled "nonrelativistic" mechanics of Newton. The equations describing a particle in Newtonian mechanics depend on the motion of the particle *relative* to the observer; in Einsteinian mechanics they do not. Einstein originally formulated his mechanics in terms of "relative variables" (such as the position and velocity of a particle relative to a given observer), but he eliminated dependence of the equations on the observer's motion by the "relativity postulate," which requires that the equations be invariant under a change of relative variables from those of one inertial observer to those of another. Minkowski's covariant formulation of Einstein's theory replaced the explicit use of variables relative to inertial observers by components relative to an arbitrary coordinate system in spacetime. The "proper formulation" used here relates particle motion directly to Minkowski's "absolute spacetime" without the intermediary of a coordinate system.

Minkowski had the great idea of interpreting Einstein's theory of relativity as a prescription for fusing space and time into a single entity "spacetime." The straightforward algebraic characterization of "Minkowski spacetime" by "spacetime algebra" makes a proper formulation of mechanics possible. The space-time algebra can be regarded as a variant of the Dirac algebra more intimately related to spacetime than the usual matrix version. Proper mechanics shows how generally useful the Dirac algebra is outside its usual domain of "relativistic quantum theory." Besides providing a simple proper formulation of all the usual equations in "classical relativistic mechanics," spacetime algebra brings spinors to bear on the subject; as will be shown, this simplifies many things and brings the subject closer, in its formulation, to quantum theory.

In Sec. 1, the spacetime algebra is introduced along with important notations needed to interpret and apply it efficiently. For later use a number of important algebraic identities are set down and the spinor formulation of Lorentz rotations is discussed.

In Sec. 2, the proper description of a material particle is given. Inertial observers are introduced on the same footing as other physical systems; the distinction between proper and relative vectors is explained, and the reformulation of proper quantities in terms of relative variables is carried out in detail.

In Sec. 3, the relation of the spacetime algebra to the Dirac and Pauli matrix algebras is briefly explained. It is shown how easily the usual covariant equations can be put in proper form and vice versa. This section pertains only to the relation of spacetime algebra to other mathematical systems, and it is not needed in the rest of the paper.

In Sec. 4, a comoving frame is associated with a particle, and its kinematics are completely described in spinor form. This gives immediately a complete and simple formulation of the kinematics of a "rigid point particle" (i.e., a rigid body of negligible dimensions). In particular, the Thomas precession is derived by a new (and hopefully, a clearer and simpler) method, along with a complete treatment of related kinematical results. A great advantage of this approach is that all results can be used directly in an analysis of Thomas precession in the Dirac electron theory, as will be demonstrated elsewhere.

1. SPACETIME ALGEBRA

In this paper *spacetime* is understood to be four-dimensional continuum (or manifold) with "Minkowski metric" of signature minus two. Spacetime derives its significance from the facts (or, hypotheses, if you will) that every elementary physical event can be uniquely labelled by a point of spacetime, and that the metric of spacetime determines a unique ordering of physical events.

Spacetime can be given a precise mathematical description by introducing appropriate rules for adding and multiplying points. A vector a is said to be *tangent to a point x in spacetime* if there is a curve $\{x(\alpha);\ 0<\alpha<\epsilon\}$ in spacetime extending from the point $x(0)=x$ such that

$$a=\lim_{\alpha\to 0}\alpha^{-1}(x(\alpha)-x). \tag{1.1}$$

The right side of (1.1) is made meaningful by the assumption that the points of spacetime can be added and multiplied by scalars according to the usual rules associated with vectors. However, it should be noted that the validity of (1.1) does not require that the sum of two spacetime points or the scalar multiple of one is again a spacetime point, in short, the spacetime is a linear vector space.

The set of all vectors tangent to a typical spacetime point x is a four-dimensional vector space $\mathcal{T}(x)$ called the *tangent space* at x. An element of such a space will sometimes be called a proper vector to avoid possible confusion with other uses of the word vector. By multiplication and addition the elements of $\mathcal{T}(x)$ generate a

 Copyright © 1974 American Institute of Physics

noncommutative associative algebra called the *space—time algebra* (at x). This algebra has been systematically discussed in Ref. 1 and since developed into a more extensive mathematical system especially by Ref. 2. However, the basic multiplication law of spacetime algebra is likely to be familiar to most readers only in the guise of the Dirac matrix algebra, so a sketchy review of the algebra is necessary to establish terminology and a few basic relations. Relation of the spacetime algebra to more familiar formalisms will be discussed in Sec. 3.

The *geometric product* of a generic proper vector a with itself is a scalar quantity describing the metric of spacetime; thus,

$$a^2>0 \text{ iff } a \text{ is a } timelike \text{ vector;} \tag{1.1a}$$

$$a^2=0 \text{ iff } a \text{ is zero or a } lightlike \text{ vector;} \tag{1.1b}$$

$$a^2<0 \text{ iff } a \text{ is a } spacelike \text{ vector.} \tag{1.1c}$$

The term "scalar" here always means "real number," The geometric product ab of proper vectors a and b can be decomposed into a a sum of commuting and anticommuting parts; thus,

$$ab=a\cdot b+a\wedge b, \tag{1.2a}$$

where

$$a\cdot b\equiv\tfrac{1}{2}(ab+ba)=b\cdot a, \tag{1.2b}$$

$$a\wedge b\equiv\tfrac{1}{2}[a,b]=-b\wedge a, \tag{1.2c}$$

and $[A,B]\equiv AB-BA$. It follows from (1.1) that $a\cdot b$ is a scalar quantity, the usual *inner product* of spacetime vectors. The quantity $a\wedge b$, called the *outer product* of a and b, is a (proper) *bivector* (or 2-vector).

Bivectors are related to vectors by multiplication. A bivector which can be expressed, as in (1.2c), as the outer product of two vectors is said to be *simple*. A bivector B in the spacetime algebra can be uniquely that every null bivector is simple and, in fact, has a null vector as a factor. Furthermore, every nonnull bivector B in the space—time algebra can be uniquely expressed as the sum of two simple bivectors or *blades*; that is, there exist unique blades B_1 and B_2 such that

$$B=B_1+B_2, \tag{1.3}$$

and B_1B_2 is a *pseudoscalar*, or equivalently B_2 is proportional to the *dual* of B_1. (The meanings of the terms "pseudoscalar" and "dual" will be explained later.)

The *inner* and *outer product* of a vector a with a bivector B can be defined respectively by

$$a\cdot B\equiv\tfrac{1}{2}[a,B]=-B\cdot a, \tag{1.4a}$$

and

$$a\wedge B\equiv\tfrac{1}{2}(aB+Ba)=B\wedge a, \tag{1.4b}$$

so

$$aB=a\cdot B+a\wedge B. \tag{1.4c}$$

Using (1.4a) together with (1.2b) and (1.2c), it is easy to prove that any three vectors a,b,c satisfy the useful identity

$$a\cdot(b\wedge c)=a\cdot bc-a\cdot cb=-(b\wedge c)\cdot a. \tag{1.5}$$

It follows that the quantity $a\cdot B$ defined by (1.4a) is a vector. On the other hand, the quantity $a\wedge B$ is a *trivector* (or 3-vector). Using (1.4c) and (1.2c), one can show that the outer product of vectors is associative, that is

$$(a\wedge b)\wedge c=a\wedge(b\wedge c)=a\wedge b\wedge c. \tag{1.6}$$

Every trivector in the spacetime algebra can be factored (but not uniquely) into an outer product of three vectors.

It is well at this point to introduce the convention that when parentheses are omitted inner and outer products have priority over the geometric product; for example, for vectors, a,b,c,d,

$$(a\cdot b)c=a\cdot bc\neq a\cdot(bc),$$

$$(a\wedge b)c=a\wedge bc\neq a\wedge(bc),$$

$$a\cdot bc\wedge d=(a\cdot b)(c\wedge d).$$

This convention is particularly useful in complicated formulas. It has already been used in (1.5).

The product of a vector a with a trivector T is the sum of a bivector $a\cdot T$ and a 4-*vector* or *pseudoscalar* $a\wedge T$; thus,

$$aT=a\cdot T+a\wedge T, \tag{1.7a}$$

$$a\cdot T\equiv\tfrac{1}{2}(aT+Ta)=T\cdot a, \tag{1.7b}$$

$$a\wedge T\equiv\tfrac{1}{2}(aT-Ta)=-T\wedge a, \tag{1.7c}$$

From (1.7b), (1.4a), and (1.2) one can establish the useful identity

$$a\cdot(b\wedge B)=a\cdot bB-b\wedge(a\cdot B), \tag{1.8}$$

where a,b are vectors and B is a bivector. Every pseudoscalar is a scalar multiple of a unique unit pseudoscalar which will always be denoted by i. Specification of i assigns an orientation to spacetime. It can be shown that

$$i^2=-1, \tag{1.9a}$$

and the geometric product of i with any vector a is anticommutative; that is,

$$ai=-ia. \tag{1.9b}$$

It follows that the outer product $a\wedge i\equiv\frac{1}{2}(ai+ia)$ vanishes, while the inner product $a\cdot i\equiv\frac{1}{2}(ai-ia)=ai$ is a trivector (called the *dual* of a). Every trivector T is the dual of some vector t, that is, $T=ti$. By (1.9a), $Ti=-t$, so the *dual* Ti of any trivector T is a unique vector. This establishes an isomorphism of the linear space of all trivectors to the space of all vectors. For this reason, trivectors are often called *pseudovectors*.

A generic element of the space—time algebra will be called a (*proper*) *multivector*. Every proper multivector M can be uniquely expressed as a sum of a 0-vector (or scalar), a 1-vector (or vector), a 2-vector (or bivector), a 3-vector (or pseudovector), and a 4-vector (or pseudoscalar); that is

$$M=[M]_0+[M]_1+[M]_2+[M]_3+[M]_4, \tag{1.10}$$

where $[M]_k$ denotes the k-vector part of M. A multivector M is said to be *even* if $[M]_1=[M]_3=0$. The even multivectors compose an important subalgebra of the full spacetime algebra.

The *reverse* $\tilde{M}$ of a multivector M can be defined by the equation

$$\tilde{M} = [M]_0 + [M]_1 - [M]_2 - [M]_3 + [M]_4. \tag{1.11}$$

It can then be shown that the reverse of a product equals the product of reverses, that is, if

$$M = AB, \quad \text{then } \tilde{M} = \tilde{B}\tilde{A}. \tag{1.12}$$

Spacetime algebra makes it possible to describe Lorentz transformations completely, without resorting to coordinates or matrices. Only Lorentz rotations (i.e., Lorentz transformations without time reversal or space inversion) are of interest here. Any *Lorentz rotation* $\mathcal{R}$ which maps a generic proper vector a into the vector $a' = \mathcal{R}(a)$ can be written in the canonical form

$$a' = \mathcal{R}(a) = Ra\tilde{R}; \tag{1.13a}$$

here R is an *even* multivector, unique except for sign, with the property

$$R\tilde{R} = 1. \tag{1.13b}$$

The multivector R is called a *spinor*. One way to establish (1.13) is to introduce an orthonormal frame of vectors γ_μ and its "reciprocal frame" $\{\gamma^\mu\}$ defined by the equations

$$\gamma^\mu \cdot \gamma_\nu = \delta^\mu_\nu, \quad \mu, \nu = 0, 1, 2, 3 \tag{1.14}$$

where δ^μ_ν is the "unit matrix." According to (1.13a) the transformation of γ_μ is given by

$$\gamma'_\mu = R\gamma_\mu\tilde{R} = a^\nu_\mu \gamma_\nu \tag{1.15}$$

(sum over repeated indices), where $a^\nu_\mu = \gamma^\nu \cdot \gamma'_\mu$ is the matrix of the transformation. These equations can be solved for R. One obtains (see Sec. 17 of Ref. 1)

$$R = \pm (\tilde{A}A)^{-1/2}A \quad \text{where } A \equiv \gamma'_\mu \gamma^\mu = a^\nu_\mu \gamma_\nu \gamma^\mu. \tag{1.16}$$

This gives R explicitly as a function of the matrix a^ν_μ, but it is of little practical use since in most applications it is easier to determine R directly from the data.

Two special classes of Lorentz rotations are of interest here, boosts and spatial rotations. A Lorentz rotation $\mathcal{L}(a) = La\tilde{L}$ which takes a unit *timelike* vector u into the vector v is said to be a *boost of u into v* if it leaves vectors orthogonal to the $v \wedge u$-plane invariant. Any vector a can be expressed as the sum of a component $a_\parallel$ in the $v \wedge u$-plane and a component $a_\perp$ orthogonal to it; thus,

$$a = a_\parallel + a_\perp \tag{1.17a}$$

where

$$a_\parallel = a \cdot (v \wedge u)(v \wedge u)^{-1}, \tag{1.17b}$$

$$a_\perp = a \quad (v \wedge u)(v \wedge u)^{-1}. \tag{1.17c}$$

By definition

$$La_\perp\tilde{L} = a_\perp \quad \text{so } La_\perp = a_\perp L, \tag{1.18a}$$

because $L\tilde{L} = 1$. It can further be shown that

$$La_\parallel\tilde{L} = L^2 a_\parallel \quad \text{or } La_\parallel = a_\parallel \tilde{L}; \tag{1.18b}$$

in particular,

$$v = Lu\tilde{L} = L^2 u \quad \text{so } L^2 = vu. \tag{1.18c}$$

The square root in (1.18c) can be taken to give L explicitly in terms of v and u [Eq. (18.14) of Ref. 1], but the result is unduly complicated and can be avoided in applications by using (1.18).

A Lorentz rotation $\mathcal{U}(a) = Ua\tilde{U}$ said to be a *spatial rotation* if it leaves a timelike vector u invariant; that is, if

$$Uu\tilde{U} = u, \tag{1.19a}$$

or, equivalently,

$$UU^\dagger = 1 \quad \text{where } U^\dagger \equiv u\tilde{U}u. \tag{1.19b}$$

The set of all Lorentz rotations satisfying (1.19) is the group of spatial rotations in the space—like hypersurface with normal u, called the *little group* of u.

Any Lorentz rotation can be uniquely expressed as a spatial rotation followed by a boost of a given timelike vector u. This decomposition can be completely characterized by factoring the spinor R defined by (1.13) into the form

$$R = LU, \tag{1.20}$$

where L and U are defined by (1.18) and (1.19), respectively.

The spacetime algebra associated with a single spacetime point has been discussed. If spacetime is geometrically flat, then, with one point chosen as the zero vector, it is identical with the tangent space at each of its points. In this case there is only one space—time algebra, and the spacetime points have all the properties of proper vectors mentioned above.

In the rest of this paper spacetime will be assumed geometrically flat. However, the basic ideas and most of the results apply with little or no modification to curved spacetime. To make such applications easier in the future, the definition of proper multivectors has been given in greater generality than is needed in this paper. The mathematical apparatus needed to apply spacetime algebra to curved spacetime is developed in Refs. 1 and 2.

2. THE PROPER POINT OF VIEW

The *history* of a material particle is a timelike curve $x = x(\tau)$ in spacetime. Particle conservation is expressed by assuming that the function $x = x(\tau)$ is single-valued and continuous, except at discrete points where particle creation and/or annihilation occurs. Only differentiable particle histories will be considered here, and τ will always refer to the *proper time* (arc length) of a particle history. After a unit of length (say centimeters) has been chosen, the physical significance of the spacetime metric is fixed by the assumption that the proper time of a material particle is equal to the time (in centimeters) recorded on a material clock traveling with the particle.

The unit tangent $v = v(\tau) = dx/d\tau \equiv \dot{x}$ of a particle history will be called the *(proper) velocity* of the particle. By the definition of proper time, $d\tau = |dx| = |(dx)^2|^{1/2}$, and

$$v^2 = 1. \tag{2.1}$$

The term "proper velocity" is to be preferred to the alternative terms "absolute velocity," "world velocity," "invariant velocity," and "four velocity." The adjective "proper" is used to emphasize that the velocity v describes an intrinsic property of the particle, independent of any observer or coordinate system. The adjective "absolute" would do the same, but it may not be free from undesirable connotations. Moreover, the word "proper" is shorter and has already been used in the same sense in the terms "proper mass" and "proper time." The adjective "invariant" is inappropriate, because no transformation group has been introduced. The velocity will not be called a "4-vector" because that term already means pseudoscalar in spacetime algebra; besides, there is no need to refer to any four components of the velocity.

The quantity $dv/d\tau \equiv \dot{v} = \ddot{x}$ will be called the (*proper*) *acceleration* of the particle. The constraint (2.1) implies that $\dot{v}$ is orthogonal to v, that is

$$\dot{v}\cdot v = 0, \tag{2.2a}$$

or, equivalently, by virtue of (1.2a),

$$\dot{v}v = \dot{v}\wedge v = -v\dot{v}. \tag{2.2b}$$

The motion of a particle is said to be *inertial* if $\dot{v}=0$.

The physical notion of an *inertial observer* (or system) is fully characterized mathematically by specifying a *constant timelike* vector field u, which, of course, can be constructed from the proper velocity u of a single inertial particle. It is often convenient to regard an inertial observer as an inertial particle with its history passing through the point $x=0$. The language can be considerably simplified by using the proper velocity of an observer as the *name* of the observer. A description of the motion of a particle according to an observer is, then, just a description of the motion of one particle relative to another.

Let u be an inertial observer and x any spacetime point (labelling some physical event). By virtue of (1.2),

$$xu = x\cdot u + x\wedge u = ct + \mathbf{x}, \tag{2.3a}$$

where

$$ct = x\cdot u, \tag{2.3b}$$

$$\mathbf{x} = x\wedge u. \tag{2.3c}$$

The quantities t and $\mathbf{x}$ are, respectively, the *time* and *position* of the event x according the observer u. For fixed t and variable x, (2.3b) is an equation for a space—like hyperplane with normal u, and each point x of the hyperplane is uniquely designated by $\mathbf{x}=x\wedge u$. For variable t, (2.3b) is an equation for a one parameter family of space—like hyperplanes. The time t designating a hyperplane is the proper time of the observer expressed in convenient units (say seconds); the constant c (with value equal to the speed of light) converts the unit of time into the unit of length.

Note that, by virtue of (1.2), (2.3) gives

$$ux = u\cdot x + u\wedge x = x\cdot u - x\wedge u = ct - \mathbf{x}.$$

Using this and $u^2=1$, one finds

$$x^2 = (xu)(ux) = (ct+\mathbf{x})(ct-\mathbf{x}) = c^2t^2 - \mathbf{x}^2, \tag{2.4}$$

a familiar expression for the "interval" between the event 0 and an event x.

Let $x = x(\tau)$ be the history of a particle with proper velocity $v = dx/d\tau$. Differentiating (2.3a), one finds

$$vu = \frac{d}{d\tau}(xu) = c\frac{dx}{d\tau} = v\cdot u + v\wedge u.$$

Introducing the abbreviation $\gamma \equiv v\cdot u = cdt/d\tau$ and defining the *relative velocity* $\mathbf{v}$ by

$$\mathbf{v} \equiv \frac{d\mathbf{x}}{dt} = \frac{d\tau}{dt}\frac{d\mathbf{x}}{d\tau} = c\frac{v\wedge u}{v\cdot u}, \tag{2.5a}$$

one obtains

$$vu = \gamma(1+\mathbf{v}/c) = L^2, \tag{2.5b}$$

where L is the spinor introduced in (1.18) to describe the boost of u into v. Since both v and u are unit vectors, one obtains from (2.5b)

$$1 = v^2 = (vu)(uv) = \gamma(1+\mathbf{v}/c)\gamma(1-\mathbf{v}/c) = \gamma^2(1-\mathbf{v}^2/c^2).$$

Hence

$$\gamma \equiv v\cdot u = c\frac{dt}{d\tau} = (1-\mathbf{v}^2/c^2)^{-1/2}. \tag{2.5c}$$

Any proper bivector which can be expressed as the outer product $a\wedge u$ of an observer u with some vector a may be called a *relative vector* (relative to u of course) and denoted by a letter in boldface type, as in (2.3c) and (2.5a). It is not difficult to show that the set of all relative vectors is a three-dimensional linear space, so that *relative position vectors* of the form (2.3c) may serve as labels for (or, indeed, as a definition of) the three-dimensional "physical space" of the observer u. The adjective "relative" serves to distinguish "relative vectors" from "proper vectors" and to emphasize that they describe a particular relation to an observer, but it may be omitted when understood from the context or the use of boldface type. Any proper vector can be reexpressed as an equivalent sum of a relative scalar and a relative vector by multiplying it by u, as has already been shown, for example, by (2.3a) and (2.5b). In this way a *proper description* of physical events can be reformulated as a *relative description* of events. Several more important examples will be given to show how easily this is accomplished with spacetime algebra.

Let p be the *proper momentum* (i.e., the energy—momentum vector) of a particle. Multiplying by u, one obtains from p the *energy* (or relative mass) E and the *relative momentum* $\mathbf{p}$; thus

$$pu = p\cdot u + p\wedge u = E + c\mathbf{p}, \tag{2.6a}$$

$$E \equiv p\cdot u, \tag{2.6b}$$

$$\mathbf{p} \equiv c^{-1}p\wedge u. \tag{2.6c}$$

For "physical particles" the *proper* (or *rest*) *mass* m is defined by the equation $p^2 = m^2c^4 \geqslant 0$. The relation of proper mass to energy and relative momemtum can be obtained from (2.6a); thus

$$p^2 = (pu)(up) = (E+c\mathbf{p})(E-c\mathbf{p}) = E^2 - c^2\mathbf{p}^2 = m^2c^4. \tag{2.7}$$

For material particles $m\neq 0$, and if the momentum is related to the velocity by the equation

$$p = mc^2v, \tag{2.8a}$$

one has, from (2.6c), the famous expressions

$$E = mc^2\gamma = mc^2(1-\mathbf{v}^2/c^2)^{-1/2}, \tag{2.8b}$$

$$\mathbf{p}=\frac{E}{c^2}\mathbf{v}=m\gamma\mathbf{v}=\frac{m\mathbf{v}}{(1-\mathbf{v}^2/c^2)^{1/2}}. \tag{2.8c}$$

Like the geometric product of proper vectors in (1.2) the geometric product of relative vectors a and b can be decomposed into an inner product $\mathbf{a}\cdot\mathbf{b}$ and an outer product $\mathbf{a}\wedge\mathbf{b}$; thus

$$\mathbf{ab}=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\wedge\mathbf{b}, \tag{2.9a}$$

$$\mathbf{a}\cdot\mathbf{b}\equiv\tfrac{1}{2}(\mathbf{ab}+\mathbf{ba}), \tag{2.9b}$$

$$\mathbf{a}\wedge\mathbf{b}\equiv\tfrac{1}{2}[\mathbf{a},\mathbf{b}]\equiv i\mathbf{a}\times\mathbf{b}. \tag{2.9c}$$

Equation (2.9c) expresses the relative *bivector* $\mathbf{a}\wedge\mathbf{b}$ as the *dual* of a *relative vector* $\mathbf{a}\times\mathbf{b}$, the i being the unit pseudoscalar already introduced in (1.8) and (1.9). The right side of (2.9c) can be regarded as a definition of the vector *cross product* $\mathbf{a}\times\mathbf{b}$. For further discussion of this relation see Refs. 3 and 1.

By multiplication and addition the relative vectors generate an algebra which is, in fact, exactly the even subalgebra of the complete spacetime algebra. Indeed, any element E of the even subalgebra can be written in the form

$$[E]=[E]_0+[E]_2+[E]_4=[E]_0+[E]_1+[E]_2+[E]_3, \tag{2.10a}$$

where

$$[E]_0=[E]_0, \tag{2.10b}$$

$$[E]_2=[E]_1+[E]_2, \tag{2.10c}$$

$$[E]_4=[E]_3. \tag{2.10d}$$

As in (1.10), $[E]_k$ indicates the *proper* k-vector part of E. Similarly, $[E]_k$ indicates the *relative* k-vector part. Of the three relations (2.10b)–(2.10d), (2.10c) is of the most interest here. It says that any proper bivector can be expressed as the sum of a relative vector and a relative bivector. To see how this decomposition can be carried out, consider the proper bivector F representing the electromagnetic field at some spacetime point. Note that, by (1.4),

$$F=Fu^2=(F\cdot u+F\wedge u)u,$$

so

$$F=\mathbf{E}+i\mathbf{B}, \tag{2.11a}$$

where

$$\mathbf{E}\equiv F\cdot uu=(F\cdot u)\wedge u=[F]_1, \tag{2.11b}$$

$$i\mathbf{B}\equiv F\wedge uu=(F\wedge u)\cdot u=[F]_2. \tag{2.11c}$$

The relative vector $\mathbf{E}$ is the *electric field according to the observer* u. The wedge in (2.11b) can be included or omitted as desired; this follows from (1.2a), since $F\cdot u$ is a proper vector which is orthonal to u, as shown by $(F\cdot u)\cdot u=F\cdot(u\wedge u)\equiv[Fu\wedge u]_0=0$. Similarly, by (1.7a) the dot in (2.11c) can be omitted or included at will because $(F\wedge u)\wedge u=F\wedge(u\wedge u)=0$. To justify the notation $\mathbf{B}$ indicating a relative vector in (2.11c), note that

$$\mathbf{B}=-iF\quad uu=(-iF)\cdot uu=[(-iF)\cdot u]\wedge u, \tag{2.12}$$

showing that the "proper expression" for $\mathbf{B}$ has the same form as the one for $\mathbf{E}$ if only the electromagnetic field F is replaced by its *dual* $-iF$, which is also a proper bivector. The relative vector $\mathbf{B}$ is the *magnetic field according to the observer* u.

In "proper notation" the classical equation of motion for a "test particle" with charge e and mass m takes the form

$$\dot{p}=mc^2\dot{v}=eF\cdot v, \tag{2.13}$$

with all symbols being defined as before, and, of course, $F=F(x(\tau))$. To reexpress (2.13) in "relative notation," it is helpful to note that

$$F^\dagger\equiv uFu=-uFu=\mathbf{E}-i\mathbf{B}. \tag{2.14}$$

So, with the help of (1.4a), (2.5b), (2.11), and (2.9),

$$\begin{aligned}(F\cdot v)u&=\tfrac{1}{2}(Fv-vF)u=\tfrac{1}{2}(Fvu+uvF^\dagger)\\&=\gamma\tfrac{1}{2}(\mathbf{E}(1+\mathbf{v}/c)+(1+\mathbf{v}/c)\mathbf{E})+\gamma\tfrac{1}{2}[i\mathbf{B},(1+\mathbf{v}/c)]\\&=\gamma[\mathbf{E}\cdot\mathbf{v}/c+\mathbf{E}+i\mathbf{B}\wedge\mathbf{v}/c].\end{aligned} \tag{2.15}$$

But (2.13) gives

$$\dot{p}u=\frac{d}{d\tau}(pu)=\frac{\gamma}{c}\frac{d}{dt}(E+c\mathbf{p})=e(F\cdot v)u.$$

So

$$c\gamma^{-1}\dot{p}\cdot u=\frac{dE}{dt}=e\mathbf{E}\cdot\mathbf{v} \tag{2.16a}$$

and

$$\gamma^{-1}\dot{p}\wedge u=\frac{d\mathbf{p}}{dt}=e(\mathbf{E}+c^{-1}\mathbf{v}\times\mathbf{B}), \tag{2.16b}$$

the usual relative vector form for the Lorentz force.

Obviously the decomposition (2.11) of the electromagnetic field F into electric and magnetic fields depends on the observer. The observer need not be inertial. Thus, the proper velocity $v=v(\tau)$ of a particle in arbitrary motion determines and instantaneous rest frame of the particle in which the electric field is

$$\mathbf{E}_v\equiv F\cdot vv=(F\cdot v)\wedge v \tag{2.17a}$$

and the magnetic field $\mathbf{B}_v$ is given by

$$i\mathbf{B}_v\equiv F\cdot vv=(F\wedge v)\cdot v, \tag{2.17b}$$

so that

$$F=\mathbf{E}_v+i\mathbf{B}_v. \tag{2.17c}$$

The subscript v indicates the rest system. Some such notation is necessary when relative vectors in more than one rest system are considered. The relative acceleration of the particle itself in its own inertial system is

$$\mathbf{a}_v\equiv c^2\dot{v}\wedge v=c^2\ddot{x}\wedge v. \tag{2.18}$$

Multiplying (2.13) by v and using (2.26) along with (2.17a) and (2.18), one finds

$$m\mathbf{a}_v=mc^2\dot{v}v=eF\cdot vv=e\mathbf{E}_v, \tag{2.19}$$

which says that a charge at (relative) rest is accelerated by an electric but not a magnetic field. Indeed, it is by (2.19) that an electric field is defined in the first place.

Now, as one more example and for later use, the proper velocity $\dot{v}$ will be expressed in relative form. From (2.5b),

$$\dot{v}u=\frac{d}{d\tau}(vu)=\dot{\gamma}(1+\mathbf{v}/c)+\gamma\dot{\mathbf{v}}/c. \tag{2.20}$$

Now

$$\dot{\mathbf{v}} = \frac{d\mathbf{v}}{d\tau} = \frac{dt}{d\tau}\frac{d\mathbf{v}}{dt} = c^{-1}\gamma\mathbf{a}$$

where

$$\mathbf{a} \equiv \frac{d\mathbf{v}}{dt} \tag{2.21}$$

is the *relative acceleration* of the particle. The quantity $\dot{\gamma}$ can be related to $\mathbf{a}$ by direct differentiation of (2.5c), but it is easier and more instructive to use (2.2). For this reason, consider

$$\begin{aligned} \dot{v}v &= (\dot{v}u)(uv) = [\dot{\gamma}(1+\mathbf{v}/c) + \gamma\dot{\mathbf{v}}/c]\gamma(1-\mathbf{v}/c) \\ &= \gamma[\dot{\gamma}(1-\mathbf{v}^2/c^2) + c^{-1}\gamma\dot{\mathbf{v}}(1-\mathbf{v}/c)]. \end{aligned}$$

The scalar part $\dot{v}\cdot v = 0 = \gamma[\dot{\gamma}(1-\mathbf{v}^2/c) - c^{-1}\gamma\dot{\mathbf{v}}\cdot\mathbf{v}/c]$, so, recalling (2.5c), one finds

$$\dot{\gamma} = c^{-2}\gamma^3\dot{\mathbf{v}}\cdot\mathbf{v} = c^{-3}\gamma^4\mathbf{v}\cdot\mathbf{a} = \dot{v}\cdot a = c\frac{d^2t}{dt^2}. \tag{2.22}$$

The bivector part is simply

$$\dot{v}v = \dot{v}\wedge v = c^{-1}\gamma^2(\dot{\mathbf{v}} - c^{-1}\dot{\mathbf{v}}\wedge\mathbf{v}) = c^{-2}\gamma^3(\mathbf{a} + c^{-1}i\mathbf{v}\times\mathbf{a}). \tag{2.23}$$

Substitution of (2.22) into the proper bivector part of (2.20) yields

$$\dot{v}\wedge u = c^{-1}(\gamma\dot{\mathbf{v}} + \dot{\gamma}\mathbf{v}) = c^{-2}\gamma^2(\mathbf{a} + c^{-2}\gamma^{-2}\mathbf{v}\cdot\mathbf{a}\mathbf{v}).$$

But a more helpful expression can be obtained from (2.23); thus,

$$\dot{v}u = (\dot{v}v)(vu) = c^{-2}\gamma^3(\mathbf{a} + c^{-1}\mathbf{v}\wedge\mathbf{a})\gamma(1 + c^{-1}\mathbf{v}),$$

the proper bivector part of which is

$$\dot{v}\wedge u = c^{-2}\gamma^4[\mathbf{a} + c^{-2}(\mathbf{v}\wedge\mathbf{a})\mathbf{v}] = c^{-2}\gamma^4[\mathbf{a} + c^{-2}\mathbf{v}\times(\mathbf{v}\times\mathbf{a})]. \tag{2.24}$$

3. THE COVARIANT POINT OF VIEW

Before continuing the proper description of mechanics, a brief discussion of its relation to more conventional formulations may be helpful.

Given an orthonormal frame $\{\gamma_\mu\}$, the coefficients $g_{\mu\nu}$ of the metric tensor (relative to that frame) are determined by the equation

$$g_{\mu\nu} = \tfrac{1}{2}(\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu) = \gamma_\mu\cdot\gamma_\nu. \tag{3.1}$$

This equation will appear familiar to anyone acquainted with the Dirac matrix algebra. Indeed, the spacetime algebra used here is algebraically isomorphic to the algebra generated by the Dirac matrices over the real numbers (a subalgebra of the full Dirac algebra over the complex numbers). It is important to understand the differences between these algebras. The γ_μ in (3.1) are regarded as vectors, whereas the corresponding Dirac matrices are ordinarily related to vectors only indirectly with the help of spinors. The Dirac matrices are hardly used except in connection with spin-$\frac{1}{2}$ particles, so one gets the impression that the Dirac algebra merely describes some property of spin. On the contrary, (3.1) is here a direct expression of the metric of spacetime as a rule for multiplying vectors, from which it follows that the full spacetime algebra directly expresses basic geometrical properties of spacetime. It is as applicable to any classical theory as it is to the quantum theory of spin-$\frac{1}{2}$ particles. The fact that the γ_μ can be represented by 4×4 matrices is irrelevent to any geometrical or physical application of spacetime algebra. Indeed, matrices introduce unnecessary mathematical complications and obscure interpretations even in the Dirac electron theory. This has been established in Refs. 4 and 5 and will be discussed more fully in a forthcoming paper.

With $u = \gamma_0$ being the proper velocity of an inertial observer, the relative vectors

$$\boldsymbol{\sigma}_i \equiv \gamma_i\gamma_0 = \gamma_i\wedge\gamma_0 \quad (i = 1,2,3) \tag{3.2}$$

compose a basis for the space of all relative vectors. The $\boldsymbol{\sigma}_i$ can be represented by the 2×2 Pauli matrices, from which it follows that the even subalgebra of the spacetime algebra is isomorphic to the Pauli matrix algebra. But again, matrices are of negative value. For example, from (3.2) one obtains

$$\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = i = \gamma_0\gamma_1\gamma_2\gamma_3, \tag{3.3}$$

where i is the unit pseudoscalar, a fundamental geometrical quantity; on the other hand, no geometrical significance is ordinarily attributed to the corresponding matrix equation. Moreover, the simple relations (3.2) and (3.3) between the γ_μ and the $\boldsymbol{\sigma}_i$ do not obtain if the γ_μ are to be represented by 4×4 matrices while the $\boldsymbol{\sigma}_i$ are represented by 2×2 matrices. For purposes of comparison with matrix representations of the Lorentz group, it should be noted that (3.2) enables one to write $A = a^\nu_\mu\gamma_\nu\gamma_0\gamma_0\gamma^\mu = a^0_0 + (a^0_i + a^i_0)\boldsymbol{\sigma}_i + a^i_j\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j$ in (1.16). So (1.16) can be represented as a 2×2 matrix, which the work of MacFarlane[6] shows immediately to be the representation of a Lorentz transformation in $SL\,2(C)$.

To transcribe proper equations into covariant tensor form, it is necessary to introduce a set of coordinates $\{x^\mu = x^\mu(x);\ \mu = 0,1,2,3\}$. It suffices to consider a set of "Cartesian coordinates," which can always be written in the form

$$x^\mu = x^\mu(x) = x\cdot\gamma^\mu, \tag{3.4a}$$

where $\{\gamma^\mu\}$ is an orthonormal frame of constant vectors with reciprocal frame $\{\gamma^\mu\}$. Equation (3.4a) expresses the coordinates as a function of the point x. The inverse function expressing the point x as a function of the coordinates $\{x^\mu\}$ is

$$x = x(x^0, x^1, x^2, x^3) = x^\mu\gamma_\mu. \tag{3.4b}$$

One readily verifies that

$$\Box x^\mu = \gamma^\mu, \tag{3.5a}$$

$$\partial_\mu x = \gamma_\mu, \tag{3.5b}$$

where

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \gamma_\mu\cdot\Box \quad\text{and}\quad \Box = \gamma^\mu\partial_\mu. \tag{3.5c}$$

Indeed, the relations of the form (3.5a, b, c) are completely general, obtaining for any set of coordinates.

As an example, the classical "Lorentz equation" (2.13) will be put in covariant form. The components of the velocity are

$$v^\mu = v\cdot\gamma^\mu \quad\text{and}\quad v_\mu = v\cdot\gamma_\mu.$$

Since the γ^μ are constant,

$$\dot{v}\cdot\gamma^\mu = \frac{d}{d\tau}(v\cdot\gamma^\mu) = \frac{dv^\mu}{d\tau}.$$

The tensor components of the electromagnetic field F are

$$F^{\mu\nu} = \gamma^\mu\cdot F\cdot\gamma^\nu = F\cdot(\gamma^\nu\wedge\gamma^\mu) = -F^{\nu\mu},$$

and the expression for F in terms of the $F^{\mu\nu}$ is

$$F = \tfrac{1}{2}F^{\mu\nu}\gamma_\mu\wedge\gamma_\nu.$$

So, with the help of (1.5),

$$F\cdot v = \tfrac{1}{2}F^{\mu\nu}(\gamma_\mu\wedge\gamma_\nu)\cdot v = \tfrac{1}{2}F^{\mu\nu}(\gamma_\mu v_\nu - v_\mu\gamma_\nu) = F^{\mu\nu}\gamma_\mu v_\nu,$$

$$\gamma^\mu\cdot F\cdot v = F^{\mu\nu}v_\nu.$$

Thus, if (2.13) is "dotted" with γ^μ, it can be given the familiar covariant tensor form

$$mc^2\frac{dv^\mu}{d\tau} = eF^{\mu\nu}v_\nu. \tag{3.6}$$

The covariant equation (3.6) describes the motion of a "test charge" *relative to* an arbitrarily chosen set of (Cartesian) coordinates. In contrast, the proper equation (2.13) is simpler because it is formulated and, as will be shown in a subsequent paper, can be solved without reference to any set of scalar coordinates.

It is a simple matter to reexpress any covariant tensor equation in proper form. But the converse is not true; for example, the important spinor representation (1.13) of a Lorentz rotation has no simple tensor form, nor, of course, does the Dirac equation. Therefore, the spacetime algebra is a more powerful mathematical tool than conventional tensor analysis.

4. PROPER KINEMATICS OF A RIGID POINT PARTICLE

As before, let v and $\dot{v}$ be, respectively, the proper velocity and the proper acceleration of a material point particle. From the fact that $\dot{v}\cdot v = 0$, it follows that it is always possible to find a bivector valued function $\Omega = \Omega(\tau)$ such that

$$\dot{v} = \Omega\cdot v. \tag{4.1}$$

Indeed, as shown by (2.17) and (2.11), Ω submits to the decomposition

$$\Omega = \boldsymbol{\alpha}_v + i\boldsymbol{\beta}_v = \dot{v}v + B, \tag{4.2a}$$

where

$$\boldsymbol{\alpha}_v \equiv \Omega\cdot vv = \dot{v}v = \dot{v}\wedge v,$$

$$i\boldsymbol{\beta}_v \equiv \Omega\wedge vv = (\Omega\wedge v)\cdot v \equiv B.$$

Noting that

$$B\cdot v = 0 \tag{4.2b}$$

and using the identity (1.5), one shows easily that (4.2a) satisfies (4.1). So any choice of B in (4.2a) will satisfy (4.1) provided only that $B\cdot v = 0$.

A *coming frame* of vectors $e_\mu = e_\mu(\tau)$ $(\mu = 0,1,2,3)$ can be introduced by the equations

$$e_\mu = R\gamma_\mu\tilde{R} \quad\text{with } e_0 = v = \dot{x}, \tag{4.3a}$$

where $\{\gamma_\mu\}$ is a fixed orthonormal frame of vectors, and $R = R(\tau)$ is a unimodular spinor, i.e.,

$$R\tilde{R} = 1, \tag{4.3b}$$

with the equation of motion

$$\dot{R} \equiv \frac{dR}{d\tau} = \tfrac{1}{2}\Omega R. \tag{4.3c}$$

Frequently, it is convenient to adopt the initial condition

$$R(0) = 1, \quad\text{or equivalently}\quad e_\mu(0) = \gamma_\mu, \tag{4.4}$$

but this will not be required in this section.

Equation (4.3a) is a Lorentz rotation of the frame $\{\gamma_\mu\}$ into the frame $\{e_\mu(\tau)\}$ determined by the spinor $R(\tau)$. From (4.3a, b) one shows easily that $e_\mu\cdot e_\nu = \gamma_\mu\cdot\gamma_\nu$, so the e_μ are orthonormal. The quantity Ω is the *angular velocity* of the spinor-valued function $R = R(\tau)$. Solving (4.3c) for Ω and differentiating (4.3b), one finds

$$\Omega = 2\dot{R}\tilde{R} = -2R\dot{\tilde{R}} = -2R\tilde{\dot{R}} = -\tilde{\Omega}. \tag{4.5}$$

Since R is an even multivector, so is Ω; more particularly, by virtue of (1.11), (4.5) implies that Ω is a bivector.

By differentiating (4.3a) one can obtain, with the help of (4.5), a set of differential equations for the e_μ which is equivalent to the single spinor equation (4.3c); thus

$$\dot{e}_\mu = \dot{R}\gamma_\mu\tilde{R} + R\gamma_\mu\dot{\tilde{R}}$$
$$= \tfrac{1}{2}(2\dot{R}\tilde{R})R\gamma_\mu\tilde{R} + \tfrac{1}{2}R\gamma_\mu\tilde{R}(2R\dot{\tilde{R}}),$$

or

$$\dot{e}_\mu = \tfrac{1}{2}[\Omega, e_\mu] = \Omega\cdot e_\mu. \tag{4.6}$$

This displays Ω as the *angular velocity* of the comoving frame, and for $\mu = 0$ it is seen to be identical to (4.1).

The arbitrariness in Ω which exists when (4.1) is considered alone obviously does not exist when the complete equations (4.6) for a comoving frame are given. But this is worth proving by solving (4.6) explicitly for Ω. Introducing the reciprocal frame $\{e^\mu\}$ defined by the equations

$$e^\mu\cdot e_\nu = \gamma^\mu\cdot\gamma_\nu = \delta^\mu_\nu \quad (\mu, \nu = 0,1,2,3), \tag{4.7}$$

one can prove the identities

$$e_\mu e^\mu = e_\mu\cdot e^\mu = 4, \tag{4.8}$$

$$e_\mu\Omega e^\mu = 0 \tag{4.9}$$

(sum over repeated indices). Identity (4.9) requires that Ω be a bivector. Multiplying (4.6) by e^μ and summing, one gets

$$\dot{e}_\mu e^\mu = \dot{e}_\mu\wedge e^\mu = (\Omega\cdot e_\mu)e^\mu$$
$$= \tfrac{1}{2}(\Omega e_\mu - e_\mu\Omega)e^\mu = \tfrac{1}{2}\Omega 4.$$

So

$$\Omega = \tfrac{1}{2}\dot{e}_\mu e^\mu = \tfrac{1}{2}\dot{e}_\mu\wedge e^\mu. \tag{4.10}$$

Clearly, there are many comoving frames which can be associated with the history of a particle, since with only the conditions set down so far the history $x = x(\tau)$ itself determines only one of the e_μ, the velocity $e_0 = v = \dot{x}$. A frame more intimately related to the history is easy to construct. Suppose the angular velocity Ω has the form

$$\Omega = \kappa_1 e_1 e_0 + \kappa_2 e_1 e_2 + \kappa_3 e_2 e_3 \tag{4.11}$$

where the κ_i $(i=1,2,3)$ are scalar quantities. Substitution of (4.11) into (4.6) yields, with the help of identity (1.5),

$$\begin{aligned}
\dot{e}_0 &= \kappa_1 e_1,\\
\dot{e}_1 &= \kappa_1 e_0 + \kappa_2 e_2,\\
\dot{e}_2 &= -\kappa_1 e_1 + \kappa_3 e_3,\\
\dot{e}_3 &= -\kappa_3 e_2.
\end{aligned} \tag{4.12}$$

These are the so-called Frenet—Serret equations for the particle history. It follows that the ith *curvature* κ_i of the history satisfies

$$\kappa_i = e_i \cdot \dot{e}_{i-1} = e_i \cdot \Omega \cdot e_{i-1} = \Omega \cdot (e_{i-1} \wedge e_i). \tag{4.13}$$

The angular velocity Ω of the *Frenet frame* $\{e_\mu\}$ satisfying (4.12) is called the *Darboux bivector*, because it generalizes the Darboux vector of classical differential geometry. It is not difficult to show that the Frenet frame determines all the derivatives of $x = x(\tau)$, and conversely, if none of the κ_i vanish the Frenet frame is uniquely determined by the derivatives of the history. One important feature of the formulation given here is that the single spinor equation (4.3c) with Ω related to the e_μ by (4.11) may be easier to solve than the simultaneous set of equations (4.12).

In spite of the geometrical significance of Frenet frames, other choices of a comoving frame are more important physically. Every material particle has some structure, usually because it approximates some extended body. A comoving frame can be used as a basic description of such structure. In particular, the comoving vectors e_1, e_2, e_3 may be used to specify a frame fixed in a rigid body (of negligible dimensions) moving with the particle; then Ω is the *proper angular velocity* of the rigid body and the spinor R completely describes any changes in orientation of the body. With this interpretation Eqs. (4.3) will be said to describe a *rigid (point) particle*. The dynamics of a rigid particle can be described by relating Ω to the motion of other physical system. But to facilitate the analysis of dynamics, it is worthwhile first to study the general kinematics of comoving frames in more detail.

The Lorentz rotation (4.3a) can be decomposed into a spatial rotation and a boost in the manner described in section 1. Take $u = \gamma_0$ and write as before

$$R = LU, \tag{4.14}$$

where L satisfies (1.18) and U satisfies (1.19). By substituting (4.14) into (4.3c), an equation of motion for the spinor U can be obtained; thus

$$\dot{R} = \dot{L}U + L\dot{U} = \tfrac{1}{2}\Omega L U.$$

So, since $\tilde{L}L = 1$,

$$\dot{U} = \tfrac{1}{2}\omega U, \tag{4.15a}$$

where

$$\omega = \tilde{L}\Omega L - 2\tilde{L}\dot{L}. \tag{4.15b}$$

The angular velocity ω can be separated into two parts

$$\omega = \omega_T + \omega_L, \tag{4.16a}$$

where

$$\omega_T \equiv \tilde{L}\dot{v}vL - 2\tilde{L}\dot{L}, \tag{4.16b}$$

$$\omega_L \equiv \tilde{L}BL = [\tilde{L}\Omega L]_2. \tag{4.16c}$$

To prove this, note that since $v = Ru\tilde{R} = Lu\tilde{L}$,

$$\tilde{L}(\Omega \cdot v + \Omega \wedge v)L = \tilde{L}\Omega v L = \tilde{L}\Omega L u = (\tilde{L}\Omega L)\cdot u + (\tilde{L}\Omega L)\wedge u.$$

Since a Lorentz rotation does not mix multivectors of different degree, one gets by separately equating vector and trivector parts

$$\tilde{L}\Omega \cdot vL = (\tilde{L}\Omega L)\cdot u \quad \text{and} \quad \tilde{L}\Omega \wedge vL = (\tilde{L}\Omega L)\wedge v,$$

which on multiplication by u gives, as in (2.11),

$$\tilde{L}\Omega \cdot vvL = (\tilde{L}\Omega L)\cdot uu = [\tilde{L}\Omega L]_1 \tag{4.17a}$$

and

$$\tilde{L}\Omega \wedge vvL = (\tilde{L}\Omega L)\wedge uu = [\tilde{L}\Omega L]_2 \tag{4.17b}$$

Recalling (4.2), one obtains (4.16) immediately by using (4.17a, b) in (4.15b).

The rigid frame $\{e_i = R\gamma_i\tilde{R};\ i = 1,2,3\}$ describes the orientation of a rigid body in the *instantaneous rest system* of the particle. The rigid frame

$$\tilde{L}e_iL = U\gamma_i\tilde{U}$$

provides an equivalent description of the rigid body in the *inertial system* U obtained by a (de)-boost from v. Alternatively, in the inertial system it is convenient to use the frame of relative vectors

$$\mathbf{e}_i \equiv U\boldsymbol{\sigma}_iU^\dagger = U\gamma_i\tilde{U}\gamma = \tilde{L}e_ivL \tag{4.18}$$

where, as before, $\boldsymbol{\sigma}_i = \gamma_i\gamma_0$ and $U^\dagger \equiv \gamma_0\tilde{U}\gamma_0$. Differentiating (4.18) and using (4.15a) as well as $U^\dagger U = 1$, one finds the equation of motion for the $\mathbf{e}_i$;

$$\dot{\mathbf{e}}_i = \omega \cdot \dot{\mathbf{e}}_i = \omega_T \cdot \mathbf{e}_i + \omega_L \cdot \mathbf{e}_i. \tag{4.19}$$

These equations describe a precession of the rigid body which according to (4.16) can be separated into two parts, the *Thomas precession* with angular velocity ω_T which is due to the acceleration of the particle, and the (generalized) *Larmor precession* with angular velocity ω_L of a nonaccelerated body.

The Thomas precession can be expressed in terms of u, v, and $\dot{v}$. Introducing the symbol w for the angular velocity of the boost, one has

$$\dot{L} = \tfrac{1}{2}wL \quad \text{or} \quad w = 2\dot{L}L = -2L\dot{\tilde{L}}. \tag{4.20}$$

Differentiating $L^2 = vu$,

$$\dot{v}u = \frac{dL^2}{d\tau} = \dot{L}L + L\dot{L} = \tfrac{1}{2}(wL^2 + LwL) = \tfrac{1}{2}(w + Lw\tilde{L})L^2,$$

then dividing by $\frac{1}{2}u$ and using $\tilde{L}v = u\tilde{L}$, one gets

$$2\dot{v} = wv + Lwu\tilde{L}. \tag{4.21}$$

Now since L is a function of u and v only, the bivector w is a function of the vectors u, v, and $\dot{v}$ only; hence the trivector $w \quad u$ must be proportional to $\dot{v} \quad v \quad u$. It follows, then, from (1.18) that $Lw \quad u\tilde{L} = w \quad u$. So the trivector part of (4.21) yields the equation

$$w \wedge (v + u) = 0. \tag{4.22}$$

This can be solved for w by dotting with v and using (1.8), thus

$$[w \wedge (v+u)]\cdot v = w(v+u)\cdot v - (w\cdot v)\wedge(u+v) = 0,$$

and since $\dot{v} = w\cdot v$, which is easily established by differentiating $v = R u\tilde{R} = L u\tilde{L}$, one obtains

$$w = 2\dot{L}\tilde{L} = \frac{\dot{v}\wedge(v+u)}{v\cdot(v+u)} = \frac{\dot{v}v + \dot{v}\wedge u}{1+v\cdot u}. \tag{4.23}$$

Now from the vector part of (4.21) one finds, again using $\dot{v} = w\cdot v$,

$$\dot{v} = Lw\cdot u\tilde{L} = w\cdot v, \tag{4.24}$$

from which one easily obtains the following expression for the relative vector part of w

$$[w]_1 = (w\cdot u)u = \tilde{L}\dot{v}vL. \tag{4.25}$$

This result can also be obtained directly from (4.16b) by using the fact that $\omega_T = [\omega_T]_2$, which can be proved from (4.15a) and (4.16c).

From (4.23) and (4.25) one obtains an expression for the relative bivector part of w:

$$[w]_2 = (w\wedge u)u = \frac{(\dot{v}\wedge v\wedge u)u}{1+v\cdot u} = 2\dot{L}\tilde{L} - \tilde{L}\dot{v}vL. \tag{4.26}$$

This is, in fact, identical to the Thomas expression (4.16b). To show this, recall from (1.18b) that $\tilde{L} = uLu$; so, by (4.20),

$$2\tilde{L}\dot{L} = u(2Lu\dot{L}) = u(2L\dot{\tilde{L}})u = -uwu = [w]_1 - [w]_2,$$

the last step being the same as in (2.14).

To sum up, the Thomas angular velocity ω_T can be written in the several different forms:

$$\omega_T = -[2\tilde{L}\dot{L}]_2 = [2\dot{L}\tilde{L}]_2 - \frac{(\dot{v}\wedge v\wedge u)u}{1+v\cdot u}$$
$$= \frac{[\dot{v}v]_2}{1+v\cdot u} = \frac{\gamma^3}{c^3(1+\gamma)}\, i\mathbf{v}\times\mathbf{a}, \tag{4.27}$$

the last expression as a relative bivector being obtained directly from (2.23); it is identical to that obtained by Thomas[7] and again by Bacry[8] in a review of Thomas' work.

The problem remains to express the Larmor bivector ω_L in terms of relative vectors. First express Ω in "relative form"

$$\Omega = \boldsymbol{\alpha} + i\boldsymbol{\beta}, \tag{4.28a}$$

$$\boldsymbol{\alpha} = \Omega\cdot uu, \tag{4.28b}$$

$$i\boldsymbol{\beta} = \Omega\wedge uu. \tag{4.28c}$$

Then, write Ω in the form

$$\Omega = \Omega_\parallel + \Omega_\perp, \tag{4.29a}$$

where

$$\hat{\mathbf{v}} \equiv v\wedge u/|v\wedge u| = \mathbf{v}/|\mathbf{v}|$$

is the unit relative velocity of the particle, and

$$\Omega_\parallel \equiv \tfrac{1}{2}(\Omega\hat{\mathbf{v}} + \hat{\mathbf{v}}\Omega)\hat{\mathbf{v}} = \boldsymbol{\alpha}_\parallel + i\boldsymbol{\beta}_\parallel, \tag{4.29b}$$

$$\Omega_\perp \equiv \tfrac{1}{2}[\Omega, \hat{\mathbf{v}}]\hat{\mathbf{v}} = \boldsymbol{\alpha}_\perp + i\boldsymbol{\beta}_\perp, \tag{4.29c}$$

with

$$\boldsymbol{\alpha}_\perp \equiv \tfrac{1}{2}[\boldsymbol{\alpha}, \hat{\mathbf{v}}]\hat{\mathbf{v}} = \boldsymbol{\alpha}\wedge\hat{\mathbf{v}}\hat{\mathbf{v}} = -(\boldsymbol{\alpha}\times\hat{\mathbf{v}})\times\hat{\mathbf{v}}, \tag{4.30a}$$

$$\boldsymbol{\alpha}_\parallel = \boldsymbol{\alpha} - \boldsymbol{\alpha}_\perp = \boldsymbol{\alpha}\cdot\hat{\mathbf{v}}\hat{\mathbf{v}}, \tag{4.30b}$$

and similar relations for $\boldsymbol{\beta}$. The significance of (4.29) lies in the fact that $\Omega_\parallel$ commutes with $\hat{\mathbf{v}}$ while $\Omega_\perp$ anticommutes with $\hat{\mathbf{v}}$, and since the bivector part of L is proportional to $\hat{\mathbf{v}}$, one has the relations

$$\tilde{L}\Omega_\parallel L = \tilde{L}L\Omega_\parallel = \Omega_\parallel, \tag{4.31a}$$

$$\tilde{L}\Omega_\perp L = \tilde{L}^2\Omega_\perp = \Omega_\perp L^2. \tag{4.31b}$$

Hence, using $L^2 = vu = \gamma(1 + c^{-1}\mathbf{v})$, one gets

$$\tilde{L}\Omega L = \Omega_\parallel + \Omega_\perp L^2$$
$$= \Omega + (\gamma - 1)\Omega_\perp + c^{-1}\gamma\Omega_\perp\mathbf{v}. \tag{4.32}$$

Now (4.30a) shows that $\boldsymbol{\beta}_\perp\hat{\mathbf{v}} = \boldsymbol{\beta}\wedge\mathbf{v} = i\boldsymbol{\beta}\times\mathbf{v}$, so

$$\Omega_\perp\mathbf{v} = \boldsymbol{\alpha}\wedge\mathbf{v} + i\boldsymbol{\beta}\wedge\mathbf{v} = -\boldsymbol{\beta}\times\mathbf{v} + i\boldsymbol{\alpha}\times\mathbf{v}. \tag{4.33}$$

Decomposing (4.30) into relative vector and bivector parts, one gets

$$\tilde{L}\Omega L = \boldsymbol{\alpha} + (\gamma - 1)\boldsymbol{\alpha}_\perp - c^{-1}\gamma\boldsymbol{\beta}\times\mathbf{v}$$
$$+ i(\boldsymbol{\beta} + (\gamma - 1)\boldsymbol{\beta}_\perp + c^{-1}\gamma\boldsymbol{\alpha}\times\mathbf{v}). \tag{4.34}$$

The relative bivector part of (4.34) gives the desired expression for the Larmor bivector:

$$\omega_L = [\tilde{L}\Omega L]_2 = i(\boldsymbol{\beta} + (\gamma - 1)\boldsymbol{\beta}_\perp + c^{-1}\gamma\boldsymbol{\alpha}\times\mathbf{v})$$
$$= i\left(\boldsymbol{\beta} - \frac{c^{-2}\gamma^2}{(\gamma - 1)}(\boldsymbol{\beta}\times\mathbf{v})\times\mathbf{v} + c^{-1}\gamma\boldsymbol{\alpha}\times\mathbf{v}\right). \tag{4.35}$$

The Thomas bivector can also be expressed in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Replacing F by Ω in (2.15), one finds

$$\dot{v}v = (\Omega\cdot v)v = (\Omega\cdot vu)uv$$
$$= \gamma(c^{-1}\boldsymbol{\alpha}\cdot\mathbf{v} + \boldsymbol{\alpha} + c^{-1}\mathbf{v}\times\boldsymbol{\beta})\gamma(1 - c^{-1}\mathbf{v})$$
$$= \gamma^2(\boldsymbol{\alpha} + c^{-2}\boldsymbol{\alpha}\cdot\mathbf{v}\mathbf{v} + c^{-1}\mathbf{v}\times\boldsymbol{\beta})$$
$$+ i\gamma^2(c^{-1}\mathbf{v}\times\boldsymbol{\alpha} + c^{-2}(\boldsymbol{\beta}\times\mathbf{v})\times\mathbf{v}). \tag{4.36}$$

Using this in (4.27), one obtains

$$\omega_T = \frac{[\dot{v}v]_2}{1+v\cdot u} = \frac{i\gamma^2}{c^2(1+\gamma)}(\boldsymbol{\beta}\times\mathbf{v})\times\mathbf{v} + \frac{i\gamma^2}{c(1+\gamma)}\mathbf{v}\times\boldsymbol{\alpha}. \tag{4.37}$$

Finally, adding (4.35) and (4.37) one gets for the total angular velocity

$$\omega = \omega_T + \omega_L = i\left(\boldsymbol{\beta} + \frac{\gamma}{c(1+\gamma)}\,\boldsymbol{\alpha}\times\mathbf{v}\right) \equiv -i\boldsymbol{\omega}, \tag{4.38}$$

and substituting this into (4.19), one gets for the equations of motion of the rigid body in the inertial system

$$\dot{\mathbf{e}}_i = \omega\cdot\mathbf{e}_i = -i\boldsymbol{\omega}\wedge\mathbf{e}_i = \boldsymbol{\omega}\times\mathbf{e}_i$$
$$= \frac{\gamma}{c}\frac{d\mathbf{e}_i}{dt} = \left(-\boldsymbol{\beta} + \frac{\gamma}{c(1+\gamma)}\mathbf{v}\times\boldsymbol{\alpha}\right)\times\mathbf{e}_i. \tag{4.39}$$

This result agrees with Thomas,[7] though it may be more general than he realized. It applies to any motion whatever of a rigid point particle. All dynamics lie in the specification of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, or equivalently of Ω.

The precession of a rigid body can be described either by equations (4.6) or by (4.37) (or better by their corresponding spinor equations). Failure to distinguish between these two different modes of description can cause confusion. The former describes the precession in the instantaneous rest frame of the rigid body, while the latter describes an equivalent motion of a rigid body in some arbitrarily chosen inertial frame. It is worthwhile to work out the relation of the (actual) axes $\mathbf{e}_i$ of the body in the instantaneous rest frame to the equivalent axes $\mathbf{e}_i$ in the inertial frame. Using (4.18) and the decomposition (4.30) with α replaced by $\mathbf{e}_i$, one finds

$$\begin{aligned} e_i u &= e_i v v u = L\mathbf{e}_i \tilde{L} L^2 = L\mathbf{e}_i L \\ &= L^2 \mathbf{e}_i^{\parallel} + \mathbf{e}_i^{\perp} \\ &= \gamma(1 - c^{-1}\mathbf{v})\mathbf{e}_i^{\parallel} + \mathbf{e}_i - \mathbf{e}_i^{\parallel} \\ &= c^{-1}\mathbf{e}_i \cdot \mathbf{v} + \mathbf{e}_i + (\gamma - 1)\mathbf{e}_i \cdot \mathbf{v}\mathbf{v}. \end{aligned}$$

Hence,

$$e_i \cdot u = c^{-1}\gamma \mathbf{v} \cdot \mathbf{e}_i, \tag{4.40a}$$

and the relative vector part is

$$e_i \wedge u = \mathbf{e}_i + \frac{(\gamma - 1)}{\mathbf{v}^2}\mathbf{e}_i \cdot \mathbf{v}\mathbf{v} = \mathbf{e}_i + \frac{\gamma^2}{c^2(\gamma + 1)}\mathbf{e}_i \cdot \mathbf{v}\mathbf{v}. \tag{4.40b}$$

[1]D. Hestenes, *Spacetime Algebra* (Gordon and Breach, New York, 1966).
[2]D. Hestenes and G. Sobczyk, *Geometric Calculus* (to be published).
[3]D. Hestenes, Am. J. Phys. **39**, 1013 (1971).
[4]D. Hestenes, J. Math. Phys. **8**, 748 (1967).
[5]D. Hestenes, J. Math. Phys. **14**, 893 (1973).
[6]A. J. MacFarlane, J. Math. Phys. **3**, 1116 (1962).
[7]L. Thomas, Phil. Mag. **3**, 1 (1927).
[8]H. Bacry, Nuovo Cimento **26**, 1164 (1962).

APPENDIX: ERRATA TO REFERENCE 1

Since this paper elaborates certain parts of Ref. 1, it is appropriate to include here the following list of errata to that monograph: The last line of Eq. (3.12) should read

$$+(-1)^{s-r}(a_1 \wedge \ldots \wedge a_r) \cdot (b_{s-r+1} \wedge \ldots \wedge b_s) b_1 \wedge b_2 \wedge \ldots \wedge b_{s-r}.$$

Delete the last minus sign on the right-hand side of Eq. (6.16). Equation (19.22) should read

$$\mathbf{E}' + i\mathbf{B}' = \mathbf{E}_{\parallel} + \beta(\mathbf{E}_{\perp} + \mathbf{v} \times \mathbf{B}) + i[\mathbf{B}_{\parallel} + \beta(\mathbf{B}_{\perp} - \mathbf{v} \times \mathbf{E})].$$

Equation (20.2) should read $\gamma^i \cdot \gamma_j = \delta^i{}_j$. Insert a factor of $\frac{1}{2}$ on the right-hand sides of Eq. (21.5) and (21.20) and in front of $R^{\mu}{}_{\alpha\beta\sigma}$ in Eq. (21.7). Delete the explicit factors of $\frac{1}{2}$ from Eq. (21.10). Dispense with the pseudoscalar part of (22.3) and delete Eq. (22.5b). Equation (23.15) should read $C_{ijk} = -C_{ikj}$. Equation (24.14) should read $C \equiv \gamma^k C_k$. The sentence following Eq. (A7) should read "where the *signature* s is the maximum number of linearly independent vectors. . . ." Replace the subscript i in (A12) by 1. Six lines after Eq. (B1), the sentence should begin "If $T \neq 0$, ..."

Proper dynamics of a rigid point particle

David Hestenes

Physics Department, Arizona State University, Tempe, Arizona 85281
(Received 1 April 1974)

A spinor formulation of the classical Lorentz force is given which describes the precession of an electron's spin as well as its velocity. Solutions are worked out applicable to an electron in a uniform field, a plane wave, and a Coulomb field.

INTRODUCTION

Every evidence indicates that the Dirac theory provides an optimal description of electron motion, but for many purposes it is unnecessarily complex. The classical model of an electron as a point charge is sometimes adequate, but of course it gives no account of electron spin. The minimal generalization of the classical model is obtained simply by expressing the Lorentz force as a spinor equation. The main objective of this paper is to study the solutions of this equation in some detail.

This approach has several advantages. As will be demonstrated, it provides a new and (it seems) simpler way of integrating the classical Lorentz force and expressing the orbit as a parametrized algebraic equation. Besides providing new insights into old results, the spinor solution describes the precession of electron spin with the same accuracy as it determines the orbit. The classical spinor equations are closely related in form to the Dirac equation. This narrows the gap between classical and quantum mechanical formulations of electron motion and hopefully will help clarify the relations between them.

Since the description of electron motion given here would be impossible without the mathematical apparatus developed in Ref. 1, familiarity with the notations and results given therein is presumed.

Section 1 shows how classical electrodynamics can be used to derive a spinor equation of motion for a localized charge distribution. As a important special case, the BMT equation is derived and shown to be already in the work of Thomas in a different form. The spinor formulation of the Lorentz force is given and it's applicability to a description of the electron is discussed.

In Secs. 2, 3, and 4 the spinor Lorentz force is integrated to describe the motion of a charge in a uniform field, in a plane wave and in a Coulomb field. The problems are worked out in considerable detail to illustrate fully the efficiency of spacetime algebra in practical computations. Though the spinor solutions for uniform and plane wave fields have been found previously by other authors, the treatment here is unique in many details. I believe the spinor solution for the Coulomb field is published here for the first time.

1. PROPER DYNAMICS

In Sec. 4 of Ref. 1 the kinematics of a rigid point particle were expressed in terms of its proper angular velocity Ω. Before the equations of motion can be solved, dynamical assumptions must be made to express Ω as a definite function of the proper time τ. These depend on the nature of the particle. As an example of great importance, typical assumptions of classical electrodynamics will be put into proper form here and related to a model of the electron.

The classical force on a localized charge distribution at rest is, after a multipole expansion,

$$\mathbf{f} = e\mathbf{E} + \mathbf{p}\cdot\nabla\mathbf{E} + \nabla\boldsymbol{\mu}\cdot\mathbf{B} + \cdots. \qquad (1.1)$$

Assume now that the charge distribution can be regarded as a particle (of zero extent) with total electric change e, intrinsic electric dipole moment $\mathbf{p}$, intrinsic magnetic dipole moment $\boldsymbol{\mu}$, and that the higher multipole moments vanish or are negligible. Assume also that (1.1) is an expression for the relative force $\mathbf{f}$ on the particle in its instantaneous rest frame, more specifically, that

$$\mathbf{f} = f\wedge v = mc^2\dot{v}v = fv. \qquad (1.2)$$

For vanishing $\mathbf{p}$ and $\boldsymbol{\mu}$, then, (1.1) and (1.2) reduce to the Lorentz force as already shown in (I.2.19). Therefore, it is only necessary to express the last two terms in proper form.

Define now the *proper moment bivector* M of the particle by the equations

$$M = -\mathbf{p} + i\boldsymbol{\mu}, \qquad (1.3a)$$

$$-\mathbf{p} = M\cdot vv, \qquad (1.3b)$$

$$i\boldsymbol{\mu} = M\wedge vv. \qquad (1.3c)$$

In the instantaneous rest system $F = \mathbf{E} + i\mathbf{B}$ where $\mathbf{E} = F\cdot vv$ and $i\mathbf{B} = F\wedge vv$; so, for instance,

$$M\cdot F = -\mathbf{p}\cdot\mathbf{E} - \boldsymbol{\mu}\cdot\mathbf{B}, \qquad (1.4)$$

which is the familiar classical expression for the energy of electric and magnetic dipoles. The second term of (1.4) by itself can be written

$$\boldsymbol{\mu}\cdot\mathbf{B} = -[(M\wedge v)\cdot v]\cdot F - (M\wedge v)\cdot(v\wedge F). \qquad (1.5)$$

The formulas used in (1.5) to rearrange the inner and outer products are established in Refs. 2 and 3. The proper form for ∇ in (1.1) is $v \quad \square$, so

$$\mathbf{p}\cdot\nabla = -((M\cdot v)\wedge v)\cdot(v\wedge\square) = -M\cdot(v\wedge\square). \qquad (1.6)$$

Substituting the proper expressions in (1.1), one gets

$$\mathbf{f} = f\wedge v = eF\cdot vv - M\cdot(v\wedge\square)F\cdot vv - v\wedge\square(M\wedge v)\cdot(v\wedge F) \qquad (1.7)$$

and substituting this in (1.2) and dividing by v one obtains finally

$$mc^2\dot{v} = f = [eF - M\cdot(v\wedge\square)F - v\wedge\square(M\wedge v)\cdot(v\wedge F)]\cdot v. \qquad (1.8)$$

The form of this equation suggests taking the term in

 Copyright © 1974 American Institute of Physics

brackets to be Ω, but as (I.4.2) shows, an expression for $\dot{v}$ determines only part of Ω, so additional dynamical assumptions are required.

In order to get equations describing the motion of an electron, assume that $\mathbf{p}=0$, or equivalently,

$$v \cdot M = 0. \tag{1.9}$$

With this condition (1.5) can be replaced by the simpler relation

$$\boldsymbol{\mu} \cdot \mathbf{B} = -M \cdot F. \tag{1.10}$$

Next assume that M has constant magnitude and is proportional to the *spin* (intrinsic angular momentum) *bivector S*, that is,

$$M = c\lambda S, \quad \text{where } \lambda \equiv ge/2mc^2, \tag{1.11}$$

the constant g being the usual *gyromagnetic ratio*. The relation (1.11) obtains if the magnetic moment arises from a circulating charge distribution. If the distribution has a constant ratio of charge to mass density, it is easy to show that $g=1$, in disagreement with the value $g=2$ which obtains for an electron. However, other assumptions about the structure of the particle will give almost any desired value for g.

From (1.9) and (1.11) it follows that

$$v \cdot S = 0. \tag{1.12}$$

There exists a unique proper vector s called the *spin vector* such that

$$S = isv = is \wedge v. \tag{1.13a}$$

This can be proved simply by solving for s; thus

$$s = -iSv = iS \wedge v. \tag{1.13b}$$

It follows from this that $s \cdot v = 0$. The spin can now be related to the kinematical equations (I.4.3) for a rigid point particle by writing

$$s = |s| e_3. \tag{1.14}$$

But to get a definite functional form for the equations, classical dynamical considerations are helpful, at least as a guide.

For a magnetic dipole at rest in a magnetic field the classical theory gives the famous equation for the *Larmor precession* of the spin,

$$\frac{d\mathbf{s}}{dt} = \boldsymbol{\mu} \times \mathbf{B}. \tag{1.15}$$

More generally, the classical theory adds a term proportional to $\nabla \times \mathbf{E}$ to the right side of (1.15), but, following Thomas, this can be neglected in the first approximation. To put (5.15) in proper form in accordance with the preceding assumptions, write

$$\mathbf{s} = sv = s \wedge v, \tag{1.16a}$$

$$\boldsymbol{\mu} = c\lambda\mathbf{s} = c\lambda sv, \tag{1.16b}$$

$$i\mathbf{B} = B \equiv (F \wedge v)v. \tag{1.17}$$

Also, it is necessary to take account of the fact that (1.15) was derived for an inertial frame rather than an instantaneous rest frame. This can be done by interpreting the left side of (1.15) as a special case of $c\dot{s} \wedge v$ [just as was done for the acceleration in (I.2.19)], rather than as $cd(sv)/d\tau$, which can be shown to be inconsistent with the condition $s \cdot v = 0$. After noting that

$$\boldsymbol{\mu} \times \mathbf{B} = -\tfrac{1}{2}[\boldsymbol{\mu}, i\mathbf{B}] = \tfrac{1}{2}[i\mathbf{B}, \boldsymbol{\mu}],$$

(1.15) can be put in the form

$$c\dot{s} \wedge v = c\lambda\tfrac{1}{2}[B, sv] = c\lambda\tfrac{1}{2}[B, s]v = c\lambda B \cdot sv.$$

Multiplying by v and using

$$(\dot{s} \wedge v)v = (\dot{s} \wedge v) \cdot v = \dot{s} - (\dot{s} \cdot v)v$$
$$= \dot{s} + (\dot{v} \cdot s)v = \dot{s} - (\dot{v}v) \cdot s,$$

which is a consequence of $s \cdot v = 0$, one obtains the equation of motion for s:

$$\dot{s} = \lambda B \cdot s - (\dot{v} \cdot s)s = (\dot{v}v + \lambda B) \cdot s \tag{1.18a}$$

This is the so-called Bargmann—Michel—Telegdi (BMT) equation.[4] Since derivatives of the field were neglected in the derivation of (1.18a), the same assumption must be made in the corresponding equation for v. Hence, in (1.18a)

$$\dot{v}v = \frac{e^2}{mc^2}(F \cdot v)v = \frac{e^2}{mc^2}(F - B). \tag{1.18b}$$

While equations (1.18a, b) hold rigorously only for a homogeneous (i.e., constant in time and uniform in space) field F, they may serve as a useful approximation under other conditions. Indeed, Thomas used them in a different form to calculate the spin precession of an electron in an atom.

According to (1.14), (1.18a) is an equation for the unit spacelike vector e_3. Comparison with Eqs. (I.4.2a), (I.4.3), and (I.4.6) suggests that Eqs. (1.18a, b) be interpreted as equations of motion for a rigid point particle with angular velocity

$$\Omega = \dot{v}v + \lambda B = \frac{e}{mc^2}(F \cdot v)v + \frac{ge}{2mc^2}(F \wedge v)v$$
$$= \frac{e}{mc^2}[F + (g/2 - 1)B]. \tag{1.19}$$

For an electron, according to atomic theory, $g=2$, in which case (1.19) reduces to the strikingly simple form

$$\Omega = \frac{e}{mc^2}F = \lambda F \tag{1.20a}$$

and the spinor equation for an electron is

$$\dot{R} = \tfrac{1}{2}\Omega R = \frac{e}{2mc^2}FR. \tag{1.20b}$$

Of course, the argument leading up to (1.20) can in no sense be regarded as a derivation from any consistent classical model of the electron as a spinning charge distribution. However, an equation exactly of the form (1.20b) has been derived as an approximation of the Dirac equation [Eq. (6.17) of Ref. 5], though the significance of the approximation is not entirely clear. Therefore, it is interesting that (1.20) can be tested directly by experiments on the spin precession of electrons moving through a constant field,[6] and that the anomolous magnetic moment of the electron can be evaluated by using (1.19).

Equation (1.18a) describes the spin precession in the instantaneous rest frame of the particle. The equivalent equation describing spin precession in an inertial frame can be obtained directly by expressing the proper angular velocity given by (1.19) in relative form and using Eq. (I.4.45); write

$$\Omega = \alpha + i\beta = \frac{e}{mc^2}F + \left(\gamma - \frac{e}{mc^2}\right)B,$$

$$F = \mathbf{E} + i\mathbf{B},$$

and, with the help of (I.4.34),

$$B = F - F\cdot vv = (1-\gamma^2)\mathbf{E} - \gamma^2(e^{-2}\mathbf{E}\cdot\mathbf{vv} + c^{-1}\mathbf{v}\times\mathbf{E}) + i\{\mathbf{B} - \gamma^2(c^{-1}\mathbf{v}\times\mathbf{E} + c^{-2}(\mathbf{B}\times\mathbf{v})\times\mathbf{v})\}.$$

From these equations, expressions for α and β can be read off directly, which, on substitution into (I.4.36) and some rearrangement of terms, yields

$$\omega = -\left[\frac{e}{mc^2} + \gamma\left(\lambda - \frac{e}{mc^2}\right)\right]\mathbf{B} - \left(\frac{e}{mc^2}\frac{\gamma^2}{c(\gamma+1)} - \frac{\lambda\gamma}{c}\right)\mathbf{v}\times\mathbf{E} - \frac{\gamma^2}{c^2(1+\gamma)}\left(\frac{e}{mc^2} - \lambda\right)\mathbf{B}\cdot\mathbf{vv}. \tag{1.21a}$$

So the equation for the spin $\sigma \equiv |s|\mathbf{e}_3$ in the inertial system is, by (I.4.39),

$$\sigma = \frac{\gamma}{c}\frac{d\sigma}{dt} = \omega\times\sigma. \tag{1.21b}$$

This is exactly the result obtained by Thomas [Ref. 7, his Eq. (4.121)], and proves directly its equivalence to the BMT equation (1.18a)

Equations equivalent to (1.20b) have been discussed by other authors.[4,8,9] However, the form (1.20b) is easier to handle than other forms because it is supported by the spacetime algebra. Equation (1.20b) describes precession of both electron spin and velocity with the same degree of accuracy that the Lorentz force describes electron motion. Even apart from equations of spin, it is sometimes easier to solve than the Lorentz equation. For these reasons, Eq. (1.20b) is important enough to be given a name and its basic solutions will be thoroughly studied in the following sections.

No attempt will be made here to generalize (1.20) to get a more precise description of the electron, since, short of the full Dirac equation, the best procedure is unclear. Equation (1.20) can be used in connection with quantum theory by taking the spin to be the quantum mechanical polarization vector. It will be referred to as "the *spinor Lorentz force*" or as "the equation of motion for a *rigid test charge*"; the adjective "test" serves to indicate that radiation of the charge is not taken into account, while the adjective "rigid" indicates that a complete comoving frame is described. Of course, the "rigid test charge" is most important as a model of the electron if the charge e is negative or a positron if e is positive.

2. RIGID TEST CHARGE IN A HOMOGENEOUS FIELD

The spinor equation of motion $\mathring{R} = \frac{1}{2}\Omega R$ for a rigid point particle with constant proper angular velocity Ω integrates immediately to

$$R = \exp(\Omega\tau/2) = \sum_{n=0}^{\infty}\frac{1}{n!}(\tfrac{1}{2}\Omega\tau)^n, \tag{2.1}$$

where the initial condition $R(0)=1$ has been adopted. With the dynamical assumption

$$\Omega = \lambda F = (e/mc^2)F, \tag{2.2}$$

the spinor (2.1) describes the motion of are rigid test charge in a homogeneous electromagnetic field F. In particular, it describes the precession of the velocity and spin of an electron. Thus, the electron velocity $v = v(\tau)$ and spin $s = s(\tau)$ are given explicitly by

$$v = Rv_0\tilde{R} = \exp(F\lambda\tau/2)v_0\exp(F\lambda\tau/2), \tag{2.3a}$$

$$s = Rs_0\tilde{R} = \exp(F\lambda\tau/2)s_0\exp(F\lambda\tau/2). \tag{2.3b}$$

The history of the electron can be obtained by integration from (2.3a). To do this, it is convenient to assume that F is nonnull. The alternative case of a homogeneous null field has little practical significance; in any event it can be treated separately if necessary.

Since F is assumed to be nonnull, in accordance with (I.1.3), it is subject to the canonical decomposition into orthogonal blades:

$$F = \alpha f + \beta i f = fz, \tag{2.4a}$$

where α and β are scalars,

$$z = \alpha + i\beta \quad \text{with } \alpha \geq 0, \tag{2.4b}$$

and f is a simple unit timelike vector, that is,

$$f^2 = 1 \quad \text{and} \quad [f]_2 = f. \tag{2.4c}$$

Substituting (2.4a) into (2.1), R can be written

$$R = \exp(f\alpha\lambda\tau)\exp(if\beta\lambda\tau) = (\cosh\alpha\lambda\tau + f\sinh\alpha\lambda\tau) \times(\cos\beta\lambda\tau + if\sin\beta\lambda\tau). \tag{2.5}$$

Now using (2.4c) and (I.1.4), the initial velocity v_0 can be decomposed into a component v_0 in the f-plane and a component v_0 orthogonal to the f-plane; thus

$$v_0 = f^2 v_0 = v_{0\parallel} + v_{0\perp}, \tag{2.6a}$$

where

$$v_{0\parallel} = f(f\cdot v_0) = \tfrac{1}{2}(v_0 - fv_0 f) = -if(if)\wedge v_0, \tag{2.6b}$$

$$v_{0\perp} = f(f\wedge v_0) = \tfrac{1}{2}(v_0 + fv_0 f) = -if(if)\cdot v_0. \tag{2.6c}$$

From (2.5b, c), one has

$$fv_{0\parallel} = f\cdot v_0 = -v_{0\parallel}f, \tag{2.7a}$$

$$fv_{0\perp} = f\wedge v_0 = v_{0\perp}f. \tag{2.7b}$$

Using this and recalling $v_0 i = -iv_0$, one finds

$$v_{0\parallel}\exp(-f\alpha\lambda\tau/2)\exp(-if\beta\lambda\tau/2) = \exp(f\alpha\lambda\tau/2) \times\exp(-if\beta\lambda\tau/2)v_{0\parallel}, \tag{2.8a}$$

$$v_{0\perp}\exp(-f\alpha\lambda\tau/2)\exp(-if\beta\lambda\tau/2) = \exp(-f\alpha\lambda\tau/2) \times\exp(if\beta\lambda\tau/2)v_{0\perp}. \tag{2.8b}$$

So, substituting (2.5) in (2.3a) and using (2.8), one obtains

$$v = \frac{dx}{d\tau} = \exp(f\alpha\lambda\tau)v_{0\parallel} + \exp(if\beta\lambda\tau)v_{0\perp}. \tag{2.9}$$

This can be integrated immediately to get the history $x = x(\tau)$:

$$x - x_0 = \frac{(\exp(f\alpha\lambda\tau) - 1)}{\alpha\lambda} f \cdot v_0 + \frac{(\exp(if\beta\lambda\tau) - 1)}{\beta\lambda} (if) \cdot v_0. \tag{2.10}$$

This solution is valid even if $\alpha = 0$ and/or $\beta = 0$, as is easily established by expressing the exponential as a power series.

It is worth noting that, more generally, integration of the equations of motion can be carried out in essentially the same way as above when z is any function of τ as long as f is constant. This situation obtains when one has fields with fixed direction but spacial and/or temporal variations in magnitude.

The problem remains to reexpress the solutions (2.9) and (2.10) in terms of relative vectors such as the electric and magnetic field strengths $\mathbf{E}$ and $\mathbf{B}$, because these quantities have direct observational significance. To accomplish this, it is necessary to relate the decomposition $F = \mathbf{E} + i\mathbf{B}$ relative to a given observer u to the canonical decomposition (2.4) which is independent of any observer. The relation is a simple one in the case that

$$f \wedge u = 0; \tag{2.11a}$$

then,

$$\hat{f} = \hat{\mathbf{E}}, \quad \alpha = |\mathbf{E}|, \tag{2.11b}$$

$$\alpha\hat{f} = \mathbf{E}, \quad \beta\hat{f} = \beta\hat{\mathbf{E}} = \mathbf{B}; \tag{2.11c}$$

all of which is equivalent to the condition

$$\mathbf{E} \wedge \mathbf{B} = 0, \tag{2.11d}$$

that is, $\mathbf{E}$ and $\mathbf{B}$ are *parallel fields*. This case is important enough in itself to work out before proceeding to the general case. Using (2.11b) in (I.2.15), one can write down immediately

$$f \cdot v_0 u = \gamma_0\left(\frac{\hat{\mathbf{E}} \cdot v_0}{c} + \hat{\mathbf{E}}\right), \tag{2.12a}$$

$$(if) \cdot v_0 u = \frac{\gamma_0}{c} i\hat{\mathbf{E}} \wedge \mathbf{v}_0 = \frac{\gamma_0}{c} \mathbf{v}_0 \times \hat{\mathbf{E}}, \tag{2.12b}$$

where

$$\gamma_0 \mathbf{v}_0 = v_0 \wedge u \quad \text{and} \quad \gamma_0 = v_0 \cdot u = (1 - \mathbf{v}_0^2/c^2)^{-1/2}. \tag{2.13}$$

Using (2.6), one obtains from (2.12)

$$v_{0\parallel} u = f(f \cdot v_0) u = \gamma_0\left(\mathbf{E}\frac{(\mathbf{E} \cdot \mathbf{v}_0)}{c} + 1\right) = \gamma_0\left(\frac{\mathbf{v}_{0\parallel}}{c} + 1\right), \tag{2.14a}$$

$$v_{0\perp} u = -if(if) \cdot v_0 u = \frac{\gamma_0}{c} \hat{\mathbf{E}} \times (\mathbf{v}_0 \times \hat{\mathbf{E}}) = \frac{\gamma_0}{c} v_{0\perp}. \tag{2.14b}$$

Equation (2.9) can now be easily expressed as an equation in relative quantities by multiplying by u and using (2.14):

$$vu = \gamma\left(1 + \frac{\mathbf{v}}{c}\right) = \exp(\mathbf{E}\lambda\tau)\gamma_0\left(\hat{\mathbf{E}}\frac{\hat{\mathbf{E}} \cdot \mathbf{v}_0}{c} + 1\right) + \exp(i\mathbf{B}\lambda\tau)\gamma_0 \mathbf{v}_{0\perp}. \tag{2.15}$$

The scalar part of (2.15) is the equation

$$\frac{\gamma}{\gamma_0} = \cosh(|\mathbf{E}|\lambda\tau) + \frac{\hat{\mathbf{E}} \cdot \mathbf{v}_0}{c} \sinh(|\mathbf{E}|\lambda\tau). \tag{2.16}$$

Writing

$$\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp \quad \text{where } \mathbf{v}_\parallel = \mathbf{v} \cdot \hat{\mathbf{E}}\hat{\mathbf{E}}, \tag{2.17a}$$

one has from the vector part of (2.16)

$$\mathbf{v}_\parallel = \gamma_0 \sinh(|\mathbf{E}|\lambda\tau) + \frac{\mathbf{v}_0 \cdot \hat{\mathbf{E}}}{c} \cosh(|\mathbf{E}|\lambda\tau)\, \hat{\mathbf{E}}, \tag{2.17b}$$

$$\gamma \mathbf{v}_\perp = \exp(i\mathbf{B}\lambda\tau)\gamma_0 \mathbf{v}_{0\perp}. \tag{2.17c}$$

It will be noted that (2.17) is simplified by expressing it in terms of the relative momentum $\mathbf{p} = m\gamma\mathbf{v}$. Now multiplying (2.10) by u and using (2.12), one obtains

$$(x - x_0)u = (t - t_0) + (\mathbf{x} - \mathbf{x}_0)$$
$$= \frac{\gamma}{\lambda\mathbf{E}^2}(\exp(\mathbf{E}\lambda\tau) - 1)\left(\frac{\mathbf{E} \cdot \mathbf{v}_0}{c} + \mathbf{E}\right) + \frac{\gamma_0}{\lambda\mathbf{B}^2}(\exp(i\mathbf{B}\lambda\tau) - 1)\frac{v_0}{c} \times \mathbf{B}. \tag{2.18}$$

The scalar part of (2.18) gives the functional relation between the "laboratory time" t and the proper time τ:

$$t - t_0 = \frac{\gamma_0}{\lambda|\mathbf{E}|}\left(\sinh(|\mathbf{E}|\lambda\tau) + \frac{\hat{\mathbf{E}} \cdot \mathbf{v}_0}{c}[\cosh(|\mathbf{E}|\lambda\tau) - 1]\right). \tag{2.19}$$

The vector part of (2.18) is a parametric equation for the orbit $\mathbf{x} = \mathbf{x}(\tau)$:

$$\mathbf{x} - \mathbf{x}_0 = \frac{\gamma_0}{\lambda\mathbf{E}^2}\,[\cosh(|\mathbf{E}|\lambda\tau) - 1] + \frac{\hat{\mathbf{E}} \cdot \mathbf{v}_0}{c} \sinh(|\mathbf{E}|\lambda\tau)\, \mathbf{E}$$
$$+ \frac{\gamma_0}{c\lambda\mathbf{B}^2}(\exp(i\mathbf{B}\lambda\tau) - 1)\mathbf{v}_0 \times \mathbf{B}. \tag{2.20}$$

If $\mathbf{E}, \mathbf{B} \neq 0$, the orbit is a spiral with decreasing radius and increasing pitch as the charge loses energy to the field.

Now returning to the general case, it is necessary to express α, β, and f in terms of $\mathbf{E}$ and $\mathbf{B}$. Squaring (2.4a), one has

$$F^2 = z^2 = \alpha^2 - \beta^2 + 2i\alpha\beta = (\mathbf{E} + i\mathbf{B})^2 = \mathbf{E}^2 - \mathbf{B}^2 + 2i\mathbf{E} \cdot \mathbf{B}.$$

Hence,

$$\alpha^2 - \beta^2 = \mathbf{E}^2 - \mathbf{B}^2, \tag{2.21a}$$

$$\alpha\beta = \mathbf{E} \cdot \mathbf{B}. \tag{2.21b}$$

Solving for α and β, one gets

$$\alpha = \left(\frac{|z|^2 + \mathbf{E}^2 - \mathbf{B}^2}{2}\right)^{1/2} > 0, \tag{2.22a}$$

$$\beta = \pm\left(\frac{|z|^{-2} - \mathbf{E}^2 + \mathbf{B}^2}{2}\right)^{1/2}, \tag{2.22b}$$

where

$$|z|^2 = \alpha^2 + \beta^2 = [(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2]^{1/2} \tag{2.22c}$$

and the sign of β is determined by the rule

$$\beta \gtrless 0 \quad \text{if } \mathbf{E} \cdot \mathbf{B} \gtrless 0. \tag{2.22d}$$

Equation (2.4a) can be solved for f by

$$f = z^{-1}F = \frac{(\alpha - i\beta)}{|z|^2}(\mathbf{E} + i\mathbf{B}).$$

So, expressing f in terms of relative vectors $\mathbf{e}$ and $\mathbf{b}$, one has

$$f = \mathbf{e} + i\mathbf{b}, \tag{2.23a}$$

where

$$\mathbf{e}=(\alpha\mathbf{E}+\beta\mathbf{B})/|z|^2, \tag{2.23b}$$

$$\mathbf{b}=(\alpha\mathbf{B}-\beta\mathbf{E})/|z|^2. \tag{2.23c}$$

It is worth noting that, from (2.4c) or from (2.23b, c),

$$f^2=\mathbf{e}^2-\mathbf{b}^2=1, \tag{2.24a}$$

$$\mathbf{e}\cdot\mathbf{b}=0. \tag{2.24b}$$

Now, using (2.23a) in (I.2.15) one gets

$$f\cdot v_0u=\gamma_0\left(\frac{\mathbf{e}\cdot\mathbf{v}_0}{c}+\mathbf{e}+\frac{\mathbf{v}_0\times\mathbf{b}}{c}\right), \tag{2.25a}$$

$$(if)\cdot v_0u=\gamma_0\left(-\frac{\mathbf{b}\cdot\mathbf{v}_0}{c}-\mathbf{b}+\frac{\mathbf{v}_0\times\mathbf{e}}{c}\right). \tag{2.25b}$$

And using (2.26b, c) with (2.25a, b), one gets

$$v_{0\parallel}u=\gamma_0\left(\mathbf{e}^2-\frac{\mathbf{v}_0}{c}\cdot(\mathbf{e}\times\mathbf{b})+\mathbf{e}\frac{\mathbf{e}\cdot\mathbf{v}_0}{c}+\mathbf{b}\frac{\mathbf{b}\cdot\mathbf{v}_0}{c}-\mathbf{b}^2\frac{\mathbf{v}_0}{c}+\mathbf{e}\times\mathbf{b}\right), \tag{2.26a}$$

$$v_{0\perp}u=-\gamma_0\left(\mathbf{b}^2-\frac{\mathbf{v}_0}{c}\cdot(\mathbf{e}\times\mathbf{b})+\mathbf{e}\frac{\mathbf{e}\cdot\mathbf{v}_0}{c}+\mathbf{b}\frac{\mathbf{b}\cdot\mathbf{v}_0}{c}-\mathbf{e}^2\frac{\mathbf{v}_0}{c}+\mathbf{e}\times\mathbf{b}\right). \tag{2.26b}$$

Finally, using (2.6) and (2.7) and multiplying by u, (2.9) and (2.10) can be put in the forms

$$vu=\gamma\left(1+\frac{\mathbf{v}}{c}\right)=v_{0\parallel}u\cosh(\alpha\lambda\tau)+f\cdot v_0u\sinh(\alpha\lambda\tau)+v_{0\perp}u\cos(\beta\lambda\tau)-(if)\cdot v_0\sin(\beta\lambda\tau), \tag{2.27}$$

$$\begin{aligned}(x-x_0)u=(t-t_0)+(\mathbf{x}-\mathbf{x}_0)&=f\cdot v_0u\left(\frac{\cosh(\alpha\lambda\tau)-1}{\alpha\lambda}\right)\\&+v_{0\parallel}u\frac{\sinh(\alpha\lambda\tau)}{\alpha\lambda}+(if)\cdot v_0u\left(\frac{\cos(\beta\lambda\tau)-1}{\beta\lambda}\right)\\&-v_{0\perp}u\sin(\beta\lambda\tau).\end{aligned} \tag{2.28}$$

Substitution of (2.25) and (2.26) into (2.27) and (2.28) followed by separation into scalar and vector parts yields the complete solutions in relative form. The fact that the resulting relative formulas appear so much more complicated than the equivalent proper formulas (2.9) and (2.10) merely shows that the relative vectors $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{v}_0$ are a poor choice of parameters for the problem.

Insight which leads to a better choice of relative vectors as parameters can be gained as follows. A boost of u into

$$w=Wu\tilde{W}=W^2u$$

can be defined by requiring that W boost $\hat{\mathbf{e}}\equiv|\mathbf{e}|^{-1}\mathbf{e}$ into f; that is

$$f=\mathbf{e}+i\mathbf{b}=W\hat{\mathbf{e}}\tilde{W}=W^2\hat{\mathbf{e}}=\hat{\mathbf{e}}\tilde{W}^2. \tag{2.29}$$

Solving for W^2, one finds

$$wu=W^2=f\hat{\mathbf{e}}=|\mathbf{e}|+i\mathbf{b}\hat{\mathbf{e}}=|\mathbf{e}|\left(1+\frac{\mathbf{e}\times\mathbf{b}}{\mathbf{e}^2}\right)\equiv\gamma_w(1+\mathbf{w}/c). \tag{2.30a}$$

Thus, with the help of (1.23)

$$\frac{\mathbf{w}}{c}=\frac{\mathbf{e}\times\mathbf{b}}{\mathbf{e}^2}=\frac{2\mathbf{E}\times\mathbf{B}}{|z|^2+\mathbf{E}^2+\mathbf{B}^2} \tag{2.30b}$$

and

$$\gamma_w=|\mathbf{e}|=(1+\mathbf{w}^2/c^2)^{-1/2}. \tag{2.30c}$$

Note that $fw=\hat{\mathbf{e}}u=|f\cdot u|^{-1}f\cdot u$, hence

$$f\cdot w=|f\cdot u|^{-1}f\cdot u, \tag{2.31a}$$

and, more important,

$$f\wedge w=0. \tag{2.31b}$$

As noted earlier, the condition (2.31b) implies that the field $F=f(\alpha+i\beta)$ will consist of parallel electric and magnetic fields relative to an observer with proper velocity $\mathbf{w}$. For this reason, the corresponding relative vector w is called the *relative drift velocity*. It is important to realize that $w=|f\cdot u|^{-1}f(f\cdot u)$ does not describe an intrinsic property of the electromagnetic field; rather, it describes a relation of the observer u to the field $F=fz$.

Now introduce an electromagnetic field

$$F'=\hat{\mathbf{e}}(\alpha+i\beta)=\mathbf{E}'+i\mathbf{B}', \tag{2.32a}$$

where α, β and $\hat{\mathbf{e}}$ are defined as before by (2.22a, b) and (2.23b). Then, from (2.29) and (2.4) it follows that

$$F=WF'\tilde{W}. \tag{2.32b}$$

Hence, from (2.1) it follows that

$$R=WR'\tilde{W} \text{ and } \tilde{R}=W\tilde{R}'\tilde{W} \tag{2.33a}$$

where

$$R'=\exp(F'\lambda\tau/2). \tag{2.33b}$$

Therefore, the equation (2.3a) for the proper velocity v of the electron can be written

$$v=Rv_0R=WR'\tilde{W}v_0W\tilde{R}'\tilde{W}=Wv'\tilde{W}, \tag{2.34a}$$

where

$$v'=R'v_0'\tilde{R}' \tag{2.34b}$$

and

$$v_0'=\tilde{W}v_0W. \tag{2.34c}$$

Now v' is the proper velocity of an electron with initial velocity v_0' accelerated by parallel fields $\mathbf{E}'$ and $\mathbf{B}'$ relative to the observer u, so explicit expressions for $v'u=\gamma'(1+v'/c)$ are known from the special case analyzed earlier. To get corresponding expressions for vu (2.34a) and (2.30a); thus

$$vu=\gamma(1+\mathbf{v}/c)=Wv'\tilde{W}u=Wv'uW=W\gamma(1+\mathbf{v}'/c)\tilde{W},$$

the scalar part of which is

$$\gamma=\gamma'\gamma_w[(1+\mathbf{v}'\cdot\mathbf{w})/c^2], \tag{2.35a}$$

while the ratio of vector to scalar part is

$$\mathbf{v}=\frac{\mathbf{v}'+\mathbf{w}+(\gamma_w^{-1}-1)\hat{\mathbf{w}}x(\mathbf{v}'\times\hat{\mathbf{w}})}{1+c^{-2}\mathbf{w}\cdot\mathbf{v}'}, \tag{2.35b}$$

the well-known velocity addition formula. Of course a similar formula will express $\mathbf{v}_0'$ in terms of $\mathbf{v}_0$ and $\mathbf{w}$. Also in a similar fashion, the general orbit can be found from the orbit of a particle in parallel fields by a boost in the direction of the drift velocity or by integrating (2.35). The formulas are easily worked out, and of course they will agree with (2.28), but now the general nature of the orbit is easily described; it consists of a tightening spiral in the relative direction $\hat{\mathbf{e}}$ [determined by (2.23b)] drifting with velocity $\mathbf{w}$ [given by (2.30b)] in a direction orthogonal to $\hat{\mathbf{e}}$.

3. RIGID TEST CHARGE IN A PLANE WAVE FIELD

The equations of motion for a rigid test charge in an electromagnetic plane wave will now be integrated.

Any plane wave field $F=F(x)$ with proper propagation vector k can be written in the canonical form

$$F=fz, \tag{3.1a}$$

where f is a constant bivector and the x-dependence of F is exhibited explicitly by

$$z=\alpha_+\exp(ik\cdot x)+\alpha_-\exp(-ik\cdot x). \tag{3.1b}$$

As explained in Ref. 3, $\alpha_\pm$ are the "complex" amplitudes for right and left circular polarization. Here "complex" means "having only scalar and pseudoscalar parts," i.e.,

$$\alpha_\pm=[\alpha_\pm]_0+[\alpha_\pm]_4=\rho_\pm\exp(\pm i\delta_\pm) \tag{3.1c}$$

where $\delta_\pm$ and $\rho_\pm>0$ are scalars. In contrast to the usual use of complex numbers in electromagnetic theory, the "unit imaginary" i, being the unit pseudoscalar, has a definite geometrical significance. Maxwell's equation $\Box F=0$ implies, since $\Box k\cdot x=k$,

$$kf=0, \quad \text{or equivalently,} \quad kF=0. \tag{3.1d}$$

Multiplying by k, one ascertains that

$$k^2=0. \tag{3.1e}$$

It can be shown further that f must have the form

$$f=ka=k\wedge a=-ak, \tag{3.1f}$$

where a is a unit spacelike vector orthogonal to k. When $\alpha_\pm$ have been specified, a is uniquely determined, but a rotation of a preserving $k\cdot a=0$ can be compensated by an overall phase change of α_+ and α_- (corresponding to a gauge transformation of the electromagnetic vector potential), so to this extent factorization of F into f and z is not unique.

Before the spinor Lorentz force $\dot{R}=\frac{1}{2}\lambda FR$ can be integrated, it is necessary to express $F=F(x)$ as a function of τ. This can be done by using special properties of F to find constants of motion. Using (3.1d), one finds

$$\frac{d}{d\tau}(kR)=k\dot{R}=\tfrac{1}{2}\lambda kFR=0, \tag{3.2}$$

that is, kR is a constant of motion. So, using the initial condition $R(0)=1$, one finds

$$k=kR=Rk=k\widetilde{R}. \tag{3.3}$$

The second equality in (3.3) follows from the first by reversion: $Rk=R\widetilde{k}=R(\widetilde{R}\widetilde{k})=k$. From (3.3) it follows that

$$Rk\widetilde{R}=k. \tag{3.4}$$

Therefore, $R=R(\tau)$ is a family of Lorentz rotations leaving the lightlike vector k invariant. Multiplying $e_\mu=R\gamma_\mu\widetilde{R}$ by k, Eq. (3.4) gives constants of motion for the e_μ:

$$k\cdot e_\mu=k\cdot\gamma_\mu; \tag{3.5}$$

in particular,

$$k\cdot v=k\cdot v_0. \tag{3.6}$$

Since $v=dx/d\tau$, this integrates to

$$k\cdot(x(\tau)-x_0)=k\cdot v_0\tau. \tag{3.7}$$

This is precisely the relation needed to express the electromagnetic field acting on the particle as a function of the proper time. Substituting (3.7) into (3.1b), one obtains

$$z=z(\tau)=\alpha_+\exp(i\omega_0\tau)+\alpha_-\exp(-i\omega_0\tau), \tag{3.8}$$

where $\omega_0\equiv k\cdot v_0$ is the frequency of the plane wave relative to an observer with proper velocity v_0, and an overall phase $\delta_0=k\cdot x_0$ has been absorbed into the phases of α_+ and α_- [or equivalently into the definition of a in (3.1f)].

Now, by (3.1) and (3.3), the spinor Lorentz force for a plane wave has the form

$$\frac{dR}{d\tau}=\tfrac{1}{2}\lambda FR=\tfrac{1}{2}\lambda F=\tfrac{1}{2}\lambda fz. \tag{3.9}$$

With the initial $R(0)=1$ and the expression (3.8) for $z=z(\tau)$, this integrates immediately to

$$R=1+\tfrac{1}{2}\lambda fz_1=\exp(\lambda fz_1/2), \tag{3.10a}$$

where

$$z_1\equiv\int_0^\tau z(\tau)d\tau=\frac{2}{\omega_0}\sin\tfrac{1}{2}\omega_0\tau(\alpha_+\exp(i\omega_0/2)-\alpha_-\exp(-i\omega_0\tau/2)). \tag{3.10b}$$

Hence the expression for the comoving frame as a function of τ is

$$\begin{aligned}e_\mu&=R\gamma_\mu\widetilde{R}=(1+\tfrac{1}{2}\lambda fz_1)\gamma_\mu(1-\tfrac{1}{2}\lambda fz_1)\\&=\gamma_\mu+\lambda\tfrac{1}{2}(fz_1\gamma_\mu-\gamma_\mu fz_1)-\lambda^2\tfrac{1}{4}fz_1\gamma_\mu fz_1,\end{aligned}$$

or, since $z_1\gamma_\mu=z_1^*\gamma_\mu$,

$$e_\mu=\gamma_\mu+\lambda(fz_1)\cdot\gamma_\mu-\lambda^2\Theta_1\tfrac{1}{2}f\gamma_\mu f, \tag{3.11a}$$

where, recalling (3.1c) and writing $\delta=\delta_++\delta_-$,

$$\Theta_1\equiv\tfrac{1}{2}|z_1|^2=\frac{\sin^2\frac{1}{2}\omega_0\tau}{2\omega_0^2}[\rho_+^2+\rho_-^2-2\rho_+\rho_-\cos(\omega_0\tau+\delta)]. \tag{3.11b}$$

Notice that in (3.12) $-\frac{1}{2}f\gamma_\mu f=\frac{1}{2}f(-f\gamma_\mu+f\cdot\gamma_\mu)=ff\cdot\gamma_\mu$, which is proportional to the component of γ_μ in the f plane.

According to (3.11a), the equation for the proper velocity is

$$\frac{dx}{d\tau}=v_0+\lambda(fz_1)\cdot v_0+\lambda^2\Theta_1 ff\cdot v_0, \tag{3.12}$$

which integrates to a parametric equation for the particle history

$$x(\tau)-x_0=v_0\tau+\lambda[fz_2]\cdot v_0+\lambda^2\Theta_2 ff\cdot v_0, \tag{3.13a}$$

where

$$z_2\equiv\int_0^\tau z_1(\tau)d\tau=-\frac{z}{\omega_0^2}+\frac{(\alpha_++\alpha_-)}{\omega_0^2}+\frac{(\alpha_+-\alpha_-)}{i\omega_0} \tag{3.13b}$$

and

$$\Theta_2\equiv\int_0^\tau\Theta_1(\tau)d\tau=\frac{1}{\omega_0^2}\Big((\rho_+^2+\rho_-^2+2\rho_+\rho_-)\frac{\sin\omega_0\tau}{\omega_0}$$

$$-\rho_+\rho_-\frac{\sin(2\omega_0\tau+\delta)}{2\omega_0}+(\rho_+^2+\rho_-^2+\rho_+\rho_-\cos\delta)\tau$$
$$+\frac{3}{2\omega_0}\sin\delta\Big). \tag{3.13c}$$

This completes the explicit solution of a rigid point charge in a plane wave. If desired, the equations can be put in relative form by the method illustrated in the last section.

4. RIGID TEST CHARGE IN A COULOMB FIELD

The spinor Lorentz force will now be integrated to describe the motion of a test charge e in the Coulomb field of a "fixed nucleus" with charge $-Ze$. Let u be the constant proper velocity of the nucleus. In terms of relative variables the Coulomb field is

$$\lambda F=\frac{e}{mc^2}\mathbf{E}=-k\frac{\mathbf{x}}{|x|^3}=\nabla\frac{k}{|\mathbf{x}|}, \tag{4.1a}$$

where

$$k=Ze\lambda/4\pi=Ze^2/4\pi mc^2. \tag{4.1b}$$

In terms of proper variables the Coulomb field is

$$\lambda F=-k\frac{x\wedge u}{|x\wedge u|^3}=\Box\left(\frac{-ku}{|x\wedge u|}\right)=u\wedge\Box\left(\frac{k}{|x\wedge u|}\right), \tag{4.1c}$$

where, of course, $x=x(\tau)$ is the position of the test particle at time τ and the origin $x=0$ has been located at some point on the history of the nucleus.

Before the spinor equation $\mathring{R}=\frac{1}{2}\lambda FR$ can be integrated, it is necessary to express F as a parametric function of the particle history. This can be done by reexpressing symmetry properties of F in terms of constants of motion. The constants of motion can be found by multiplying F by the available vectors x, u, v and using the Lorentz force.

"Dotting" (4.1c) by u, one finds, with the help of (I.1.5),

$$\lambda u\cdot F=(\Box-uu\cdot\Box)\frac{k}{|x\wedge u|}=\Box\left(\frac{k}{|x\wedge u|}\right), \tag{4.2}$$

since $u\cdot\Box|x\wedge u|^{-1}=c^{-1}\partial_t|x|^{-1}=0$. So, dotting the Lorentz force $\mathring{v}=\lambda F\cdot v$ by u, one finds

$$\frac{d}{d\tau}(u\cdot v)=u\cdot\mathring{v}=\lambda u\cdot F\cdot v=v\cdot\Box\left(\frac{k}{|x\wedge u|}\right)=\frac{d}{d\tau}\left(\frac{k}{|x\wedge u|}\right).$$

Hence,

$$W\equiv u\cdot v-k/|x\wedge u|=\gamma-k/|\mathbf{x}| \tag{4.3}$$

is a constant of motion. The sum of particle kinetic and potential energies is $E=mc^2(W-1)=mc^2(\gamma-1)-Ze^2/4\pi|\mathbf{x}|$.

From (4.1c) it follows that the Coulomb field (in fact any central field) has the properties

$$F\wedge u=0, \tag{4.4a}$$

$$F\wedge x=0. \tag{4.4b}$$

By virtue of (I.1.8), it follows that $(F\wedge u)\cdot v=Fu\cdot v-(F\cdot v)\wedge u=0$, which, on substituting $\lambda F\cdot v=\mathring{v}$, gives

$$\frac{d}{d\tau}(v\wedge u)=\lambda Fv\cdot u. \tag{4.5a}$$

In terms of relative variables, this is just the usual equation for an electric force on a particle, i.e.,

$$\frac{d}{d\tau}(\gamma\mathbf{v})=c^2\lambda\mathbf{E}=\frac{e}{m}\mathbf{E}. \tag{4.5b}$$

Now applying (4.4b) to (4.5a), one finds

$$\frac{d}{d\tau}(v\wedge u\wedge x)=0, \tag{4.6}$$

hence the dual of the trivector $v\wedge u\wedge x$,

$$l=iv\wedge u\wedge x, \tag{4.7}$$

is also a constant of motion. Using the general duality relation $(iT)\cdot x=iT\wedge x$, one obtains immediately from (4.7)

$$l\cdot x=l\cdot u=l\cdot v=0. \tag{4.8}$$

In terms of the proper vector l, one can define a relative vector $\mathbf{l}\equiv l\wedge u$ which is obviously also a constant of motion. From (4.8) and (4.7)

$$\mathbf{l}\equiv l\wedge u=lu=-i(v\wedge x\wedge u)\cdot u=-i[v\wedge x]_2. \tag{4.9a}$$

Since

$$vx=(vu)(ux)=\gamma(1+\mathbf{v}/c)(ct-\mathbf{x})$$
$$=\gamma(ct-x)\cdot\left(\frac{\mathbf{v}}{c}\right)+\gamma(\mathbf{v}t-\mathbf{x})-\frac{\gamma}{c}\mathbf{v}\wedge\mathbf{x},$$

so $[v\wedge x]_2=c^{-1}\gamma\mathbf{x}\wedge\mathbf{v}=c^{-1}\gamma i\mathbf{x}\times\mathbf{v}$, and (4.9a) yields

$$\mathbf{l}\equiv\frac{\gamma}{c}\mathbf{x}\times\mathbf{v}=\mathbf{x}\times\frac{dx}{dt}=\frac{\mathbf{x}\times\mathbf{p}}{mc}. \tag{4.9b}$$

Thus $\mathbf{l}$ is the usual (relative) angular momentum per unit mc.

The constants of motion have been found; the problem now is to use them effectively. From (4.9b) one finds

$$\mathbf{l}\cdot\mathbf{v}=\mathbf{l}\cdot\mathbf{x}=0, \tag{4.10}$$

which says that the relative motion is in a plane orthogonal to $\mathbf{l}$. The unit bivector $i\hat{\mathbf{l}}$ is the generator of rotations in that plane. Hence one can write

$$\hat{\mathbf{x}}=\frac{\mathbf{x}}{|\mathbf{x}|}=\mathbf{a}\exp(i\hat{\mathbf{l}}\Theta), \tag{4.11a}$$

where $\mathbf{a}\cdot\mathbf{l}=0$ and $\Theta(\tau)$ is the angle of rotation of a fixed unit vector $\mathbf{a}$ into the direction $\hat{\mathbf{l}}$. The requirement $\mathring{\Theta}=d\Theta/d\tau>0$ entails that the rotation has the same "sense" as the particle motion. The sense of the rotation is described by the vector

$$\frac{d\hat{\mathbf{x}}}{d\Theta}=\mathbf{b}\exp(i\hat{\mathbf{l}}\Theta), \tag{4.11b}$$

where

$$\mathbf{b}=\mathbf{a}i\hat{\mathbf{l}}=\mathbf{a}\cdot(i\hat{\mathbf{l}})=i\mathbf{a}\wedge\hat{\mathbf{l}}=\hat{\mathbf{l}}\times\mathbf{a}. \tag{4.11c}$$

Thus, the vectors $\mathbf{a}$, $\mathbf{b}$, $\hat{\mathbf{l}}$ form a right-handed orthonormal frame. So do the vectors $\mathbf{x}$, $d\hat{\mathbf{x}}/d\Theta$, $\hat{\mathbf{l}}$, since

$$\hat{\mathbf{x}}\frac{d\hat{\mathbf{x}}}{d\Theta}=\mathbf{a}\exp(i\hat{\mathbf{l}}\Theta)\mathbf{b}\exp(i\hat{\mathbf{l}}\Theta)=\mathbf{ab}=i\hat{\mathbf{l}}=i\hat{\mathbf{x}}\times\frac{d\hat{\mathbf{x}}}{d\Theta}. \tag{4.11d}$$

Conservation of $\mathbf{l}$ implies that the direction and the magnitude of $\mathbf{l}$ are conserved separately. The implication of the directional conservation has been expressed by (4.11). The implication of the magnitude conservation can be obtained by substituting

$$\frac{\gamma \mathbf{v}}{c}=\frac{d\mathbf{x}}{d\tau}=\frac{\Theta d\mathbf{x}}{d\Theta}=\Theta\left(\mathbf{x}\frac{d|\mathbf{x}|}{d\Theta}+|\mathbf{x}|\frac{d\hat{\mathbf{x}}}{d\Theta}\right) \tag{4.12}$$

into (4.9b) and using (4.11d); thus

$$1=|\mathbf{x}|^2\,\mathring{\Theta}\mathbf{x}\times\frac{d\hat{\mathbf{x}}}{d\Theta}=|\mathbf{x}|^2,$$

or

$$|1|=|\mathbf{x}|^2\mathring{\Theta}. \tag{4.13}$$

Now, with (4.13) and (4.11a), the Coulomb field (4.1a) can be put in the simple parametric form

$$\lambda F=-\kappa\hat{\mathbf{x}}\mathring{\Theta}=-\kappa\mathbf{a}\exp(i\hat{1}\Theta)\mathring{\Theta} \tag{4.14a}$$

where

$$\kappa=k/|1|=Ze^2/4\pi mc|1|. \tag{4.14b}$$

Hence on changing variables from τ to Θ, the spinor equation $\mathring{R}=\frac{1}{2}\lambda FR$ assumes the simple form

$$\frac{dR}{d\Theta}=-\frac{\kappa}{2}\hat{\mathbf{x}}R=-\frac{\kappa}{2}\mathbf{a}\exp(i\hat{1}\Theta)R \tag{4.15}$$

Of course (4.14) and (4.15) assume $|1|\neq 0$; the case $1=0$ is easily integrated separately, since then the direction of the field is a constant of motion.

To solve (4.15), guess that the solution has the general form

$$R=\exp(-B\Theta/2)\exp(-A\Theta/2)R_0 \tag{4.16a}$$

where B, A, and R_0 are independent of Θ, and, to satisfy the conditions that R be even and $R\tilde{R}=1$, B and A must be proper bivectors and $R_0\tilde{R}_0=1$. It may be noted that no generality is gained by adding "phases" to the angles in (4.16a) since they can be "absorbed" in the definitions of the constants A, B, R_0. Substituting (4.16a) into (4.15) one obtains conditions on A and B; thus

$$-\kappa\mathbf{a}\exp(i\hat{1}\Theta)=2\frac{dR}{d\Theta}\tilde{R}=-B-\exp(-B\Theta/2)A\exp(B\Theta/2)$$
$$=-B-A_+-A_-\exp(B\Theta),$$

where to carry out the last step, A has been expressed as the sum of a part A_+ which commutes with B and a part A_- which anticommutes with B. Equating independent parts of the equation, one finds

$$B=i\hat{1}, \tag{4.16b}$$

$$A=\kappa\mathbf{a}-i\hat{1}. \tag{4.16c}$$

Hence (4.16a) subject to (4.16b, c) is a general solution of (4.15). The form of this solution is peculiar to the Coulomb field and does not apply to any other central field.

The "initial value" R_0 of the *Coulomb spinor* (4.16a) can be written $R_0=L_0U_0$ where L_0 determines a boost and U_0 a spatial rotation. By an appropriate choice of the initial conditions for the comoving frame, the spinor U_0 can be set equal to unity. The spinor L_0 is determined from the velocity v_0 at $\Theta=0$ by the equation

$$v_0=R_0u\tilde{R}_0=L_0u\tilde{L}_0=L_0^2u$$

or

$$v_0u=L_0^2=\gamma_0\left(1+\frac{\mathbf{v}_0}{c}\right)\quad\text{where } \gamma_0=(1-\mathbf{v}_0^2/c^2)^{-1/2}. \tag{4.17}$$

[The use of the symbol γ_0 in (4.17) should not be confused with the use of the same symbol to represent a vector elsewhere in this paper.] According to (4.10) $1\cdot\mathbf{v}_0=0$, although it is not necessary, it is convenient to require also $\mathbf{a}\cdot\mathbf{v}_0=\hat{\mathbf{x}}_0\cdot\mathbf{v}_0=0$; so by (4.11)

$$\mathbf{v}_0=|\mathbf{v}_0|\mathbf{b}. \tag{4.18}$$

This eliminates previous arbitrariness in the choice of the direction $\mathbf{a}$ and the zero for Θ. The constants of motion $|1|$ and W can be expressed in terms of the initial values $|\mathbf{v}_0|$ and $|\mathbf{x}_0|$ and vice versa. Because of (4.18), (4.9) and (4.11) imply

$$|1|=|\mathbf{x}_0|\frac{|d\mathbf{x}_0|}{d\tau}=|\mathbf{x}_0|\gamma_0\frac{|\mathbf{v}_0|}{c}. \tag{4.19a}$$

Using this in (4.3) one obtains

$$W=\gamma_0\left(1-\kappa\frac{\mathbf{v}_0}{c}\right)=\gamma_0^{-1}-\kappa(\gamma_0^2-1)^{1/2} \tag{4.19b}$$

Solving (4.19b) for $|\mathbf{v}_0|$ and γ_0, one obtains

$$\frac{|\mathbf{v}_0|}{c}=\frac{\kappa\pm W\sqrt{W^2+\kappa^2-1}}{W^2+\kappa^2}, \tag{4.20a}$$

$$\gamma_0=\frac{W\pm\kappa(W^2+\kappa^2-1)^{1/2}}{1-\kappa^2}, \tag{4.20b}$$

The physical roots must, of course, satisfy the condition $0<|\mathbf{v}_0|<c$.

The Coulomb spinor (4.16a) gives immediately the explicit expression for the particle proper velocity

$$v=Ru\tilde{R}=\exp(-B\Theta/2)\exp(-A\Theta/2)v_0\exp(A\Theta/2)\exp(B\Theta/2). \tag{4.21}$$

Three classes of motion can be distinguished: when A^2 is zero, positive or negative. If $A^2=0$ then $\exp\frac{1}{2}A\Theta=1+\frac{1}{2}A\Theta$ and the motion is most easily analyzed by substituting this in (4.21). If $A^2\neq 0$ the motion is most easily analyzed by decomposing v_0 into a component

$$v_{0\parallel}=\frac{v_0\cdot AA}{A^2} \tag{4.22a}$$

in the A plane and a component

$$v_{0\perp}=\frac{v_0\wedge AA}{A^2}=-\frac{v_0\cdot(iA)iA}{A^2} \tag{4.22b}$$

orthogonal to the A plane, so (4.21) becomes

$$v=\exp(-B\Theta/2)(v_{0\perp}+v_{0\parallel}\exp(A\Theta))\exp(B\Theta/2). \tag{4.23}$$

From (4.16c) one finds $A^2=\kappa^2-1$. If $\kappa^2>1$, then $\exp A\Theta=\cosh|A|\Theta+|A|^{-1}A\sinh|A|\Theta$ where $|A|=(\kappa^2-1)^{1/2}$; this is characteristic of the "scattering states" of the particle. If $\kappa^2<1$, then

$$\exp(A\Theta)=\cos|A|\Theta+\hat{A}\sin|A|\Theta, \tag{4.24}$$

where

$$|A|=(1-\kappa^2)^{1/2}\quad\text{and}\quad\hat{A}=|A|^{-1}A. \tag{4.25}$$

This is a necessary (but not sufficient) condition for "bound states" of the particle. In the following, this case will be studied in more detail.

To find an expression for the relative velocity of the particle, substitute (4.24) into (4.23) and note that by (4.16b) that u commutes with B, so

$$vu=\gamma(1+\mathbf{v}/c)=\exp(-B\Theta/2)\Big(v_0+v_0\cos|A|\Theta + \frac{v_0\cdot A}{|A|}\sin|A|\Theta\Big)u\exp(-B\Theta/2). \tag{4.26}$$

The terms involving v_0 can be put in relative form with the help of the general formula (I.2.15); thus, since $A = \kappa\mathbf{a} - i\hat{1}$ and $v_0u=\gamma_0(1+\mathbf{b}|\mathbf{v}_0|/c)$ where $\mathbf{ab}=i\hat{1}$, one obtains

$$\begin{aligned} v_0\cdot Au &= -A\cdot v_0u = -\gamma_0\Big(\kappa\frac{\mathbf{a}\cdot\mathbf{v}_0}{c}+\kappa\mathbf{a}-\hat{i1}\frac{\mathbf{v}_0}{c}\Big)\\ &= \gamma_0\Big(-\kappa+\frac{|\mathbf{v}_0|}{c}\Big)\mathbf{a}=\pm(W^2-|A|^2)^{1/2}\mathbf{a}, \end{aligned} \tag{4.27}$$

where in the last step (4.19b) and (4.20b) were used to convert initial values to constants of motion. The two signs in (4.27) correspond to an arbitrariness in the choice of orientation of $\mathbf{a}$ and $\mathbf{b}$. It is convenient to choose the *positive* sign. Repeating the procedure which lead to (4.27) with $iA=\hat{1}+i\kappa\mathbf{a}$ instead of A, one obtains

$$\begin{aligned} (iA)\cdot v_0u &= \gamma_0\Big(\hat{1}+i\kappa\mathbf{ab}\frac{|\mathbf{v}_0|}{c}\Big)\\ &= \gamma_0\Big(1-\kappa\frac{|\mathbf{v}_0|}{c}\Big)\hat{1}=W\hat{1}. \end{aligned} \tag{4.28}$$

From (4.27) one gets

$$\begin{aligned} v_{0\parallel}u &= \frac{A(A\cdot v_0)u}{-|A|^2}=\frac{(W^2-|A|^2)^{1/2}}{|A|^2}(\kappa\mathbf{a}-i\hat{1})\mathbf{a}\\ &= \frac{(W^2-|A|^2)^{1/2}}{|A|^2}(\kappa+\mathbf{b}) \end{aligned} \tag{4.29}$$

and from (4.28)

$$v_{0\perp}u=\frac{-iA(iA)\cdot v_0}{-|A|^2}=\frac{W}{|A|^2}(\hat{1}+i\kappa\mathbf{a})\hat{1}=\frac{W}{|A|^2}(1+\kappa\mathbf{b}). \tag{4.30}$$

Substituting (4.27, 29, 30) into (4.36), one has

$$\begin{aligned} \gamma\Big(1+\frac{\mathbf{v}}{c}\Big) &= \exp(-B\Theta/2)\Big(\frac{W}{|A|^2}(1+\kappa\mathbf{b})\\ &\quad+\frac{(W^2-|A|^2)^{1/2}}{|A|^2}(\kappa+\mathbf{b})\cos|A|\Theta\\ &\quad+\frac{(W^2-|A|^2)^{1/2}}{|A|}\mathbf{a}\sin|A|\Theta\Big)\exp(B\Theta/2). \end{aligned} \tag{4.31}$$

The scalar part of (4.31) is

$$\gamma=\frac{1}{|A|^2}(W+\kappa(W^2-|A|^2)^{1/2}\cos|A|\Theta), \tag{4.32}$$

while the vector part of (4.31) is

$$\begin{aligned} \frac{d\mathbf{x}}{d\tau} &= \frac{\gamma\mathbf{v}}{c}=\Big(\frac{\mathbf{b}}{|A|^2}(\kappa W+(W^2-|A|^2)^{1/2}\cos|A|\Theta)\\ &\quad+\mathbf{a}\frac{(W^2-|A|^2)^{1/2}}{|A|}\sin|A|\Theta\Big)\exp(B\Theta). \end{aligned} \tag{4.33}$$

This is the desired equation for the relative velocity.

An equation for the orbit of the particle can be obtained immediately from the radial component of (4.33) without integration. Using (4.11) and (4.13) in (4.12), one gets

$$\frac{d\mathbf{x}}{d\tau}=|1|\Big(\frac{1}{|\mathbf{x}|}\mathbf{b}-\mathbf{a}\frac{d}{d\Theta}\Big(\frac{1}{|\mathbf{x}|}\Big)\Big)\exp(B\Theta). \tag{4.34}$$

Equating (4.34) to (4.33), one obtains from the radial component

$$\frac{|1||A|^2}{|\mathbf{x}|}=\kappa W+(W^2-|A|^2)^{1/2}\cos|A|\Theta. \tag{4.35}$$

This is the well-known equation for a precessing ellipse derived long ago by Sommerfeld.

When $\kappa<1$, the bivector $-A=-\kappa\mathbf{a}+i\hat{1}$ can be obtained from $B=i\hat{1}$ by a boost. Thus, as in (2.29) and (2.30)

$$-A=-\kappa\mathbf{a}+i\hat{1}=|A|Ki\hat{1}\tilde{K}=|A|K^{-2}i\hat{1}. \tag{4.36}$$

So

$$K^2=|A|^{-1}Ai\hat{1}=|A|^{-1}(1+\kappa\mathbf{a}i\hat{1})$$

or, by (4.11c),

$$K^2=|A|^{-1}(1+\kappa\mathbf{b}). \tag{4.37}$$

Notice that K produces a boost in the direction of the initial velocity $\mathbf{v}_0=|\mathbf{v}_0|\mathbf{b}$. Indeed, from (4.34) it is obvious that the orbit is circular if $W^2=|A|^2$, and according to (4.27) this is equivalent to the condition $\kappa = |\mathbf{v}_0|/c$, which implies that K is equal to the initial boost L_0 in (4.17). Using (4.36) the Coulomb spinor, (4.16a) can be put in the form

$$R=\exp(-B\Theta/2)K\exp(B|A|\Theta/2)\tilde{K}L_0, \tag{4.38}$$

which, for circular motion, reduces to

$$R=\exp(-B\Theta/2)K\exp(B|A|\Theta/2)=K'\exp(-(1-|A|)B\Theta/2), \tag{4.39}$$

where

$$K'\equiv\exp(-B\Theta/2)K\exp(B\Theta/2).$$

The right side of (4.34) displays R factored into a boost by K' preceded by a spatial rotation through an angle $(1-|A|)$, which evaluated for a period gives the Thomas precession immediately. The Thomas precession for arbitrary angular momentum can be obtained algebraically by factoring $\exp(-\frac{1}{2}B\Theta)\exp(-\frac{1}{2}A\Theta)$ into a boost preceded by a spatial rotation.

ACKNOWLEDGMENT

A solution to (4.15) equivalent to the Coulomb spinor (4.38) was first found by L. Cummings (unpublished).

[1]D. Hestenes, J. Math. Phys. 15, preceding (1974).
[2]D. Hestenes, *Spacetime Algebra* (Gordon and Breach, New York, 1966).
[3]D. Hestenes, Am. J. Phys. 39, 1013 (1971).
[4]F. Gursey, Nuovo Cimento 5, 784 (1957).
[5]D. Hestenes, J. Math. Phys. 14, 893 (1973).
[6]V. Bargmann, L. Michel, and V. Telegdi, Phys. Rev. Lett. 2, 435 (1959).
[7]L. Thomas, Phil. Mag. 3, 1 (1927).
[8]A. Taub, Rev. Mod. Phys. 21, 388 (1949).
[9]I. Plebanski, in *Perspectives in Geometry and Relativity*, edited by B. Hoffmann (Indiana U.P., Bloomington, 1966), p. 310.

Complete sets of commuting operators and $O(3)$ scalars in the enveloping algebra of $SU(3)$

B. R. Judd*, W. Miller Jr.†, J. Patera, and P. Winternitz

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, P.Q., Canada
(Received 13 December 1973)

We consider the "missing label" problem for basis vectors of $SU(3)$ representations in a basis corresponding to the group reduction $SU(3) \supset O(3) \supset O(2)$. We prove that only two independent $O(3)$ scalars exist in the enveloping algebra of $SU(3)$, in addition to the obvious ones, namely the angular momentum L^2 and the two $SU(3)$ Casimir operators $C^{(2)}$ and $C^{(3)}$. Any one of these two operators (of third and fourth order in the generators) can be added to $C^{(2)}$, $C^{(3)}$, L^2, and L_3 to form a complete set of commuting operators. The eigenvalues of the third and fourth order scalars $X^{(3)}$ and $X^{(4)}$ are calculated analytically or numerically for many cases of physical interest. The methods developed in this article can be used to resolve a missing label problem for any semisimple group G, when reduced to any semisimple subgroup H.

1. INTRODUCTION

The general problem that we touch upon in this article is that of providing a complete labeling for the states transforming under an irreducible representation of a given Lie group G. In a certain sense this problem has been completely solved for the classical semisimple groups,[1] corresponding to the Cartan algebras A_n, B_n, C_n, and D_n. Indeed the Gel'fand–Tseitlin patterns[2] provide us precisely with such a set of labels, and the corresponding "canonical basis" consists of a complete nondegenerate set of orthonormal basis functions. The basis functions are the common set of eigenfunctions of a complete set of commuting operators, consisting of the Casimir operators of the group G and of all the Casimir operators of a "canonical" chain of subgroups of G. Thus, e. g., for the group $SU(n)$ the canonical chain is

$$SU(n) \supset S[U(n-1) \times U(1)] \supset S[U(n-2) \times U(1) \times U(1)] \supset \cdots \supset S[U(1) \times \cdots \times U(1) \times U(1)] \qquad (1)$$

so that the complete set of commuting operators consists of all the Casimir operators of $SU(n)$, $SU(n-1)$, $\ldots, SU(2)$ and of the $(n-1)$ linear operators (the Cartan subalgebra), corresponding to the $U(1)$ subgroups. Similarly, the problem is solved for the orthogonal and symplectic groups (and also for some of the noncompact groups, corresponding to the same algebras[3]).

Unfortunately, in physics one is often interested in other operators, which may correspond to subgroups, not figuring in the canonical reduction, or may lie in the enveloping algebra of the Lie algebra of G, without being Casimir operators of any subgroup of G. Hence it is important to study other bases and indeed to perform a systematic study of possible bases for representations of various Lie groups.

In this article we restrict ourself to a very simple case, which is, however, of considerable physical interest, namely the group $SU(3)$. The standard application of $SU(3)$ in particle physics, namely the "eightfold way"[4] does indeed make use of the canonical chain of subgroups $SU(3) \supset S[U(2) \times U(1)] \supset S[U(1) \times U(1)]$. However, in nuclear physics[5-7] and more generally in group theoretical treatments of the many-body problem,[8] the quantity of prime interest is angular momentum, associated with the group $O(3)$ that is imbedded into $SU(3)$ in an irreducible manner [this $O(3)$ is the intersection of $SU(3)$ and $SL(3,R)$]. The corresponding chain of subgroups is

$$SU(3) \supset O(3) \supset O(2). \qquad (2)$$

Basis functions of $SU(3)$, corresponding to the reduction (2) are eigenfunctions of the second $C^{(2)}$ and third $C^{(3)}$ order Casimir operators of $SU(3)$ and of the angular momentum operators L^2 and L_3. There is one label missing to characterize the states completely and indeed there can be more than one state, characterized by given $O(3)$ quantum numbers (l, m) within a given representation (k_1, k_2) of $SU(3)$. Several different methods have been proposed to resolve this degeneracy problem, and they can be divided into two classes.

The first type of solution leads to a simple labeling of the states (by integers), but to nonorthogonal basis functions that are not eigenfunctions of any complete set of commuting operators.[5,6,9] The other type of solution of the degeneracy problem for $O(3)$ states in $SU(3)$ representations leads to orthonormal states, that are eigenfunctions of $C^{(2)}$, $C^{(3)}$, L^2, L_3 and an additional Hermitian operator X in the enveloping algebra of $SU(3)$.[6,10] The eigenvalues of X provide the missing label for the state vectors; they are, however, not integer numbers and must in general be obtained by solving certain algebraic equations. What is more, Racah has proven[10] that it is not possible to construct any operator in the enveloping algebra of $SU(3)$ that would resolve this missing label problem and have integer eigenvalues.

The purpose of this article is to investigate further the second of the above approaches, that is, in general to study all possible complete sets of commuting operators, the eigenfunctions of which will provide an orthonormal basis for the representations of the group G [in this case $G = SU(3)$]. Investigations along these lines have been carried out,[11] e. g., for the rotation groups $O(3)$ and $O(4)$, the Euclidean groups $E(2)$ and $E(3)$, and the Lorentz groups $O(2,1)$ and $O(3,1)$. Each nonequivalent complete set of commuting operators (consisting of operators from the enveloping algebra of the given algebra that may or may not be Casimir operators of subalgebras, and possibly of some further reflection type operators) provides us with a different set of basis functions. In particular the "nonsubgroup" type opera-

 Copyright © 1974 American Institute of Physics

tors lead to the appearance of many new types of special functions in group theoretical studies[11,12] (e.g., Lamé and Heun functions).

In this article we consider the reduction of $SU(3)$ to $O(3)$ as in Eq. (2) and study the complete set of commuting operators

$$C^{(2)}, C^{(3)}, L^2, L_3, \text{ and } X, \tag{3}$$

where X is the additional "degeneracy lifting" operator, supplying the label missing in the reduction (2). In order to commute with L^2 and L_3, the operator X must be an $O(3)$ scalar. We shall search for X in the enveloping algebra of $SU(3)$—hence it will automatically commute with the $SU(3)$ Casimir operators $C^{(2)}$ and $C^{(3)}$.

Our main result is that we have shown that only a very small number of independent $O(3)$ scalars X exists in the enveloping algebra of $SU(3)$. Indeed only one third order $X^{(3)}$ and one fourth order $X^{(4)}$ independent operator of this type can be found. All other $O(3)$ scalars can then be written as polynomials in $C^{(2)}$, $C^{(3)}$, L^2, $X^{(3)}$, and $X^{(4)}$ (this result was probably well known, e.g., to Racah, but we are not aware of any general proof).

In Sec. 2 we show for an arbitrary connected Lie group G and an arbitrary (compact or semisimple) Lie subgroup $H\subset G$ that the number of independent scalars with respect to H in the enveloping algebra of G is finite. We then identify G with $SU(3)$, H with $O(3)$, and derive a generating function for the number of $O(3)$ scalars of each order. Finally we present the independent $O(3)$ scalars explicitly. At this stage it is appropriate to stress that the method presented for deriving the generating function for the number of subgroup scalars of a definite order in the enveloping algebra of a given group is quite general and can be applied to many cases of physical interest.

In Sec. 3 we discuss the operator $X^{(3)}$ in detail, derive formulas for its eigenvalues for the cases when the $O(3)$ representation J occurs at most twice in the representation (k_1, k_2). We present a numeric method, making use of the Gel'fand–Tseitlin states, for calculating the $X^{(3)}$ and $X^{(4)}$ eigenvalues for arbitrary representations. The method, which turns out to be quite simple, is then applied to calculate the eigenvalues on a computer for a large number of representations. The results are presented in Tables I and II. A different method for calculating the eigenvalues of $X^{(3)}$ was quite recently presented by Hughes.[13] For those four representations that he considered our results coincide (up to a normalization factor equal to $2\sqrt{6}$). His operator Q_t^0 differs from $X^{(4)}$ by an algebraic combination of the lower order $O(3)$ scalar operators so that the eigenvalues cannot be easily compared. Still another method for calculating these eigenvalues was essentially contained in the by now classical articles of Bargmann and Moshinsky.[6]

2. SUBGROUP INVARIANTS IN THE ENVELOPING ALGEBRA OF THE GROUP

A. Proof that the algebra of invariants is finitely generated

Let H be a connected Lie group, compact or semisimple, with Lie algebra $\mathcal{H}$ and let the matrices $T(h)$, $h\in H$, be an $n\times n$ matrix representation of H. The mapping $\mathbf{x}\to T(h)\mathbf{x}$, where $\mathbf{x}=(x_1,\ldots,x_n)$ is a column vector, induces a representation of H in the space $P[\mathbf{x}]$ of all polynomials in the indeterminants $x_1,\ldots,x_n$ over the complex field. Clearly, the subspaces $P_m[\mathbf{x}]$ consisting of homogeneous polynomials of degree m in the x_j are invariant under the group action, $m=0,1,2,\cdots$.

An *invariant* in $P[\mathbf{x}]$ is a polynomial $p(\mathbf{x})$ which is fixed under the group action: $p(T(g)\mathbf{x})=p(\mathbf{x})$ for all $h\in H$. Clearly, the invariants in $P[\mathbf{x}]$ form an associative algebra $I[\mathbf{x}]$. In particular $a_1p_1(\mathbf{x})+a_2p_2(\mathbf{x})\in I[\mathbf{x}]$ and $p_1(\mathbf{x})p_2(\mathbf{x})\in I[\mathbf{x}]$ for any invariants $p_1,p_2\in I[\mathbf{x}]$ and constants $a_1,a_2\in C$. Furthermore, $I[x]=\sum_{m=0}^{\infty} I_m[\mathbf{x}]$, where $I_m[\mathbf{x}]=I[\mathbf{x}]\cap P_m[\mathbf{x}]$.

A fundamental fact about $I[\mathbf{x}]$ is that it is finitely generated. That is, there exists a finite set $i_1,\ldots,i_q$ of nonconstant invariants such that for every $p(\mathbf{x})\in I[\mathbf{x}]$ it is possible to find a polynomial $h(y_1,\ldots,y_q)$ with the property $p(\mathbf{x})\equiv h(i_1(\mathbf{x}),\ldots,i_q(\mathbf{x}))$. Clearly one can choose $i_1,\ldots,i_q$ as homogeneous polynomials in the x_j. Furthermore, if one of the generators, say i_q, can be expressed as a polynomial in the remaining generators, then we can remove it and $i_1,\ldots,i_{q-1}$ will still generate $I[\mathbf{x}]$.

Proceeding in this way, we eventually obtain a minimal set of nonconstant homogeneous polynomial invariants $i_1',\ldots,i_{q'}'$ which generate $I[\mathbf{x}]$. Such a minimal generating set for $I[\mathbf{x}]$ is called an *integrity basis*. A proof of the existence of a finite integrity basis can be obtained by a slight modification of that given by Weyl,[14] and will not be repeated here.

Let G be a connected Lie group containing H as a Lie subgroup. Then $\mathcal{H}$ is a subalgebra of the Lie algebra $\mathcal{G}$ of G. Let $\mathcal{U}$ be the universal enveloping algebra[1] of $\mathcal{G}$. If $X_1,\ldots,X_n$ is a basis for $\mathcal{G}$, it follows from the Poincaré–Birkhoff–Witt (PBW) theorem[1] that as a vector space $\mathcal{U}\approx\sum_{m=0}^{\infty}\oplus\,\mathcal{U}_m$, where $\mathcal{U}_0=C$, $\mathcal{U}_1=\mathcal{G}$, and $\mathcal{U}_m$ is the space of all symmetric polynomials $p(X_1,\ldots,X_n)$ in the Lie algebra generators which are homogeneous of degree m (see Ref. 1). Furthermore, H (and $\mathcal{H}$) act on $\mathcal{U}$ by means of the adjoint representation, and the subspaces $\mathcal{U}_m$ are invariant under this action. In this paper we are interested in computing the elements in $\mathcal{U}$ which are fixed under the adjoint action of H. If we denote the set of all such elements by $\mathcal{J}$, we see easily that $\mathcal{J}$ is an associative algebra and $\mathcal{J}\approx\sum_{m=0}^{\infty}\oplus\,\mathcal{J}_m$, where $\mathcal{J}_m\subseteq\mathcal{U}_m$.

Note that as a vector space $\mathcal{U}$ is isomorphic to $P[\mathbf{x}]$. Indeed, by the PBW theorem every $p\in\mathcal{U}$ can be written uniquely as $p=\sum_{m=0}^{\infty}p_m(X_1,\ldots,X_n)$, $p_m\in\mathcal{U}_m$. Moreover, the assignment $p_m(X_1,\ldots,X_n)\longleftrightarrow p_m(x_1,\ldots,x_n)$ yields an isomorphism of $\mathcal{U}_m$ and $P_m[\mathbf{x}]$. Finally, if we define the $n\times n$ matrix representation T of H to be that induced by the adjoint action of H on the basis $X_1,\ldots,X_n$ of $\mathcal{G}$, we see that there is a one-to-one correspondence between invariants in $\mathcal{U}$ and polynomial invariants in $P[\mathbf{x}]$.

We can define the notion of an integrity basis for the invariants $\mathcal{J}$ in $\mathcal{U}$ in exact analogy with the definition for the invariants $I[\mathbf{x}]$ in $P[\mathbf{x}]$. An *integrity basis* for $\mathcal{J}$ is a finite set $\{i_1,\ldots,i_q\}$ such that: (1) Each $i_j\in\mathcal{J}$ is homogeneous of degree $m_j\geq 1$ and symmetric in

$X_1, \ldots, X_n$, i.e., each $i_j \in \mathcal{J}_{m_j}$. (2) Every $i \in \mathcal{J}$ can be expressed as a polynomial in $i_1, \ldots, i_q$. (Here we must take into account the fact that the X_k hence the i_j may not commute.) (3) No one of the i_k may be expressed as a polynomial in the remaining i_j, $j \neq k$.

Due to the noncommutativity of the X_j it is not immediately obvious that $\mathcal{J}$ has a finite integrity basis. However, the following holds.

Theorem: If $i_1(x), \ldots, i_q(\mathbf{x})$ is an integrity basis of homogeneous polynomials for $I(\mathbf{x})$, then $i_1(X_1, \ldots, X_n), \ldots, i_q(X_1, \ldots, X_n)$ contains an integrity basis for $\mathcal{J}$. Here, $i_j(X_1, \ldots, X_n)$ is the homogeneous symmetric polynomial in $\mathcal{U}$ corresponding to $i_j(X_1, \ldots, X_n)$.

Proof: We will show that any $C \in \mathcal{J}$ can be expressed as a polynomial in $i_1, \ldots, i_q$. Without loss of generality we can assume $C = C_m \in \mathcal{J}_m$. The proof now proceeds by induction on m. The case $m = 0$ is obvious. Suppose C_m can be expressed as a polynomial in $i_1, \ldots, i_q$ for any $m < m_0$ and consider some $C_{m_0} \in \mathcal{J}_{m_0}$. Since $\{i_j(\mathbf{x})\}$ is an integrity basis for $I[\mathbf{x}]$, if follows that the polynomial $C_{m_0}(\mathbf{x}) \in \mathcal{J}_{m_0}[\mathbf{x}]$ can be expressed as a polynomial in the $i_j(\mathbf{x})$.

Suppose for example that $C_{m_0}(\mathbf{x}) = i_1(\mathbf{x}) i_2(\mathbf{x})$ where $i_1 \in I_{m_1}[x]$, $i_2 \in I_{m_2}[x]$, and $m_0 = m_1 + m_2$. Now consider the elements $C_{m_0}(X_j)$ and $i_1(X_j) i_2(X_j)$ in $\mathcal{U}$. We have $C_{m_0}(X_j) \in \mathcal{J}_{m_0}$ while in general

$$i_1(X_j) i_2(X_j) \subseteq \sum_{m=0}^{m_0} \oplus \mathcal{J}_m.$$

However, it is easy to see that the component of $i_1 i_2$ in $\mathcal{J}_{m_0}$ is just $C_{m_0}(X_j)$. Thus,

$$C_{m_0}(X_j) - i_1(X_j) i_2(X_j) = \sum_{m=0}^{m_0-1} C_m(X_j).$$

Since each $C_m(X_j)$ for $m < m_0$ can be expressed as a polynomial in the invariants $i_1, \ldots, i_q$, the induction step is complete. Our example easily extends to the general case. QED

In general $i_1(X_j), \ldots, i_q(X_j)$ is not an integrity basis for $\mathcal{J}$ but rather a subset $i_1', \ldots, i_q'$ is an integrity basis. This is because there may exist algebraic relations between $i_1(X_j), \ldots, i_q(X_j)$ in $\mathcal{J}$ which have no counterpart in $I(\mathbf{x})$. Such relations are consequences of the commutation relations of $\mathcal{G}$. Indeed, if $i_1(X_j)$ and $i_2(X_j)$ do not commute, then $i(X_j) = [i_1(X_j), i_2(X_j)]$ is also an invariant and the relation $i = i_1 i_2 - i_2 i_1$ is not obtainable from $I(\mathbf{x})$.

In conclusion: To find an integrity basis for $\mathcal{J}$ we first find an integrity basis $i_1, \ldots, i_q$ for $I(\mathbf{x})$. Then, forming all possible commutators $[i_s(X_j), i_p(X_j)]$, we determine a minimal subset of the i_k which are independent.

B. Generating function for the number of $O(3)$ invariants of arbitrary finite order in the enveloping algebra of $SU(3)$

In this paper we are concerned with the example $G = SU(3)$, $H = O(3)$.

Under the adjoint representation of $O(3)$ the eight-dimensional Lie algebra $SU(3)$ splits into a direct sum of the irreducible three- and five-dimensional representations of $O(3)$. The elements L_j, T_{ij}, $1 \leq i, j \leq 3$ form a basis for $SU(3)$ where the vector L_j transforms according to the three-dimensional representation D_1 and the symmetric traceless tensor T_{ij} transforms according to the five-dimensional representation D_2 of $O(3)$.

In more physical terms we can identify $L = \{L_i\}$ and $T = \{T_{ik}\}$ with the angular momentum and quadrupole moment operators, putting

$$L_j = \epsilon_{jlk} x_l p_k, \qquad T_{jk} = \tfrac{1}{2}(p_j p_k + x_j x_k) - \tfrac{1}{6}(\vec{p}^2 + \vec{x}^2)\delta_{jk}, \tag{4}$$

where x_j are the coordinates of a particle and $p_j = -i\,\partial/\partial x_j$ its momentum. These operators satisfy the $SU(3)$ commutation relations

$$\begin{aligned} &[L_j, L_k] = i\epsilon_{jkl} L_l, \\ &[L_j, T_{kl}] = i\epsilon_{jkm} T_{lm} + i\epsilon_{jlm} T_{km}, \\ &[T_{jk}, T_{lm}] = \tfrac{1}{4} i(\delta_{jl}\epsilon_{kmn} + \delta_{jm}\epsilon_{kln} + \delta_{kl}\epsilon_{jmn} + \delta_{km}\epsilon_{jln}) L_n. \end{aligned} \tag{5}$$

In the defining representation of $SU(3)$ these generators can be identified as follows:

$$\begin{aligned} &L_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad L_2 = \frac{i}{\sqrt{2}}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \\ &L_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \\ &T_{11} = \frac{1}{6}\begin{pmatrix} -1 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & -1 \end{pmatrix}, \quad T_{22} = \frac{1}{6}\begin{pmatrix} -1 & 0 & -3 \\ 0 & 2 & 0 \\ -3 & 0 & -1 \end{pmatrix}, \\ &T_{33} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ &T_{12} = \frac{i}{2}\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad T_{23} = \frac{i}{2\sqrt{2}}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \\ &T_{31} = \frac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}. \end{aligned} \tag{6}$$

By our theorem, to find an $O(3)$ integrity basis for the enveloping algebra of $SU(3)$ it is enough to find an integrity basis for the space of all polynomials in the eight indeterminants l_i, t_{jk}, $1 \leq i, j, k \leq 3$, where $t_{jk} = t_{kj}$ and $t_{jj} = 0$. Here the l_i transform under $O(3)$ according to D_1 and the t_{jk} according to D_2. In this case it is clear that the subspace P_{nm} of polynomials homogeneous of degree n in the l_i and degree m in the t_{jk} is invariant under the group action. Thus we can classify polynomial invariants $C^{(n,m)}$ in terms of their degrees of homogeneity n, m.

It is very easy to construct examples of polynomial invariants, e.g.,

$$\begin{aligned} &C^{(2,0)} = l_i l_i, \quad C^{(2,1)} = l_i t_{ij} l_j, \\ &C^{(0,2)} = t_{ij} t_{ij}, \quad C^{(0,3)} = t_{ij} t_{jk} t_{ki}, \\ &C^{(2,2)} = l_i t_{ij} t_{jk} l_k, \quad C^{(3,3)} = \epsilon_{abc} t_{bk} t_{cj} t_{jh} l_a l_k l_h. \end{aligned} \tag{7}$$

The basic problem is to find all such independent invariants, or more specifically, to construct an integrity basis. We will show below that the above list of six invariants is in fact an integrity basis and thus solve our problem.

First of all it is useful to apply Lie's theory of invariants to this problem, e.g., Ref. 15. It follows easily from this theory that the action of the three-dimensional algebra $so(3)$ on eight-parameter functions $F(l_i, t_{jk})$ implies the existence of exactly five functionally independent invariants $h_a(l_i, t_{jk})$, $a=1,\ldots,5$. By this we mean that there exist five invariant functions analytic (but not necessarily polynomials) in the variables l_i, t_{jk} such that every other invariant is an analytic function of these five. Furthermore, no one of the h_a can be expressed as an analytic function of the remaining four.

By inspection one can show that the invariants $C^{(2,0)}$, $C^{(2,1)}, C^{(0,2)}, C^{(0,3)}, C^{(2,2)}$ are functionally independent, so that all other invariants must be analytic functions of these five. However, the remaining invariants would have to be expressible as *polynomials* in these five invariants for them to be an integrity basis. $C^{(3,3)}$ is not so expressible. Indeed a direct computation yields

$$\begin{aligned}
[C^{(3,3)}]^2 &= C^{(2,0)}C^{(2,1)}C^{(0,3)}C^{(2,2)} + \tfrac{1}{2}C^{(2,0)}C^{(0,2)}[C^{(2,2)}]^2 \\
&\quad - \tfrac{1}{4}C^{(2,0)}[C^{(0,2)}]^2[C^{(2,1)}]^2 - \tfrac{1}{9}[C^{(2,0)}]^3[C^{(0,3)}]^2 \\
&\quad - \tfrac{1}{3}[C^{(2,0)}]^2C^{(0,2)}C^{(2,1)}C^{(0,3)} + \tfrac{1}{2}C^{(0,2)}[C^{(2,1)}]^2C^{(2,2)} \\
&\quad - \tfrac{1}{3}[C^{(2,1)}]^3C^{(0,3)} - [C^{(2,2)}]^3,
\end{aligned} \tag{8}$$

i.e., $[C^{(3,3)}]^2$ is a polynomial in the first five invariants but $C^{(3,3)}$ is not.

To show explicitly that we have found an integrity basis we generalize a technique found in Ref. 14, p. 181, and Ref. 16, to derive a generating function for the number of invariants of rank (n,m). For this we recall that the irreducible representations of $O(3)$ can be denoted by D_j, $j=0,1,2,\cdots$, and that the character $\chi_j(\theta)$ of D_j corresponding to a rotation through the angle θ is

$$\chi_j(\theta) = \sum_{k=-j}^{j} \exp(ik\theta). \tag{9}$$

By choosing a weight basis it is straightforward to check that the character $\chi_{n,m}(\theta)$ of $O(3)$ acting on the subspace $P_{n,m}$ is

$$\chi_{n,m}(\theta) = \sum_{a,\ldots,h} \exp[i\theta(a-c+2d+e-g-2h)], \tag{10}$$

where the sum is taken over all nonnegative integers $a,\ldots,h$ such that $a+b+c=n$, $d+e+f+g+h=m$. It follows from this that

$$\begin{aligned}
&F[\exp(i\theta), P, D] \\
&\quad \equiv [(1-\exp(i\theta)P)(1-P)(1-\exp(-i\theta)P)(1-\exp(2i\theta)D) \\
&\qquad \times (1-\exp(i\theta)D)(1-D)(1-\exp(-i\theta)D)(1-\exp(-2i\theta)D)]^{-1} \\
&\quad = \sum_{n,m=0}^{\infty} \chi_{n,m}(\theta)P^nD^m,
\end{aligned} \tag{11}$$

i.e., $F[\exp(i\theta), P, D]$ is a generating function for the character $\chi_{n,m}(\theta)$. Note that the number of invariants of degree (n,m) is just the multiplicity of the identity representation D_0 of $O(3)$ in $P_{n,m}$. Thus, using the orthogonality relations

$$\frac{1}{\pi}\int_0^{2\pi} \chi_n(\theta)\overline{\chi_m(\theta)} \sin^2\frac{\theta}{2}\, d\theta = \delta_{nm}, \tag{12}$$

we find

$$\frac{1}{\pi}\int_0^{2\pi} \sin^2\frac{\theta}{2}\, F(\exp(i\theta), P, D)\, d\theta = \sum_{n,m=0}^{\infty} N_{n,m}P^nD^m, \tag{13}$$

where the integer $N_{n,m}$ is the number of linearly independent $O(3)$ invariants of rank (n,m). Setting $\exp(i\theta) = \lambda$, we can regard the left-hand side as a contour integral about a unit circle in the complex λ plane. Evaluating the integral by residues and employing some tedious algebra, we finally obtain

$$\frac{1+P^3D^3}{(1-P^2)(1-D^2)(1-D^3)(1-P^2D^2)(1-P^2D)} = \sum_{n,m=0}^{\infty} N_{n,m}P^nD^m. \tag{14}$$

It is illuminating to compare this expression with our earlier results. Since $C^{(2,0)}$, $C^{(2,1)}$, $C^{(0,2)}$, $C^{(0,3)}$, and $C^{(2,2)}$ are functionally independent, we can construct invariants of the form $[C^{(2,0)}]^a[C^{(2,1)}]^b[C^{(0,2)}]^c[C^{(0,3)}]^d \times [C^{(2,2)}]^e$, where $a,\ldots,e$ run over the nonnegative integers and the set of all such invariants is linearly independent. If these were all possible invariants, then the generating function (14) would be

$$\frac{1}{(1-P^2)(1-D^2)(1-D^3)(1-P^2D)(1-P^2D^2)}. \tag{15}$$

However, the actual $N_{n,m}$ is in general larger than that predicted by (15) which shows that there are additional invariants. Indeed $N_{3,3}=1$, while it is impossible to construct a $(3,3)$ invariant out of $C^{(2,0)},\ldots,C^{(2,2)}$. Thus, there must exist a new $(3,3)$ invariant. This new invariant is clearly $C^{(3,3)}$. We can now obtain new invariants of the form $C^{(3,3)}[C^{(2,0)}]^a,\ldots,[C^{(2,2)}]^e$. This accounts for all terms in (14) and completely solves the problem of finding all $O(3)$ invariants. (It is not possible to obtain independent invariants by taking higher power of $C^{(3,3)}$ because $[C^{(3,3)}]^2$ can be expressed as a polynomial in $C^{(2,0)},\ldots,C^{(2,2)}$.)

C. The $O(3)$ invariants and the $SU(3)\supset O(3)$ reduction

It was shown above that there are at most six algebraicly independent $O(3)$ scalars in the enveloping algebra of $SU(3)$. They can easily be expressed in terms of the generators L_i and T_{ik} of Eqs. (4)–(6) and indeed they are given by Eq. (7) with l_i and t_{ij} replaced by the operators L_i and T_{ij}.

The two Casimir operators[17] $C^{(2)}$ and $C^{(3)}$ of $SU(3)$ are, of course, also $O(3)$ scalars and must be contained among those found. Indeed, it is easy to check that we have

$$\begin{aligned}
C^{(2)} &= (\tfrac{3}{4})^2(L^2+2T^2) = (\tfrac{3}{4})^2(L_iL_i + 2T_{ik}T_{ik}), \\
\text{const}\; C^{(3)} &= LTL - \tfrac{4}{3}TTT = L_iT_{ik}L_k - \tfrac{4}{3}T_{ik}T_{kl}T_{li}.
\end{aligned} \tag{16}$$

It is also easy to verify that the operator

$$X^{(6)} = \epsilon_{abc}T_{bd}T_{ce}T_{ef}L_aL_dL_f$$

can be expressed in terms of the commutator of the two operators

$$X^{(3)} = L_a T_{ab} L_b \quad \text{and} \quad X^{(4)} = L_a T_{ab} T_{bc} L_c \tag{17}$$

and lower order terms.

In addition to the angular momentum L^2 and the two Casimir operators $C^{(2)}$ and $C^{(3)}$ we thus only have two new independent $O(3)$ invariants $X^{(3)}$ and $X^{(4)}$ [see (17)].

Note that the scalars of this section do not quite coincide with those listed in Eq. (7) because they are not all symmetrized. However, they do agree in the highest order terms and they provide an alternative integrity basis which is computationally easier to deal with.

Let us note here that the operator $X^{(3)}$ is equivalent to an operator used in a similar context by Bargmann and Moshinsky.[6]

Returning to the problem of representations in the $SU(3) \supset O(2)$ basis, we see that the basis functions of irreducible representations of $SU(3)$ can be chosen to be eigenstates of the operators $C^{(2)}$, $C^{(3)}$, L^2, L_3, and X, where X is in principle an arbitrary function of the operators (17).

If we make the natural restriction that X be an operator of a definite order in the enveloping algebra of $SU(3)$, we find that only one third-order and one fourth-order are available. Some physical implications of this fact will be discussed in the final section.

In conclusion, the operators L^2, $C^{(2)}$, $C^{(3)}$, $X^{(3)}$, and $X^{(4)}$ form an integrity basis for the $O(3)$ scalars in the enveloping algebra of $SU(3)$.

3. SPECTRUM OF THE $O(3)$—SCALAR OPERATORS

The purpose of this section is to calculate the spectrum of the third and fourth order operators $X^{(3)}$ and $X^{(4)}$ and to demonstrate some of their general properties. Indeed, for any practical use of the present state labeling method it is essential to know the spectrum of the operators for all $SU(3)$ representations likely to appear in applications.

The $SU(3) \supset O(3)$ case is only the simplest of many group—subgroup pairs of physical interest where some labels are missing. Higher order operators can resolve these labeling problems not only in principle, but in our opinion are the most practical way to approach the problem. It is therefore natural to perform the (computer) calculations of the spectra in a way which is not limited to the $SU(3) \supset O(3)$ case but can readily be extended to cases like $SU(4) \supset SU(2) \times SU(2)$, $G_2 \supset O(3)$, and others. The basis we use for deriving the secular equations is that of Gel'fand and Tseitlin,[2] with $U(3)$ generators E_{ik} satisfying the commutation relations

$$[E_{ij}, E_{kl}] = \delta_{jk} E_{il} - \delta_{il} E_{kj}, \tag{18}$$

where δ_{jk} is the Kronecker delta. An explicit form of the matrix elements of the $U(3)$ generators can be found in the second example of Ref. 18 [Eq. (22)]; correspondence between the notations in the present paper and Ref. 18 is established by putting $E_{ik} \equiv C_i^k$ and $m_{ik} \equiv h_{ik}$, where m_{ik} are the elements of each pattern-basis vector.

It is convenient to replace the generators L_1, L_2, and L_3 of (6) by equivalent ones:

$$L_1 = E_{12} + E_{23}, \quad L_0 = E_{11} - E_{33}, \quad L_{-1} = E_{21} + E_{32}, \tag{19}$$

whose commutations relations

$$[L_1, L_{-1}] = L_0, \quad [L_0, L_1] = L_1, \quad [L_0, L_{-1}] = -L_{-1} \tag{20}$$

follow from (18). The generators (19) can be realized as 3×3 matrices:

$$L_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad L_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad L_{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{21}$$

With the choice (19) the five components of the operator T_{ik} then can be taken as

$$\begin{aligned} &T_2 = E_{13}, \quad T_1 = E_{12} - E_{23}, \quad T_0 = E_{11} - 2E_{22} + E_{33}, \\ &T_{-2} = E_{31}, \quad T_{-1} = E_{21} - E_{32}. \end{aligned} \tag{22}$$

Realized as 3×3 matrices, these are

$$\begin{aligned} &T_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ &T_{-2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad T_{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}. \end{aligned} \tag{23}$$

Using (18), one readily verifies that T_i indeed is the rank two $O(3)$-tensor operator:

$$\begin{aligned} &[T_2, L_1] = 0, \quad [T_2, L_0] = -2T_2, \quad [T_2, L_{-1}] = T_1, \\ &[T_1, L_1] = 2T_2, \quad [T_1, L_0] = -T_1, \quad [T_1, L_{-1}] = T_0, \\ &[T_0, L_1] = 3T_1, \quad [T_0, L_0] = 0, \quad [T_0, L_{-1}] = -3T_{-1}, \text{ etc.} \end{aligned} \tag{24}$$

The second order operators $C^{(2)}$, L^2, and T^2 are then

$$\begin{aligned} &L^2 = L_1 L_{-1} + L_{-1} L_1 + L_0^2, \\ &T^2 = T_2 T_{-2} + T_{-2} T_2 + \tfrac{1}{2}(T_1 T_{-1} + T_{-1} T_1) + \tfrac{1}{6} T_0^2, \\ &C^{(2)} = \sum_{i,k=1}^{3} E_{ik} E_{ki} = (\tfrac{3}{4})^2 (L^2 + 2T^2). \end{aligned} \tag{25}$$

The labeling operators then are

$$\begin{aligned} X^{(3)} = {} & 3(L_1 T_{-2} L_1 + L_{-1} T_2 L_{-1}) \\ & + \tfrac{3}{2}(L_{-1} T_1 L_0 + L_0 T_1 L_{-1} + L_1 T_{-1} L_0 + L_0 T_{-1} L_1) \\ & - \tfrac{1}{2}(L_1 T_0 L_{-1} + L_{-1} T_0 L_1) + L_0 T_0 L_0 \end{aligned} \tag{26}$$

and

$$\begin{aligned} X^{(4)} = {} & 2T_0 L_0 L_0 T_0 + (-T_0 L_1 L_{-1} T_0 + \tfrac{3}{2} T_0 L_1 L_0 T_{-1} + \tfrac{3}{2} T_0 L_0 L_1 T_{-1} \\ & - 6T_0 L_1 L_1 T_{-1} + 9T_1 L_{-1} L_{-1} T_1 + \tfrac{3}{2} T_1 L_{-1} L_0 T_0 \\ & + \tfrac{3}{2} T_1 L_0 L_{-1} T_0 - \tfrac{3}{2} T_1 L_{-1} L_1 T_{-1} + 3T_1 L_0 L_0 T_{-1} \\ & - \tfrac{3}{2} T_1 L_1 L_{-1} T_{-1} + 9T_1 L_1 L_0 T_{-2} + 9T_1 L_0 L_1 T_{-2} \\ & - 6T_2 L_{-1} L_{-1} T_0 + 9T_2 L_{-1} L_0 T_{-1} + 9T_2 L_0 L_{-1} T_{-1} \\ & + 6T_2 L_{-1} L_1 T_{-2} - 12 T_2 L_0 L_0 T_{-2} + 6T_2 L_1 L_{-1} T_{-2}) \\ & + (\cdots), \end{aligned} \tag{26'}$$

where $(\cdots)$ stands for terms with signs of indices opposite to those in the first bracket. Here $X^{(3)}$ and $X^{(4)}$ are normalized so that their eigenvalues are integers whenever possible. The operators $X^{(4)}$ in (17) and (26′) differ by $O(3)$-scalars of order lower than four. By a straightforward calculation one verifies that $X^{(i)}$, indeed, are $O(3)$ scalars:

$$[X^{(i)}, L_1] = [X^{(i)}, L_0] = [X^{(i)}, L_{-1}] = 0, \quad i = 3 \text{ or } 4. \tag{27}$$

An irreducible representation of $U(3)$ is denoted by integers (m_{13}, m_{23}, m_{33}) such that $m_{13} \geq m_{23} \geq m_{33}$. If $m_{33} = 0$, a $U(3)$ representation reduces to that of $SU(3)$ with $p = m_{13} - m_{23}$ and $q = m_{23}$. The patterns

$$\left|\begin{matrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \\ & & m_{11} & & \end{matrix}\right\rangle,$$

$$m_{i,k+1} \geq m_{ik} \geq m_{i+1,k}, \quad m_{ik} \text{ integers}, \tag{28}$$

transformed by the generators E_{ik} according to (22) of Ref. 18, form an orthonormal basis in a space in which an irreducible unitary representation of the group $U(3)$ acts. If $m_{33} = 0$, the space is irreducible with respect to $SU(3)$. Since m_{13}, m_{23}, and m_{33} are fixed throughout an irreducible representation of $U(3)$, we shall omit them when writing the patterns.

The $C^{(2)}$ and $C^{(3)}$ operators are[17] diagonal in the basis (28) because they are the Casimir operators of $U(3)$ [and $SU(3)$]. Since E_{11}, E_{22}, and E_{33} are diagonal in (28) too, L_0 and T_0 are also diagonal. One has, in particular

$$L_0 \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle = (m_{11} + m_{12} + m_{22} - m_{13} - m_{23} - m_{33}) \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle \tag{29a}$$

and

$$T_0 \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle = (m_{13} + m_{23} + m_{33} + 3m_{11} - 3m_{12} - 3m_{22}) \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle. \tag{29b}$$

An arbitrary $SU(3)$ pattern for a given representation is a linear combination of $O(3)$ states $|JM\rangle$:

$$\left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle = \sum_{JK} a_J |JMK\rangle. \tag{30}$$

Here a_J are some coefficients,

$$M = m_{11} + m_{12} + m_{22} - m_{13} - m_{23} - m_{33} \tag{31}$$

is the eigenvalue of L_0, J denotes an $O(3)$-irreducible subspace, and K are the eigenvalues of $X^{(i)}$ which we want to find. The values of J for any $U(3)$ representation are well known.[19,20] The summation in (30) extends over all $J \geq M$ which occur in the $SU(3)$ space labeled by m_{13} and m_{23} ($m_{33} = 0$). There is no summation over M in (30) because both the Gel'fand–Tseitlin and $|JMK\rangle$ states are eigenvectors of the $O(2)$ generator L_0. When $X^{(i)}$ acts on both sides of (30), one gets

$$\sum_{m_{11}+m_{12}+m_{22}=M+m_{13}+m_{23}} x_m(m_{12}, m_{22}, m_{11}) \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle = \sum_{J,K} a_J K |JMK\rangle. \tag{32}$$

The coefficients x_M are matrix elements of $X^{(i)}$ between the patterns with the same value (31) of M. They are calculated using (20), (22), (26), and (22) of Ref. 18. For example, the diagonal matrix element of $X^{(3)}$ is

$$\begin{aligned}\left\langle\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right| X^{(3)} \left|\begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}\right\rangle = {} & \frac{3}{m_{12} - m_{22} + 1} \left(\frac{(m_{13} - m_{12})(m_{12} - m_{23} + 1)(m_{12} - m_{33} + 2)(m_{12} - m_{11} + 1)[2(m_{11} - m_{22}) - 2M - N/3]}{(m_{12} - m_{22} + 2)} \right. \\ & \left. + \frac{(m_{13} - m_{22} + 1)(m_{23} - m_{22})(m_{22} - m_{33} + 1)(m_{11} - m_{22})[2(m_{11} - m_{12} - 1) - 2M - N/3]}{(m_{12} - m_{22})} \right) \\ & + 3(m_{12} - m_{11})(m_{11} - m_{22} + 1) \left(2M - \frac{N}{3} + 2 \frac{(m_{13} - m_{12} + 1)(m_{12} - m_{23})(m_{12} - m_{33} + 1)}{(m_{12} - m_{22} + 1)(m_{12} - m_{22})} \right. \\ & \left. - 2 \frac{(m_{13} - m_{22} + 2)(m_{23} - m_{22} + 1)(m_{22} - m_{33})}{(m_{12} - m_{22} + 1)(m_{12} - m_{22} + 2)} \right) + \tfrac{1}{2} N(M+1)(2M+3),\end{aligned} \tag{33}$$

where M is given by (31) and N is the eigenvalue of T_0:

$$N = 3(m_{11} - m_{12} - m_{22}) + m_{13} + m_{23} + m_{33}. \tag{34}$$

Substituting (30) into the left side of (32), and comparing the coefficients of the linearly independent vectors $|JMK\rangle$, we arrive at the secular equation

$$|x_M(m_{11}, m_{12}, m_{22}) - K| = 0. \tag{35}$$

The roots $K_1, K_2, \cdots$ of (35) are real because $X^{(i)}$ is Hermitian. The value of M in (30) is a fixed parameter. Hence we have secular equation (35) for every value of M which occurs in the $U(3)$ representation (m_{13}, m_{23}, m_{33}). [For $SU(3)$ we still have $m_{33} = 0$.] Equation (35) is of the first order when M equals its highest (smallest) value within the inequalities (28), i.e., $M = m_{13} - m_{33}$ ($M = m_{33} - m_{13}$). Then indeed, there is only one pattern, namely $m_{11} = m_{12} = m_{13}$, $m_{22} = m_{23}$ ($m_{11} = m_{22} = m_{33}$, $m_{12} = m_{23}$). Consequently, (30) has the form

$$\left|\begin{matrix} m_{13} & & m_{23} \\ & m_{13} & \end{matrix}\right. = |m_{13} - m_{33}, m_{13} - m_{33}, K\rangle. \tag{36}$$

The order of Eq. (35) increases, in general, when the absolute value $|M|$ diminishes, and for $M = 0$, (35) is of the highest order. The order of (35) in this case equals to the number of different patterns (28) with $M = 0$, or, what is the same, it equals the number of $O(3)$ representations contained in (m_{13}, m_{23}, m_{33}).

From the property

$$X^{(i)} |JMK\rangle = K |JMK\rangle \quad \text{for } M = J, J-1, \ldots, -J,$$

of $X^{(i)}$ it follows that the eigenvalue will occur as a root of the secular equation (35) for any M. Similarly, an eigenvalue, say K', calculated from (35) with $M = M'$, will be a root of every secular equation with $|M| \leq |M'|$. One has thus two alternative ways for computating the spectrum of $X^{(i)}$ for a given representation (m_{13}, m_{23}, m_{33}). First is solution of (high order) equation (35) for $M = 0$ in order to get all the eigenvalues K at once. The second way is the solving of equation (35) first for $M = m_{13} - m_{33}$, then for $M = m_{13} - m_{33} - 1$, $M = m_{13}$

$-m_{33}-2$, and so on. In this manner the order of the secular equation for a given M is drastically reduced because most of its roots are known from solving the secular equation for $M+1$. To illustrate this point, let us notice that, e.g., for the $SU(3)$ representation $(12,6,0)$ of dimension 343, the order of (35) at $M=0$ equals 25. Proceeding the second way, one would have to solve 4, 4, 3, and 1 secular equations of orders 1, 2, 3, and 4, respectively.

The tables contain the eigenvalues of $X^{(3)}$ and $X^{(4)}$ calculated by a computer for the lower $SU(3)$ representations. For pairs of contragredient representations [i.e., representations $(m_{13}, m_{23}, 0)$ and $(m_{13}, m_{13}-m_{23}, 0)$] the $O(3)$ branching rules coincide and the eigenvalues of $X^{(3)}$ differ by a sign, and those of $X^{(4)}$ are the same. Therefore, the tables contain only one representation of each pair. The computer time needed for construction of the tables was negligible. Thus in order to verify the eigenvalues we have obtained, the secular equation (35) was solved for all $M \geq 0$ for both $X^{(3)}$ and $X^{(4)}$.

The numerical results presented in the Table I for the $SU(3)$ representations

$$(k_1, k_2, 0) = (m_{12}-m_{33}, m_{23}-m_{33}, 0) \tag{37}$$

[note that $k_1 \geq k_2 \geq 0$ and k_1 and k_2 are the lengths of the first and second row in Young patterns for $SU(3)$] were obtained using the above algorithm, starting from Gel'fand–Tseitlin states. For the particular case considered in this article, i.e., the $SU(3) \supset O(3) \supset O(2)$ group–subgroup chain a different method could also be used for calculating the eigenvalues K. Indeed, Bargmann and Moshinsky[6] and Elliott[5] have calculated the matrix elements of the operator $X^{(3)} = LTL$ in certain nonorthogonal bases. All we have to do is take these matrices and diagonalize them. For analytic calculations (as opposed to computer ones) this procedure is simpler.

Since in many applications it is convenient to have explicit formulas for the eigenvalues K, rather than only numeric tables, we present below expressions for K in special cases, when the $O(3)$ representation J occurs in the $SU(3)$ representation (k_1, k_2) once (hence K is uniquely determined as a solution of a linear equation) or twice (then K is the solution of a quadratic equation).

To do this, we choose to make use of the Bargmann–Moshinsky basis vectors $P_{k_1 k_2 J q}$ in which we have

$$X^{(3)} P_{k_1 k_2 J q} = -3 \sum_{q'} \beta_{q' q} P_{k_1 k_2 J q'} \tag{38}$$

[see formula (62) of the second of Refs. 6; the factor (-3) is due to a difference in the normalization of our $X^{(3)}$ and their operator Ω]. The matrix elements $\beta_{q'q}$ are given by formulas (66) and (67) of Ref. 6 and restrictions on the region of summation on (38) are given by their formula (59).

All we have to do is restrict ourselves to cases when only one or two values of the label q exist (no degeneracy or twofold degeneracy). If there is no degeneracy, then $K = -3\beta_{qq}$; if there is a degeneracy, then we obtain the eigenvalues K by diagonalizing the matrix $\beta_{q'q}$.

By inspecting the Bargmann–Moshinsky formulas, we see immediately that the representation J is contained in representation (k_1, k_2) at most once in any of the following cases:

$$J = 0, 1, k_1 - 1 \text{ or } k_1 \quad (k_1 \text{ and } k_2 \text{ arbitrary}), \tag{39}$$

or

$$k_2 = 0, 1, k_1 - 1, k_1 \quad (J \text{ arbitrary}). \tag{40}$$

The degeneracy is at most twofold if

$$J = 2, 3, k_1 - 2 \text{ or } k_1 - 3 \quad (k_1 \text{ and } k_2 \text{ arbitrary}), \tag{41}$$

or

$$k_2 = 2, 3, k_1 - 2 \text{ or } k_1 - 3 \quad (J \text{ arbitrary}). \tag{42}$$

Proceeding as described, we obtain the following expressions for the eigenvalues K in nondegenerate cases.

$J=0$: We have

$$K = 0 \tag{43}$$

for k_1 and k_2 both even [the representation $J=0$ is not contained in (k_1, k_2) otherwise].

$J=1$: We obtain

$$\begin{aligned} K &= -k_1 + 2k_2 && \text{for } k_1 \text{ even, } k_2 \text{ odd} \\ &= 2k_1 - k_2 + 3 && \text{for } k_1 \text{ odd, } k_2 \text{ even} \\ &= -(k_1 + k_2 + 3) && \text{for } k_1 \text{ odd, } k_2 \text{ odd} \end{aligned} \tag{44}$$

($J=1$ is not present for k_1 even, k_2 even).

$J=k_1$: We have

$$K = \tfrac{1}{2}(k_1+1)(2k_1+3)(k_1-2k_2). \tag{45}$$

$J=k_1-1$: We have

$$K = \tfrac{1}{2}(k_1+3)(2k_1+1)(k_1-2k_2). \tag{46}$$

$k_2=0$: We have

$$K = \tfrac{1}{2}(2k_1+3)J(J+1) \quad \text{for } k_1 - J \text{ even} \tag{47}$$

and J is not contained in $(k_1, 0)$ for $k_1 - J$ odd.

$k_2=1$: We have

$$\begin{aligned} K &= -3(k_1+1) + (k_1 - \tfrac{1}{2})J(J+1) && \text{for } k_1 - J \text{ even} \\ &= -3(k_1+1) + (k_1 + \tfrac{5}{2})J(J+1) && \text{for } k_1 - J \text{ odd.} \end{aligned} \tag{48}$$

$k_2 = k_1$ and $k_2 = k_1 - 1$: These are contragradient to $k_2 = 0$ and $k_2 = 1$; hence formulas (47) and (48) apply with reversed signs.

In the cases when at most a twofold degeneracy can occur, we obtain:

$J=2$: We have

$$\begin{aligned} K &= \pm 3[(2k_1+3)^2 - 4k_2(k_1-k_2)]^{1/2} \\ & \qquad \text{for } k_1 \text{ even, } k_2 \text{ even, } 2 \leq k_2 \leq k_1 - 2, \\ &= -3(2k_1+3) && \text{for } k_1 \text{ even, } k_2 \text{ even, } k_1 = k_2, \\ &= 3(2k_1+3) && \text{for } k_1 \text{ even, } k_2 \text{ even, } k_2 = 0, \\ &= 3(k_1 - 2k_2) && \text{for } k_1 \text{ even, } k_2 \text{ odd} \\ &= -3(2k_1 - k_2 + 3) && \text{for } k_1 \text{ odd, } k_2 \text{ even} \\ &= 3(k_1 + k_2 + 3) && \text{for } k_1 \text{ odd, } k_2 \text{ odd.} \end{aligned} \tag{49}$$

$J=3$: We have

TABLE I. Eigenvalues K of the third order operator $X^{(3)} = L\,T\,L$. The first column gives the representations of $SU(3)$, the rows give all possible values of the $O(3)$ label J and of K within the corresponding representation of $SU(3)$.

(0,0,0)	J	0											
	K	0											
(1,0,0)	J	1											
	K	5											
(2,0,0)	J	2	0										
	K	21	0										
(2,1,0)	J	2	1										
	K	0	0										
(3,0,0)	J	3	1										
	K	54	9										
(3,1,0)	J	3	2	1									
	K	18	21	−7									
(4,0,0)	J	4	2	0									
	K	110	33	0									
(4,1,0)	J	4	3	2	1								
	K	55	63	6	−2								
(4,2,0)	J	4	3	2	2	0							
	K	0	0	30.741	−30.741	0							
(5,0,0)	J	5	3	1									
	K	195	78	13									
(5,1,0)	J	5	4	3	2	1							
	K	117	132	36	27	−9							
(5,2,0)	J	5	4	3	3	2	1						
	K	39	44	79.573	−13.573	−33	11						
(6,0,0)	J	6	4	2	0								
	K	315	150	45	0								
(6 1,0)	J	6	5	4	3	2	1						
	K	210	234	89	81	12	−4						
(6,2,0)	J	6	5	4	4	3	2	2	0				
	K	105	117	160	22	0	41.677	−41.677	0				
(6,3,0)	J	6	5	4	4	3	3	2	1				
	K	0	0	58.864	−58.864	92.223	−92.223	0	0				
(7,0,0)	J	7	5	3	1								
	K	476	255	102	17								
(7,1,0)	J	7	6	5	4	3	2	1					
	K	340	375	171	166	54	33	−11					
(7,2,0)	J	7	6	5	5	4	3	3	2	1			
	K	204	225	277.903	82.097	60	100.723	−10.723	−45	15			
(7,3,0)	J	7	6	5	5	4	4	3	3	2	1		
	K	68	75	137.827	−17.827	186.231	−56.231	−108	30	39	−13		
(8,0,0)	J	8	6	4	2	0							
	K	684	399	190	57	0							
(8,1,0)	J	8	7	6	5	4	3	2	1				
	K	513	561	288	288	123	99	18	−6				
(8,2,0)	J	8	7	6	6	5	4	4	3	2	2	0	
	K	342	374	439.120	172.880	153	195.426	42.574	0	53.075	−53.075	0	
(8,3,0)	J	8	7	6	6	5	5	4	4	3	3	2	1
	K	171	187	256.966	49.034	321.259	8.741	87.827	−67.827	109.763	−121.763	6	−2
(8,4,0)	J	8	7	6	6	5	5	4	4	4	3	2	2
	K	0	0	−92.223	92.223	148.704	−148.704	214.075	−214.075	0	0	51.701	−51.701
	J	0											
	K	0											
(9,0,0)	J	9	7	5	3	1							
	K	945	588	315	126	21							
(9,1,0)	J	9	8	7	6	5	4	3	2	1			
	K	735	798	446	453	225	200	72	39	−13			
(9,2,0)	J	9	8	7	7	6	5	5	4	3	3	2	1
	K	525	570	649.542	300.458	285	331.966	124.034	76	122.795	−8.795	−57	19

TABLE I. (*continued*)

(9,3,0)	J	9	8	7	7	6	6	5	5	4	4	3	3
	K	315	342	422.117	147.883	503.385	108.615	182.223	−2.223	219.031	−69.031	−137.223	47.223
	J	2	1										
	K	45	−15										
(9,4,0)	J	9	8	7	7	6	6	5	5	5	4	4	3
	K	105	114	208.882	−18.882	284.014	−80.014	365.586	−167.567	56.981	−241.259	71.259	127.426
	J	3	2	1									
	K	−25.426	−51	17									
(10,0,0)	J	10	8	6	4	2	0						
	K	1265	828	483	230	69	0						
(10,1,0)	J	10	9	8	7	6	5	4	3	2	1		
	K	1012	1092	651	667	366	342	157	117	24	−8		
(10,2,0)	J	10	9	8	8	7	6	6	5	4	4	3	2
	K	759	819	915.101	470.899	462	239.837	516.163	189	231.906	62.094	0	64.692
	J	2	0										
	K	−64.692	0										
(10,3,0)	J	10	9	8	8	7	7	6	6	5	5	4	4
	K	506	546	639	285	738.624	249.376	320.882	93.118	373.216	16.784	119	−79
	J	3	3	2	1								
	K	−151.460	127.460	12	−4								
(10,4,0)	J	10	9	8	8	7	7	6	6	6	5	5	4
	K	253	273	374.279	87.721	468.452	25.548	568.366	149.892	−91.257	182.014	−182.014	247.375
	J	4	4	3	2	2	0						
	K	−269.112	21.737	0	−62.426	62.426	0						
(10,5,0)	J	10	9	8	8	7	7	6	6	6	5	5	5
	K	0	0	130.111	−130.111	213.169	−213.169	315.728	−315.728	0	409.805	−409.805	0
	J	4	4	3	3	2	1						
	K	92.223	−92.223	138.942	−138.942	0	0						
(11,0,0)	J	11	9	7	5	3	1						
	K	1650	1125	700	375	150	25						
(11,1,0)	J	11	10	9	8	7	6	5	4	3	2	1	
	K	1350	1449	909	936	552	531	279	234	90	45	−15	
(11,2,0)	J	11	10	9	9	8	7	7	6	5	5	4	3
	K	1050	1127	1241.755	690.245	690	753.821	396.179	345	387	165	92	145.426
	J	3	2	1									
	K	−7.426	−69	23									
(11,3,0)	J	11	10	9	9	8	8	7	7	6	6	5	5
	K	750	805	913.394	466.606	1032.968	437.032	223.721	510.279	578.506	141.494	230.023	9.977
	J	4	4	3	3	2	1						
	K	252.168	−82.168	−166.763	64.763	51	−17						
(11,4,0)	J	11	10	9	9	8	8	7	7	7	6	6	5
	K	450	483	594.225	233.775	708.135	173.865	828.332	287.538	18.131	339.169	−87.169	417.241
	J	5	5	4	4	3	3	2	1				
	K	−200.262	98.020	−296.648	86.648	147.906	−21.906	−63	21				
(11,5,0)	J	11	10	9	9	8	8	7	7	7	6	6	6
	K	150	161	292.172	−16.172	397.298	−103.298	517.824	−231.787	91.962	633.480	−345.871	111.391
	J	5	5	5	4	4	3	3	2	1			
	K	−454.366	205.257	−35.892	273.216	−83.216	−155.636	41.636	57	−19			
(12,0,0)	J	12	10	8	6	4	2	0					
	K	2106	1485	972	567	270	81	0					
(12,1,0)	J	12	11	10	9	8	7	6	5	4	3	2	1
	K	1755	1875	1226	1266	789	773	444	396	191	135	30	−10
(12,2,0)	J	12	11	10	10	9	8	8	7	6	6	5	4
	K	1404	1500	1635.477	964.523	975	599.201	1050.799	550	305.969	594.031	225	269.154
	J	4	3	2	2	0							
	K	80.846	0	76.426	−76.426	0							
(12,3,0)	J	12	11	10	10	9	9	8	8	7	7	6	6
	K	1053	1125	1251.147	698.853	1392.403	677.597	756.225	395.775	840.970	311.030	388.597	133.403
	J	5	5	4	4	3	3	2	1				
	K	425.641	24.359	151.231	−91.231	145.245	−181.245	18	−6				

TABLE I. (*continued*)

(12,4,0)	J	12	11	10	10	9	9	8	8	8	7	7	6
	K	702	750	874.359	425.641	1009.088	370.912	1151.439	477.878	164.683	549.298	48.702	642.696
	J	6	6	5	5	4	4	4	3	2	2	0	
	K	212.121	−95.817	216.187	−216.187	−324.526	281.695	42.831	0	73.546	−73.546	0	
(12,5,0)	J	12	11	10	10	9	9	8	8	8	7	7	7
	K	351	375	511.507	138.493	638.709	51.291	779.553	227.926	−110.479	918.293	270.150	−243.443
	J	6	6	6	5	5	5	4	4	3	3	2	1
	K	−377.824	371.787	48.038	−498.902	460.440	8.463	120.023	−100.023	−168.361	156.361	6	−2
(12,6,0)	J	12	11	10	10	9	9	8	8	8	7	7	7
	K	0	0	−172.049	172.049	−284.747	284.747	429.367	−429.367	−0	566.960	−566.960	0
	J	6	6	6	6	5	5	4	4	4	3	2	2
	K	−698.844	698.844	138.430	−138.430	213.169	−213.169	−301.257	301.257	0	0	72.560	−72.560
	J	0											
	K	0											

TABLE II. Eigenvalues K of the fourth order operator $X^{(4)}=TLLT$. The first column gives the representations of $SU(3)$, the rows give all possible values of the O(3) label J and of K within the corresponding representation of $SU(3)$.

(0,0,0)	J	0								
	K	0								
(1,0,0)	J	1								
	K	−35								
(2,0,0)	J	2	0							
	K	63	−840							
(2,1,0)	J	2	1							
	K	63	−315							
(3,0,0)	J	3	1							
	K	342	−1323							
(3,1,0)	J	3	2	1						
	K	222	−105	−1043						
(4,0,0)	J	4	2	0						
	K	898	−1881	−2352						
(4,1,0)	J	4	3	2	1					
	K	490	438	−1617	−1547					
(4,2,0)	J	4	3	2	2	0				
	K	354	270	−297	−1449	−2016				
(5,0,0)	J	5	3	1						
	K	1875	−2562	−3347						
(5,1 0)	J	5	4	3	2	1				
	K	963	1458	−2466	−1881	−2691				
(5,2,0)	J	5	4	3	3	2	1			
	K	507	922	−2307.729	543.729	−1185	−2699			
(6,0,0)	J	6	4	2	0					
	K	3465	−3366	−4689	−4536					
(6,1,0)	J	6	5	4	3	2	1			
	K	1785	3147	−3638	−2154	−4065	−3419			
(6,2,0)	J	6	5	4	4	3	2	2	0	
	K	777	1995	−3678.851	2114.851	−1578	−3825	−1617	−3864	
(6,3,0)	J	6	5	4	4	3	3	2	1	
	K	441	1611	−3654	1530	−954	774	−3825	−3051	
(7,0,0)	J	7	5	3	1					
	K	5908	−4245	−6522	−6107					
(7,1,0)	J	7	6	5	4	3	2	1		
	K	3148	5745	−5133	−2318	−6114	−4425	−4979		
(7,2,0)	J	7	6	5	5	4	3	3	2	1
	K	1308	3681	−5596.600	4642.600	−2094	−5836.759	−1543.241	−2745	−5091

$$K = 0 \quad \text{for } k_1 \text{ even, } k_2 \text{ even, } 2 \le k_2 \le k_1 - 2,$$
$$= -3\{k_1 - 2k_2 \pm [(k_1 - 2k_2)^2 + 15(k_1+1)(k_1+3)]^{1/2}\}$$
$$\text{for } k_1 \text{ even, } k_2 \text{ odd, } 3 \le k_2 \le k_1 - 3,$$
$$= -9(k_1+3) \quad \text{for } k_1 \text{ even, } k_2 \text{ odd, } k_2 = k_1 - 1,$$
$$= 9(k_1+3) \quad \text{for } k_1 \text{ even, } k_2 \text{ odd, } k_2 = 1,$$
$$= 3\{2k_1 - k_2 + 3 \pm [(2k_1 - k_2 + 3)^2 + 15k_2(k_2+2)]^{1/2}\}$$
$$\text{for } k_1 \text{ odd, } k_2 \text{ even, } 2 \le k_2 \le k_1 - 3,$$
$$= -9(k_1 - 1) \quad \text{for } k_1 \text{ odd, } k_2 \text{ even, } k_2 = k_1 - 1,$$
$$= 6(2k_1 + 3) \quad \text{for } k_1 \text{ odd, } k_2 \text{ even, } k_2 = 0,$$
$$= -3\{k_1 + k_2 + 3 \pm [(k_1 + k_2 + 3)^2 + 15(k_1 - k_2)(k_1 - k_2 + 2)]^{1/2}$$
$$\text{for } k_1 \text{ odd, } k_2 \text{ odd, } 3 \le k_2 \le k_1 - 2,$$
$$= -6(2k_1+3) \quad \text{for } k_1 \text{ odd, } k_2 \text{ odd, } k_1 = k_2,$$
$$= 9(k_1 - 1) \quad \text{for } k_1 \text{ odd, } k_2 \text{ odd, } k_2 = 1. \qquad (50)$$

$J = k_1 - 2$: We have

$$K = -\tfrac{1}{2}\{(2k_1+1)(k_1+1)(2k_2 - k_1) \pm 6[-4k_1^2 k_2(k_1 - k_2) + k_1^4 + 2k_1^3 - k_1^2 - 2k_1 - 1]^{1/2}\}, \qquad (51)$$

valid for $2 \le k_2 \le k_1 - 2$ [if k_2 is outside these bounds, there is no degeneracy and we can use Eqs. (47) and (48)].

$J = k_1 - 3$: We have

$$K = -\tfrac{1}{2}\{(2k_1+1)(k_1+3)(2k_2 - k_1) \pm 6[-4k_1^2 k_2(k_1-k_2) + k_1^4 + 6k_1^3 - 9k_1^2 - 6k_1 + 9]^{1/2}\}$$
$$\text{for } 3 \le k_2 \le k_1 - 3$$
$$= \pm\tfrac{1}{2}(2k_1+1)(k_1+1)(k_1 - 6) \quad \text{for } k_2 = 2 \text{ or } k_1 - 2. \qquad (52)$$

For $k_2 = 0, 1, k_1 - 1$ or k_1, see (47) and (48).

$k_2 = 2$: We have

$$K = \tfrac{1}{2}\{(2k_1+1)(J-2)(J+3) \pm 6[J(J-1)(J+1)(J+2) + (2k_1+1)^2]^{1/2}\} \quad \text{for } k_1 - J \text{ even}$$
$$= \tfrac{1}{2}(2k_2+1)[J(J+1) - 12] \quad \text{for } k_1 - J \text{ odd} \qquad (53)$$

[the first formula holds for $2 \le J \le k_1 - 2$; otherwise there is no degeneracy—see (43)—(46)].

$k_2 = 3$: We have

$$K = \tfrac{1}{6}\{30k_1 - (2k_1 - 3)J(J+1) \pm 6[16k_1^2 - 4k_1 J(J+1) + J^4 + (J-3)(J-1)(2J+3)]^{1/2}\}$$
$$\text{for } k_1 - J \text{ even, } 3 \le J \le k_1 - 3, \qquad (54)$$
$$K = \tfrac{1}{6}\{30k_1 - (2k_1+3)J(J+1) \pm 6[16k_1^2 + 4k_1 J(J+1) + J^4 + (J-3)(J-1)(2J+3)]^{1/2}\}$$
$$\text{for } k_1 - J \text{ odd, } 3 \le J \le k_1 - 3.$$

For $J \le 2$ or $J \ge k_1 - 2$ see earlier formulas.

$k_2 = k - 2$ *and* $k_1 - 3$: These are contragredient to $k_2 = 2$ and $k_2 = 3$. Hence formulas (53) and (54) apply with reversed signs.

Further explicit formulas (for $J = 4, 5, k_1 - 4, k_1 - 5$, $k_2 = 4, 5, k_1 - 4, k_1 - 5$) could be obtained by solving cubic equations (that may in some cases reduce to quadratic or linear ones), and we could proceed even further by solving quartic equations. We have, however, decided not to proceed in this direction.

Let us make a few further comments:

1. The eigenvalues of the operators $X^{(i)}$ coincide for the $U(3)$ representation (m_{13}, m_{23}, m_{33}) and the $SU(3)$ representation $(m_{13} - m_{33}, m_{23} - m_{33}) \equiv (k_1, k_2)$.

2. For any self-contragredient representation, i.e., such that $m_{13} - m_{33} = 2(m_{23} - m_{33})$, and for any fixed value of J, the sum of all eigenvalues of $X^{(3)}$ corresponding to J equals zero. More precisely, one has

$$\sum_{m_{11}+m_{12}+m_{23}=J+3m_{33}} \left\langle \begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix} \middle| X^{(3)} \middle| \begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix} \right\rangle = 0. \qquad (55)$$

This property is evidently connected to the automorphism $T_i \to -T_i$, $L_i \to L_i$, for which $X^{(3)} \to -X^{(3)}$.

3. A given eigenvector $|JMK_t\rangle$ of $X^{(i)}$ belonging to a representation space of (m_{13}, m_{23}, m_{33}) is readily constructed if one knows all eigenvalues K_j belonging to (m_{13}, m_{23}, m_{33}). Indeed,

$$|JMK_t\rangle \sim \prod_{j \ne t} (X^{(i)} - K_j)\psi, \qquad (56)$$

where ψ is an arbitrary vector from the representation space of (m_{13}, m_{23}, m_{33}) such that

$$\langle \psi | JMK_t \rangle \ne 0.$$

4. CONCLUSIONS

The contents of this article can be summarized as follows:

1. We have shown that for an arbitrary semisimple group G and its semisimple subgroup H there exists only a finite number of independent scalars with respect to H in the enveloping algebra of G.

2. We have derived a generating function for the number of $O(3)$ invariants of any given order in the enveloping algebra of $SU(3)$. The method is quite general and can be applied to any (semisimple) group G and its (semisimple) subgroup H.

3. We have used the above results to prove that besides the Casimir operators of $SU(3)$ and angular momentum L^2 only two other independent $O(3)$ scalars exist in the enveloping algebra of $SU(3)$, namely $X^{(3)} = L_a T_{ab} L_b$ and $X^{(4)} = L_a T_{ab} T_{bc} L_c$ (both of these operators have already made an appearance in the literature[6,10,13]). Either of these operators (or an arbitrary nontrivial polynomial in $C^{(2)}$, $C^{(3)}$, L^2, $X^{(3)}$, $X^{(4)}$, and L_3) can be used to resolve the missing label problem in the $SU(3) \supset O(3) \supset O(2)$ reduction.

4. We consider the basis functions

$$|(m_{13}, m_{23}, m_{33}) JMK\rangle \qquad (57)$$

for irreducible representations of $U(3)$, where (m_{13}, m_{23}, m_{33}) label the $U(3)$ representation [$k_1 = m_{13} - m_{33}$, $k_2 = m_{23} - m_{33}$ for $SU(3)$], J is an eigenvalue of L^2, M of L_3, and K of $X^{(i)}$, i.e.,

$$X^{(i)}|(m_{13}m_{23}m_{33})JMK\rangle = K|(m_{13}m_{23}m_{33})JMK\rangle, \quad i = 3 \text{ or } 4.$$

We also make use of the Gel'fand—Tseitlin formalism to derive a simple algorithm for evaluating K for any representation. The values of K are computed numerically for a large number of representations of $SU(3)$ and presented in the tables (containing all representations of known physical interest). In the case when the multiplicity of the $O(3)$ representation J in the $SU(3)$ representation (k_1, k_2) is 1 or 2, we give explicit formulas for the eigenvalues K of $X^{(3)}$, in terms of k_1, k_2, and J [see Eqs. (43)–(54)]. They are, of course, in agreement with Table I.

The eigenvalues K are integer whenever there is no degeneracy in J. If there are two or more multiplets with the same J, then the sum of the eigenvalues is integer, although the individual K's are solutions of algebraic equations of order equal to the multiplicity of J in the given $SU(3)$ representation. The eigenvalues K corresponding to the same J in contragredient representations of $SU(3)$ differ by a sign in the case of $X^{(3)}$ and remain unchanged for $X^{(4)}$. For self-contragradient representations the sum of all $X^{(3)}$ eigenvalues corresponding to $O(3)$ multiplets with the same J equals zero: if there is only one multiplet with a given $J=J_0$, then its K equals zero.

We have chosen to present a smaller number of computer calculated eigenvalues of $X^{(4)}$ in Table II, than for $X^{(3)}$ in Table I. The computer programs we have used are available on request and are suitable for arbitrary representations of $SU(3)$. Similarly we have running programs for explicit construction of eigenvectors of $X^{(3)}$ and $X^{(4)}$ as linear combinations of Gel'fand–Tseitlin patterns, and also a program for calculating matrix elements of any polynomial of $U(3)$ generators relative to both the basis of patterns and to the basis (57).

It should also be mentioned that a large amount of literature related *inter alia* to the $SU(3)\supset O(3)\supset O(2)$ missing label problem exists. Besides the articles already quoted we mention the work of Biedenharn,[21] the review by Louck and Galbraith[22] (containing numerous references) and the recent article by Asherova and Smirnov.[23]

Let us make a few comments on physical applications of the results of this paper.

1. The fact that the basis functions (57) form an orthonormal set is particularly helpful, e.g., if we are interested in calculating matrix elements of some operator Q (a Hamiltonian, a term in a Hamiltonian, a transition operator, etc.) that commutes with $X^{(i)}$ since we will then obtain selection rules with respect to K. Similarly, if some polynomial $P(X^{(3)}, X^{(4)}, C^{(2)}, C^{(3)}, L^2, L_3)$ commutes with Q, rather than $X^{(3)}$ itself, then this operator P should be used to provide the missing label. It is certainly of interest that the algebra of such polynomials is finitely generated.

2. Various $O(3)$ scalars in the enveloping algebra of $SU(3)$ have been succesfully used as models for two- and three-body forces.[7,24] One implication of the present results is the following:

The only "fundamental" forces that can be introduced in an $SU(3)$ scheme with an $O(3)$ invariant interaction are two-body forces involving $C^{(2)}$ and L^2, three-body forces involving $C^{(3)}$ and $X^{(3)}$, and four-body forces, involving $X^{(4)}$. Any other forces can be represented as polynomials in the fundamental ones.

3. The formalism developed in this article has an amusing application in elementary particle physics. Indeed, the problem of constructing a state vector for N identical pions in a state with definite isospin T can be solved by embedding an $O(3)$ group, related to the isospin, into an $U(3)$ group.[25,26] The N-pion state will be characterized by the $U(3)$ labels N_1, N_2, N_3 (with $N=N_1+N_2+N_3$, $N_1\geq N_2\geq N_3\geq 0$), the isospin T, charge $Q=T_3$ and the degeneracy label K (the correspondence with the notations of the present article is $N_1=m_{13}$, $N_2=m_{23}$, $N_3=m_{33}$, $T=L$, $Q=L_0$). If K is identified with the eigenvalue of operator $X^{(3)}$ as in this article, it is possible to obtain rigorous limits on the charge distribution of pions in N-pion production, following from isospin conservation and Bose statistics alone. This can then be done for arbitrary values of the isospin T; previous considerations[25,26] were restricted to $T=0$ and 1, when no degeneracies occur. The results are presented in a separate article.[27]

Other group–subgroup chains of physical interest with a missing label problem are presently being considered. Work in progress on the Wigner supermultiplet scheme $SU(4)\supset SU(2)\times SU(2)$ (two missing labels) and also the schemes $SO(5)\supset SU(2)\times U(1)$ (one missing), $SO(5)\supset SO(3)$ (two missing), and $G_2\supset SO(3)$ (four labels missing).

ACKNOWLEDGMENTS

In conclusion we thank Professor M. Moshinsky for a very enlightening discussion and Dr. Boyer for discussions and for calling Ref. 13 to our attention. We also thank Dr. C.S. Kalman, Dr. E.G. Kalnins, Dr. R.T. Sharp, and Dr. K.B. Wolf for their comments. We are greatly indebted to Mrs. W. McKay who performed all the computer calculations, the results of which are presented in this article.

*Permanent address: Department of Physics, Johns Hopkins University, Baltimore, Maryland

†On leave from School of Mathematics, University of Minnesota, Minneapolis, Minn.

[1]N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).

[2]I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); 71, 1017 (1950); D. P. Zhelobenko, Usp. Mat. Nauk 17, 27 (1962) [Russian Math. Surv. 17, 1 (1962), Chap. 13].

[3]I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSR, Ser. Mat. 29, 1329 (1965) [Am. Math. Soc. Transl. Ser. 2 64, 116 (1967)].

[4]M. Gell-Mann, Phys. Rev. **125**, 1067 (1962); Y. Neéman, Nucl. Phys. **26**, 222 (1961).

[5]J. P. Elliott, Proc. Roy. Soc. A **245**, 128, 562 (1958).

[6]V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 967 (1960); **23**, 177 (1961).

[7]M. Harvey, *Advances in Nuclear Physics*, *Vol.* 1, edited by M. Baranger and E. Vogt (Plenum, New York, 1968).

[8]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968).

[9]R. T. Sharp, H. C. von Baeyer, and S. C. Pieper, Nucl. Phys. A **127**, 513 (1969); T. C. Haskell, B. G. Wybourne, and S. Feneuille, Physica **53**, 64 (1971); V. S. Devi, J. Math. Phys. **12**, 1732 (1971); J. Patera and R. T. Sharp, Nuovo Cimento A 12, 365 (1972).

[10]G. Racah, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).
[11]P. Winternitz I. Lukáč, and Ya. A. Smorodinsky, Yad. Fiz. 7, 192 (1968) [Sov. J. Nucl. Phys. 7, 139 (1968)]; A.A. Makarov, Ya. A. Smorodinsky, Kh. V. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061 (1967); N.W. Macfadyen and P. Winternitz, J. Math. Phys. **12**, 281 (1971)]; J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130 (1973); Ya. A. Smorodinsky and I.I. Tugov, Zh. Eksp. Teor. Fiz. 50, 653 (1966) [Sov. Phys. JETP **23**, 434 (1966)].
[12]W. Miller, Jr., "Lie Theory and Separation of Variables, Parts 1 and 2," SIAM J. Math. Anal. (to appear); E.G. Kalnins, "On the Separation of Variables for the Laplace Equation $\Delta\psi + X^2\psi = 0$ in Two and Three Dimensional Minkowski Space," Preprint CRM-319 (Université de Montréal, 1974); E.G. Kalnins and W. Miller, Jr. "Lie Theory and Separation of Variables. 3," J. Math. Phys. **15**, 1025 (1974).
[13]J.W.B. Hughes, J. Phys. A **6**, 48, 281 (1973).
[14]H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), pp. 251–52, 274–75.
[15]L.P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), p. 62.
[16]B. Meyer, Can J. Math. **6**, 155 (1953).
[17]I.M. Gel'fand, Mat. Sb. **26**, 103 (1950).
[18]J. Patera, J. Math. Phys. **14**, 279 (1973).
[19]B.G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970).
[20]J. Patera and D. Sankoff, *Tables of Branching Rules for Representations of Semisimple Lie Algebras* (Presses Université de Montréal, Montreal, 1973).
[21]L.C. Biedenharn, Phys. Letters **28B**, 537 (1969).
[22]J.D. Louck and H.W. Galbraith Rev. Mod. Phys. **44**, 540 (1972).
[23]R.M. Asherova and Yu F. Smirnov, Rep. Math. Phys. 4, 83 (1973).
[24]G.F. Nash, Nuovo Cimento B **42**, 8 (1966); **61**, 261 (1969).
[25]C.H. Llewellyn Smith and A. Pais, Phys. Rev. D 6, 2625 (1972).
[26]E. Chacon and M. Moshinsky, Phys. Rev. D 7, 2783 (1973).
[27]M. Moshinsky, J. Patera, R.T. Sharp, and P. Winternitz, Phys. Rev. D **10** (September, 1974).

Some topological and graphical aspects of phase contours

A. Shafee

International Centre for Theoretical Physics, Trieste, Italy
(Received 1 June 1973)

Using the topological properties of phase contours, the phase contour diagram of a function in the complex plane is reduced to the simpler form of a bichromatic multigraph, which contains all the combinatorial characteristics regarding links between zeros and singularities of the function. It is found that the saddle points of the phase play an essential role in any such qualitative global description of the function, as can be anticipated from Morse theory. In the case of the scattering amplitude, constraints from symmetries, periodicity, or unitarity are seen to yield simplifications with interesting results concerning factorization and asymptotic behavior.

1. INTRODUCTION

Qualitative methods of analysis[1] have been known to provide useful insight into the nature of the solutions of complicated dynamical problems, where a quantitative solution turns out to be impossible or extremely difficult to obtain. They are valuable in establishing existence and consistency conditions and in indicating peculiarities of the solution, so that the subsequent application of quantitative methods for an exact or approximate solution may be facilitated. It is of course also possible that the particular problem of interest has no quantitative content, so that qualitative methods alone can suffice.

The richness of the singularity spectrum of the strong interaction amplitude, which makes an analytical solution difficult, provides a nontrivial and interesting field for the application of certain qualitative methods, especially those of a topological nature. Of all the different approaches to the study of symmetry and consistency conditions for scattering amplitudes, that which most successfully exploits the global topological properties is probably the phase contour method,[2-9] where patterns of singularities and zeros of amplitude functions related by crossing or some other symmetry have been investigated using specific dynamical models, like Regge asymptotic behavior, or more general physical principles, like unitarity or hermiticity, but without resorting to numerics for most results. On the other hand, by analyzing the data with dispersion relations, one can obtain[10,11] the phenomenological picture of such patterns, showing intimate relations between specific zeros and poles of the amplitude. Such relations probably have more topological than geometric significance, because the complications from a complex background or kinematic peculiarities usually destroy the usefulness of metric-dependent geometrical concepts like linearity or convexity, while topological invariants are less likely to be affected by errors in the data or by perturbative correction terms in a theoretical study.

In previous applications of the phase contour method more emphasis was laid on the physical content than on the topological side of the approach. This left certain mathematically undesirable features in the treatment, e.g., ambiguity in the labeling of phase and arbitrariness in the spacing of phase contours. Besides, the crucial dependence on visual presentation made any generalization of such techniques to complex spaces of multiple dimensions obscure and impracticable. The usefulness of topological methods, especially of the critical point theory of Morse, [12-14,1] in the global analysis of complex functions has been of great interest to mathematicians in recent years. In this paper we intend to introduce the concept of topological invariants in the study of phase contour diagrams and show the relevance of critical points and of Morse theory in this context. In the case of a single complex variable we shall show how phase contour diagrams (PCD's) can be reduced to graphs containing all the combinatorial information relating poles, zeros, and critical points. We shall follow a rather heuristic line so that the formalism may remain sufficiently transparent to show its relation to the previous works with the phase contour method.

In Sec. 2 the phase contour diagram is redefined on a more abstract footing than previously. We use the concept of bundles of phase contours connecting sources of opposite polarity. In Sec. 3 we show how homotopical equivalence of contours allows us to derive bichromatic multigraphs[15] from the PCD's. We indicate the importance of the critical points of the function in a comprehensive topological description of these graphs and hence also of the original PCD's. In Sec. 4 we look into the consequences of physical constraints, like hermiticity, positivity over the cut from unitarity, asymptotic behavior, and periodicity in the patterns of zeros and poles.

2. PHASE CONTOUR DIAGRAMS

Let F be a complex function defined in an n-dimensional complex space C^n with points $z=(z_1, z_2, \ldots, z_n)$. The phase of the function is defined, as usual, as the real-valued function

$$\phi(z) = \operatorname{Im}\log F(z). \tag{2.1}$$

We define phase contours P_i in the space C^n as the connected sets of points $P_i \subset C^n$, such that $\phi(z)=K_i$ (a constant) for all $z \in P_i$. Since the n-dimensional complex space corresponds to a $2n$-dimensional real space, the P_i are, because of the single constraint (2.1), $(2n-1)$-dimensional real subspaces in this real space. If we consider functions of a single complex variable $z \equiv z_1$, the phase contours are simply the familiar contour curves in the complex z plane. Usually[2-11] a PCD is defined as the collection of P_i drawn at regular intervals Δ with $K_i = K_0 + i\Delta$ $(i=0, 1, 2, \ldots)$, on the whole z space or a section of this space. However, the arbitrariness of K_0 and Δ introduces an unsatisfactory discrete description of a generally continuous function ϕ. In this

 Copyright © 1974 American Institute of Physics

paper we shall define a phase contour diagram D as the collection of all P_i corresponding to every value K_i in the range of ϕ. According to this definition, any point where the phase ϕ is defined must lie on a phase contour P_i.

In a similar way one can define a modulus contour diagram with the modulus μ of F given by

$$\mu(z) = \mathrm{Re}\,\log F(z), \tag{2.2}$$

for all $z \in C^n$ where $\log F$ is defined. These again are $(2n-1)$-dimensional real subspaces, or one-dimensional curves if $n=1$. In the latter case, phase and modulus contours form two orthogonal families of curves, by Cauchy—Riemann conditions, spanning all points of the complex space where $\log F$ is defined. The logarithm has branch points at the origin and at infinity. Hence ϕ and μ are defined everywhere, except where F has a singularity or a zero and along cuts joining the zeros and the singular points. The behavior of phase contours in the neighborhood of a singular point or a zero depends on its type and order. In the neighborhood of an essential singularity or an exponential zero, phase or modulus contours are not well defined. We shall consider only isolated singular points and zeros of finite order. Let F be factorizable into the form

$$F(z) = (z - z_i)^{\alpha_i} f(z), \tag{2.3}$$

where α_i is a positive or negative integer, $f(z)$ being regular and nonzero at z_i. The phase ϕ is the sum of the phases of the two factors

$$\phi(z) = \alpha_i\,\mathrm{Im}\log(z - z_i) + \mathrm{Im}\log f(z). \tag{2.4}$$

Since the second term is regular near z_i, by choosing a disk small enough around z_i its variation can be made as small as desired, so that we can effectively replace it by a constant within the disk and on its boundary. At the boundary of the disk $z = z_i + re^{i\theta}$ we have

$$\phi(\theta) \approx \alpha_i\theta + \text{const}. \tag{2.5}$$

On a complete rotation around z_i, ϕ changes by $2\alpha_i\pi$, necessitating a branch cut. However, the different sheets of this logarithmic singularity differ only by multiples of this constant $2\alpha_i\pi$, changing the label ϕ_i of every phase contour P_i, leaving the P_i themselves and, hence, the PCD invariant. It is, therefore, sufficient to consider just one sheet.

If $\alpha_i > 0$, i.e., z_i is a zero, ϕ increases for an anticyclic rotation around z_i. We shall call z_i a *source* of strength α_i. Similarly, if $\alpha_i < 0$, i.e., if z_i is a singularity of $F(z)$, ϕ decreases for an anticyclic rotation around z_i, which will be called a *sink* of strength α_i, or a source of strength $-\alpha_i$. These terms are borrowed from the obvious electrostatic and hydrodynamic analogies. We shall also use the term "source" generally to indicate either a source or a sink.

Lemma 2.1: P_i are continuous open sets with sources and sinks as their limit points.

Proof: The continuity of P_i at regular points follows from ϕ being a harmonic function and satisfying Laplace's equation. Closed loops of P_i are not possible because then the orthogonal modulus contours entering the region bounded by P_i must either converge to one or more points, which is impossible even at singularities of F, or pass out of the region through another point of P_i. In the latter case, modulus contours must exist whose two points of intersection with P_i approach coincidence, i.e., the modulus contour becomes tangential to P_i. This too is forbidden by the orthogonality condition. Hence all P_i must be open and have singularities of $\log F$ as limit points.

Let B be the set of all P_i such that the subspace $X \subset C^n$ covered by the set $B = \{P_i\}$ is arcwise connected. We call B a *phase bundle*, in loose analogy with fiber bundles, as the connected and continuous set B may be considered to fibrate the space X, though the definition of the base space becomes complicated because different ends of different $P_i \in B$ may have different singularities as limit points. If all $P_i \in B$ have the same two limit points, we shall call it a closed bundle. Henceforth only closed bundles will be considered and for brevity these will be called simply bundles. We shall call the difference between the extremal phases in B its *thickness*.

Lemma 2.2 The limit points of a phase bundle of nonzero thickness are sources of opposite polarity.

Proof: If D_A is a small disk around the limit point A, with the boundary ∂D_A, we have seen in Eq. (2.5) that the P_i form a monotonic sequence on an arc of ∂D_A. Since different P_i cannot intersect, on a continuous deformation and translation of the arc from ∂D_A to ∂D_B, the boundary of a small disk around the other limit point B, the sequence will remain unaltered with respect to A, but the orientation with respect to B will be opposite to that at A. Hence, again from Eq. (2.5), the α at B must have a sign opposite to that of A.

Lemma 2.3: If B_{ab} and B_{bc} are two different bundles sharing the limit point $\underline{b}$, but with the other end points $\underline{a}$ and $\underline{c}$ different, and if $B_{ab} \cap B_{bc} \neq \phi$, then this intersection contains a critical point of the function.

Proof: If the intersection is nonzero, it must be the whole or part of the adjacent boundary contour. It cannot be the whole contour because, by assumption, the other limit points $\underline{a}$ and $\underline{c}$ are different. Hence it can only be part of the boundary contour which must at some point split into two branches to join the different limit points $\underline{a}$ and $\underline{c}$. At the branching point, a tangent to the contour must become indeterminate. From the definition of the phase, Eq. (2.1), and from the Cauchy—Riemann conditions, it can be seen that at the branching point

$$d(\ln F) = \frac{dF}{F} = 0. \tag{2.6}$$

However, F cannot be infinite at an analytic point of the function, so that we must have $dF = 0$, which makes the branching point of the contour a critical point of the functions ϕ and μ. If it is a nondegenerate critical point, i.e. if the Hessian

$$\left|\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\right| (x_1 = x,\ x_2 = y)$$

does not vanish, then it would have[12] two eigenvalues of opposite sign and the critical point would be a saddle point. We shall call the contour attached to the critical point, the critical value contour. Obviously the thickness of a bundle is equal to the difference between the phases on the critical value contours bounding it.

If α_i is a nonintegral real number, F has an algebraic branch point at z_i with a finite (if α_i is a rational fraction) or an infinite (if α_i is irrational) number of Riemannian sheets. If F has only algebraic singularities, it can be seen from Eqs. (2.3) to (2.5) that the PCD's on all sheets of F are identical, except for a constant shift in the labeling of all contours on each sheet. If the branch cuts are taken along the phase contours, no contours move from one sheet to another, so that each sheet becomes self-contained. Similarly, the logarithmic cut associated with poles and zeros of integral order are also made harmless by choosing it along a P_i in a bundle B. Then B becomes effectively the union of two bundles

$$B = B_1 \cup B_2 \tag{2.7}$$

separated by the cut. However, if explicit labeling of the phase is avoided (e.g., if we consider only properties of the differential $d\phi$, which is well defined and has no logarithmic cut at regular points of F, and is sufficient to determine the singularities and zeros of F as well as the critical points of ϕ), then B can be considered to be a single effective bundle with a thickness t_B given by

$$t_B = t_{B_1} + t_{B_2},$$

which is invariant with respect to the choice of the P_i chosen for the cut, i.e., with respect to the particular partition of B into B_1 and B_2.

By mapping a cut sheet stereographically in the usual way onto the unit sphere, the "point at infinity" can be treated as any other point. From Eq. (2.3) we can get a source–sink *duality* on this sphere; every source of strength α (positive or negative) at a finite point must be accompanied by a source of strength $-\alpha$ at the point at infinity.

Theorem 2.1: For a function with only algebraic singularities (including nonrational orders) the algebraic sum of the thickness of the phase bundles leaking into adjoining sheets through the cuts is zero.

Proof: Since F is by assumption completely factorizable into the form of Eq. (2.3), the point at infinity is a source of strength equal, but opposite in sign, to the algebraic sum of the strengths of sources at finite points. By Laplace's equation, no other sources exist on a particular sheet; hence, bundles crossing into another sheet must return to join a source of opposite strength on the same sheet, or be cancelled by a bundle of opposite thickness emerging from the uncompensated source passing into another sheet, or the bundle may have both end points on other sheets. In each case we can see that the net thickness of flux bundles passing out of the branch cuts is zero.

Corollary 2.1: The algebraic sum of the thickness of phase bundles passing into each algebraic branch cut is zero. This follows from the fact that all the sheets of an algebraic branch point are interconnected through the same branch cut, and the net flow of phase bundles into each sheet is zero (by Theorem 2.1). Therefore, all leakages into other sheets may be regarded as ineffective crossings of the branch cut which were mentioned by Eden *et al.*[2,3] However, this will not be true in general when different sheets have different singularity structures, as in the case of the physical scattering amplitude or a properly unitarized model of it.

3. GRAPHICAL REPRESENTATIONS

Given a bundle B, all $P_i \in B$ (except the extremals, which, being critical value contours, have branches and hence a more complicated topology) can be continuously transformed into one another, i.e., there exist mappings homotopic relative to the end points giving all the $P_i \in B - \partial B (\equiv \tilde{B})$. In other words, all $P_i \in \tilde{B}$ are derivable from mappings belonging to a unique homotopic equivalence class. This would also extend to the case of the compound bundle [Eq. (2.7)] of an algebraic function, if contours on other sheets are identified with their images on the original sheet. So far as the combinatorial problem of the linkages between zeros and poles is concerned, for many purposes it is sufficient to consider only the equivalence classes of contuours, as individual contours within a bundle are of little topological interest.

We shall call the set of singular points of the phase and the homotopically equivalent classes of P_i (or the open sets $\tilde{B}$) represented by arcs connecting the relevant limit points, a graphical representation of the PCD, or a phase contour graph (PCG). A PCG has the same homotopy type as the PCD from which it is derived.

In general a PCG would be a multigraph, i.e., multiple arcs between the same vertices would exist, provided they are homotopically inequivalent. We can give an arc a weight equal to the thickness of the bundle it represents. Obviously the algebraic sum of the weights of the arcs of ω_i^j meeting at a vertex i indicates the strength and type of its singularity:

$$\sum_j \omega_i^j = 2\pi\alpha_i, \tag{3.1}$$

where α_i is the index used in Eq. (2.3) to show the nature of the phase singularity at z_i. We shall use the sign convention that an anticyclic rotation of increasing phase in the PCD corresponds to an outgoing positive weight for the arc representing the bundle, in conformity with our use of the terms "source" and "sink" in the previous section.

Since $\tilde{B}_1 \cap \tilde{B}_2 = \phi$ for any two bundles B_1 and B_2 in a PCD, the arcs representing $\tilde{B}_1$ and $\tilde{B}_2$ cannot intersect either. Hence a PCG is a planar graph. If only arcs of finite weight are considered, we can see from Lemma 2.2, which states that bundles of finite thickness must connect sources of opposite sign, that the PCG must be a bichromatic graph, i.e., a graph where the vertices belong to two distinct classes and the arcs join vertices of only different types. The PCD's of an algebraic function being identical on all sheets, so will be the PCG's.

Lemma 3.1: The faces of a PCG are homotopic to the critical value contours and the associated critical points.

Proof: Let D be a PCD with sinks and sources S_i and phase bundles B_{ij}^k connecting S_i with S_j. Let G be the PCG of D with vertices v_i and arcs a_{ij}^k connecting v_i with

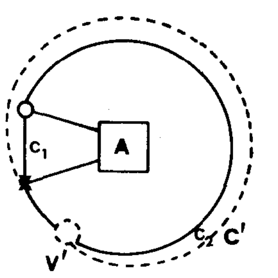

FIG. 1. A maximal bichromatic multigraph with periphery $C_1 \cup C_2$. A third vertex V' would result in a new arc C'. A is the inner part of the graph.

v_j. Let h be a set of homotopic mappings $h: \tilde{B}_{ij}^k \to a_{ij}^k$. But $D = \cup S_i \cup \tilde{B}_{ij}^k \cup \partial\tilde{B}_{ij}^k$ and $G = \cup v_i \cup a_{ij}^k \cup f_l$, where f_l are the faces. The mappings h being homotopic relative to the set $\{S_i\}$, v_i can be identified with S_i. Hence the rest of D is mapped onto the rest of G, i.e., $\cup\, \partial\tilde{B}_{ij}^k \to \cup f_l$ Matching individual connected components we get a one-to-one correspondence between the critical value contours of D and the faces of G.

Theorem 3.1: The PCG is maximally connected in the absence of degenerate critical points.

Proof: Let D be the PCD of the function and G its graphical representation. If G is not maximally connected, there exists a source–sink pair $(v_i, v_j) \in G$, which can be connected by an arc a_{ij} homotopically distinct from the arcs already present. Let h be the set of homotopic mappings $h: D \to G$. Now, $(h^{-1} a_{ij}) \cap_{m,n} \tilde{B}_{mn} = \phi$ (for $\tilde{B}_{mn}$ with $m=i$, $n=j$ does not exist). Hence $h^{-1} a_{ij} \subset \partial\tilde{B}_{ik} \cup \partial\tilde{B}_{jl}$, for some $k \neq j$ and some $l \neq i$. Let ∂D_i be the boundary of a small disk around i. A point on ∂D_i on the other side of B_{ik} must belong to another bundle, $\tilde{B}_{ik'}$. Similarly, there is another bundle $\tilde{B}_{jl'}$ separated, in part, by $\partial\tilde{B}_{jl'}$. So a_{ij} is the image of part of a critical value contour which has at least six endpoints: i, j, k, l, k', and l'. But by Morse theory a nondegenerate critical value contour in a two-dimensional manifold can only have four branches. Hence, if there is no degeneracy, $\tilde{B}_{ij}$ and its graphical image a_{ij} must already exist, making G a maximally connected graph.

Corollary 3.1: The faces of a PCG are quadrilaterals. The proof is similar to the case for maximal bichromatic graphs without multiple arcs. First, it is obvious that a face of a bichromatic graph can have only an even number of vertices and sides. As a face cannot be bounded by two homotopically distinct arcs, the minimum number of sides is four. If there are more than two vertices of either color, then, in addition to its two arcs connecting a vertex with adjacent vertices of opposite color, a diagonal can be drawn from it to a third vertex of opposite color on the periphery of the face. If the graph is maximal, such an arc must already exist, but the face would then split into two faces. Hence each face must contain exactly four vertices (two of each color) on its boundary.

From the PCG, two graphs of interest can be derived—the *dual graph* and the *medial graph*. The dual graph is constructed by joining the midpoint (in the topological sense, meaning any interior point) of every face with the midpoints of its boundary arcs. We shall also consider the domain exterior to the graph as a face, though unlike the other faces this will not be a quadrilateral.

Lemma 3.2: The outermost circuit of a maximal bichromatic multigraph consists of only two arcs.

Proof: If there is a second vertex of either color on this circuit, it can be joined to the adjacent vertex of opposite color with an arc surrounding the entire graph except the original linkage between the two vertices (Fig. 1) that forms a new periphery. But this is impossible because the graph is already maximally connected.

Theorem 3.2: The dual graph of the PCG is the graphical equivalent of the modulus contour diagram. By a graphical equivalent we mean that nonintersecting cycles of this dual graph represent all the different homological equivalence classes of modulus contour cycles of the modulus contour diagram.

Proof: The modulus contours, being orthogonal to the phase contours, should encircle the sources and sinks. The homological equivalence class of a modulus contour depends only on the particular pole or zero it encloses. If the modulus increases away from the region bounded by the contour we shall give the contour a cyclic sense. In this case we have a pole inside, and with opposite orientation and growth rate we can associate a zero. Since each face represents a critical point, joining adjacent critical points by lines bisecting the arcs of the original PCG, we essentially generate loops around the sources and sinks, and hence the homology classes of the modulus contours. The proper orientation of the cycles follows by giving each segment between successive critical points a direction agreeing with the Cauchy–Riemann conditions.

It is easy to verify, by simple counting, that the topology of the PCG, even when accompanied by the specification of the arc weights, cannot determine the function completely. For example, the function

$$F(z) = C[(z - z_1)(z - z_2)/(z - z_3)(z - z_4)] \tag{3.2}$$

will have a graph of four vertices, two of each color, and in general five topologically-distinct arcs (Fig. 2). There are only two independent weights corresponding to the loops, but the function F has ten real parameters from the five complex parameters C, z_1, z_2, z_3, and z_4. This indicates the existence of a class of transformations under which the topology and the weights of the arcs remain invariant.

A change in the topology occurs when one or more of the arc weights go to zero, which can happen (as we shall see in the next section) whenever the sinks and sources move into a pattern with a reflection symmetry. The weights remain unchanged if the critical values remain invariant under the transformation which may change the positions of the sources and the sinks and also of the critical points.

If $\bar{z} = \bar{z}(z_i)$ is a multivalued function giving the critical points $\bar{z}$ in terms of the sources and sinks z_i, it is ob-

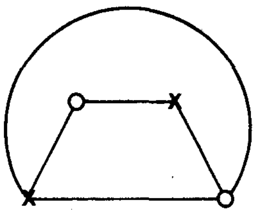

FIG. 2. A maximal bichromatic multigraph with four vertices.

tained by solving the condition for the critical point: $(df/dz)=0$, with $f=\log F$. If we now consider $f(\bar{z}, z_i)$ with $\bar{z}=\bar{z}(z_i)$, we get

$$df=\left(\frac{\partial f}{\partial \bar{z}}\frac{\partial \bar{z}}{\partial z_i}+\frac{\partial f}{\partial z_i}\right)dz_i=\frac{\partial f}{\partial z_i}\,dz_i, \tag{3.3}$$

because $(\partial f/\partial \bar{z})=0$ as $\bar{z}$ is a critical point. Hence for stationary critical values the infinitesimal transformations of the poles and zeros dz_i must satisfy the condition

$$\frac{\partial f(\bar{z}, z_i)}{\partial z_i}\,dz_i=0, \tag{3.4}$$

where $\bar{z}$ is obtained from the solution, in z, of

$$\frac{\partial f(z, z_i)}{\partial z}=0. \tag{3.5}$$

If $F(z, z_i)$ is factorizable into its sources and sinks, as in the case of an algebraic function

$$F(z, z_i)=\Pi F_i(z, z_i), \tag{3.6}$$

then the critical point is given by [with $F_i' \equiv (\partial F_i/\partial z)$]

$$\sum[F_i'(z-z_i)/F_i(z-z_i)]=0, \tag{3.7}$$

and the condition for weight-invariant transformations becomes

$$\sum[F_i'(\bar{z}-z_i)/F_i(\bar{z}-z_i)]\,dz_i=0. \tag{3.8}$$

All linear transformations

$$z_i \to az_i+b, \tag{3.9}$$

where a and b are complex constants indicating translation, rotation, or dilatation of the PCD on the complex plane, keep the weights invariant. This reduces the number of parameters of the function from $2V+2$ (where V is the total number of vertices) to $2V-2$, indicating that many other solutions to Eqs. (3.7) and (3.8) remain outside the class given by the transformations of (3.9).

The only nontrivial Betti number of a graph is the number of faces F in it. As we have seen before, this is equal to the number of saddle points of the phase, in the absence of any degeneracy resulting from symmetries or other special relations between the strengths and positions of the zeros and singularities of the function, and preventing the graph from being maximally connected.

Theorem 3.3: For a maximally connected bichromatic multigraph we have for the number of faces $F=V-2$, where V is the total number of vertices of either color.

The proof is by induction. We have seen in Lemma 3.2 that the circumference of a maximally connected bichromatic multigraph contains one vertex of each color. If a new vertex appears outside this graph, it can be connected to the vertex of opposite color by two homotopically distinct arcs, which, together with the two arcs of the previous circumference, form the four edges of a new quadrilateral face. With $V=4$ we get $F=2$, as can be verified by explicit construction. Hence, constructing graphs with $V>4$ by adding new vertices and consequently faces from one with $V=4$, we get the general relation $F=V-2$.

4. PHYSICAL CONSTRAINTS

Our discussion of the PCGs in the previous section was mostly concerned with algebraic functions with similar Riemannian sheets. Models for the scattering amplitudes with only meromorphic functions exist, which satisfy important constraints like asymptotic Regge behavior, crossing symmetry, direct and crossed channel poles (though on the real axis of the single sheet). Unitarity involves the introduction of branch points of nonalgebraic nature—except the elastic branch point.[16] However, treating the cuts as effective sinks and sources according to the amount of phase leaking into or out of them, it is possible to consider only the physical sheet or a submanifold in it instead of the complete Riemannian surface and apply some of our results to the subgraph belonging to this region.

A. Symmetries and factorization

We make the following observations:

(a) Hermitian analyticity—$A(S)=A^*(S^*)$—makes the PCD on the physical sheet, and hence the PCG on this sheet, symmetric on reflection by the real axis. Only the labels of the P_i change.

(b) Other symmetries may exist, giving left–right symmetry of the graph, e.g., the crossing symmetry of the A'^+ or B^- amplitudes of πN scattering or the $\pi^0\pi^0$ scattering amplitude. Crossing antisymmetry as in A'^- and B^+ πN-amplitudes would also produce symmetry on either side of the imaginary axis.

Theorem 4.1: Every reflection symmetry of the function leads to an increase in the number of components of the graph and the appearance of a degenerate critical value contour.

Proof: Let S be the line of symmetry dividing the graph G into identical subgraphs A and B. We prove that A and B are disconnected. If any arc connects A and B, it must either be symmetrical itself under reflection on S, or have a symmetrical partner. Since an arc (which represents a phase bundle of finite thickness) can connect only a source with a sink—by Lemma 2.2—reflection symmetry rules out a symmetrical arc. Similarly, a symmetric partner is made impossible by planarity. Hence A and B must be disconnected components of the graph of the function. Since no phase bundles cross S, S itself must be a phase contour, say with phase ϕ. Even if A (or B) is maximally connected, its periphery must contain a source and a sink (from Lemma 3.3). The phase contours with phase ϕ from these singulari-

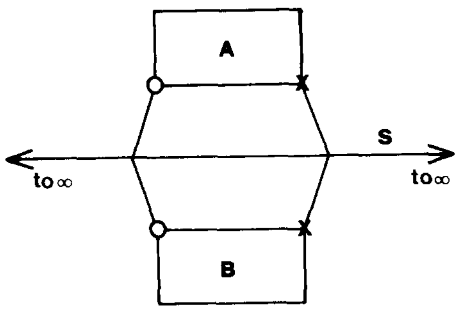

FIG. 3. Reflection symmetry: the subgraphs A and B are reflection symmetric with respect to the line S.

ties must be connected with S, because (by Lemma 2.1) all phase contours must have sources and sinks as their limit points. Hence S must be part of a contour with at least six components (Fig. 3). By Morse theory, a nondegenerate critical value contour can have only four components. So S must be a part of a degenerate critical value contour. The critical points can be either distinct or coincident. By the same arguments, it can be proved, in the case with vertices on the symmetry line, that though the graph does not split into disconnected components, it ceases to be maximally connected.

The existence of symmetries indicates the possibility of factorization of the amplitude into functionally similar components. Let $F(z)$ be a crossing-symmetric amplitude, i.e., $F(z)=F(-z)$. We can write the phase representation[17]

$$F(z)=\frac{\Pi_i[(z-z_i)(z+z_i)]}{\Pi_j[(z-z_j)(z+z_j)]}\exp\left(\frac{2}{\pi}\int_{z_0}^{\infty}\frac{\delta(z')dz'}{z-z'}\right), \qquad (4.1)$$

where $\pm z_i$ are the zeros, $\pm z_j$ the poles, and $\delta(z')$ the phase of the contours leaking out of the physical sheet through the cuts. In Eq. (4.1) we can see that $F(z)$ factorizes into $f(z)$ and $f(-z)$, with the two symmetric branch cuts and the symmetric poles and zeros separating into the two factors in any complementary combination.

However, factorization of the amplitude does not in general lead to factorization of the graph into subgraphs corresponding to the factors. Because of the requirement of planarity, the graph of a function must be different in general from the superposition or any simple interconnection of the graphs of its factorial components. Connectivity, being a global property, depends not only on the individual strengths of the relevant sources and sinks, but also on the nature and position of all other sources. We mentioned towards the end of Sec. 3 how the weights of the arcs of the graph of an algebraic function are controlled by the positions and strengths of all the sources on any sheet. For a more complicated function like the scattering amplitude, where each sheet is expected to have a different pole—zero structure, the poles and zeros on other sheets influence the weights of the arcs on the sheet of interest in a complicated way. On the other hand, we know from the phase representation that the amplitude is fully known on the physical sheet if its poles and zeros on the same sheet and its phase along the branch cuts of this sheet are known.

Since the full content of analyticity is contained in the branch cuts, the weaker topological constraints on the PCG's can also be obtained by replacing the cuts (and therefore the connected sheets) by an effective combination of sources and sinks with strength related to the phase leak into the cut. If there are no oscillations in the phase along the cut, then obviously a single source or sink would suffice for the whole cut. In the presence of oscillations the cut may act as a source or a sink locally, but, if the scale of such oscillations is small compared with the distance between the zeros and poles, we may take semilocal averages and reduce it to a single vertex of an effective strength.

B. Periodicity and infinite graphs

Periodic pole and zero structures in amplitudes induce periodicity in the PCG's with simplifications resulting in interesting predictions. Zero width dual models provide a nontrivial test case. However, the linearity of the graph destroys all faces and removes the possibility of obtaining some knowledge of the nondegenerate critical points, which undoubtedly exist in the physical amplitude and would appear in properly unitarized models, though at the price of breaking exact linearity and periodicity.

Let us take a function F, with poles along the real axis at $x_i=x_0+ia$ $[i=0,1,2,\ldots,\infty]$, and zeros also along the real axis and with the period b, $x_i'=x_0'+ib$ $[i=0,1,2,\ldots,\infty]$. If $a=b$, we shall have an infinite linear graph (Fig. 4) with alternate poles and zeros, except for a possible sequence of only poles or zeros at the finite end. At sufficiently high z values we can expect the end effects to be minimal and the weights of the arcs connecting adjacent poles and zeros to settle down to constant values depending on the separation between adjacent poles and zeros. That there are no arcs going to the point at infinity from the vertices far from the finite end can be visualized as follows. If we take a large circle with its whole circumference far from the finite end and with equal numbers of poles and zeros inside, then because of the constancy of the arc weights ω_1 and ω_2 (Fig. 4) the net phase leaking to adjacent vertices will be zero. Consequently, because of the equality of poles and zeros within the circle, the amount of phase going to infinity will also be zero. Hence we must have $\omega_1+\omega_2=2\pi$. The power behavior at infinity is essentially determined by the uncompensated zeros or poles at the finite end as well as some phase leakage to infinity from all vertices near the origin. If we sum the total weight α from all the poles and zeros, except the point at infinity, with α_0 equal to the number of unpaired zeros or poles,

$$\alpha=\alpha_0+1-1+1-1+\cdots. \qquad (4.2)$$

The sum of the series is indeterminate but bounded between α_0+1 and α_0. If we use the factorized form for the function, then

$$F=C\,\frac{\Pi_{n=0}^{\infty}(z-x_0'-na)}{\Pi_{m=0}^{\infty}(z-x_0-ma)}=C'\,\frac{\Gamma((x_0/a)-(z/a))}{\Gamma((x_0'/a)-(z/a))}, \qquad (4.3)$$

with $C'=Ca^k$, k being the integral part of $(x_0-x_0')/a$. This has an asymptotic power behavior off the real axis:

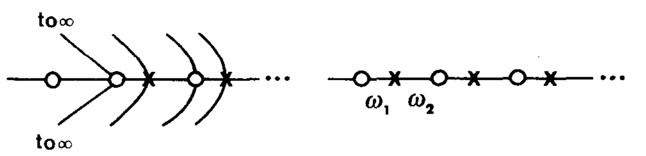

FIG. 4. Periodic linear graph with alternate poles and zeros.

$$F(z) \xrightarrow{|z|\to\infty} (z/a)^{(x_0-x_0')/a}, \tag{4.4}$$

where the index is actually the number of unmatched zeros at the finite end together with a fraction indicating the leakage of phase bundles from the nearby matched pairs as an end effect. That the analytic method is more powerful and gives the exact index, compared with the partial indeterminacy of the graphical method, can be expected from the fact that the latter does not use the separation between the zeros and the poles, but only the number of the excess. This uncertainty could possibly be remedied if the weights could be calculated without, of course, using full analyticity properties. However, as we have already seen, the exact weights of arcs cannot be determined by only graphical and topological means, because different functions may have the same graphs but different, or even the same, weights. The constraints from periodic symmetry would work only for vertices away from the finite end.

C. Unitarity and asymptotic behavior

We have just seen how even a purely meromorphic function could have a nonintegral asymptotic power behavior due to an infinite number of poles and zeros. An algebraic function also, in general, must have nonintegral power behavior, because the source or sink at the point at infinity must be equal and opposite in sign to the algebraic sum of the strengths of all the sources and sinks in the finite plane, on each sheet as the sheets are identical.

For a scattering amplitude the nature of the branch points are generally unknown. From the unitarity equation it can be shown[16] that the elastic branch point is of the square root type. The other branch points are expected to be infinite-sheeted, with no symmetry to make the PCD's on different sheets identical, as in the case of the purely algebraic function.

However, we can still think of the total phase flux moving out of the physical sheet into the branch points or into the singularities of the unphysical sheets to be absorbed by an effective sink. For an amplitude without any Born poles and with identical left- and right-hand cuts, e.g., the symmetric pion—pion amplitude $A(\pi^0\pi^0 \to \pi^0\pi^0)$, we get a very simple PCG (Fig. 5) with two equal effective sources representing the two cuts, and one at infinity that balances them and gives the asymptotic behavior. For $4m_\pi^2 > t \geq 0$, the imaginary part of the amplitude must remain positive throughout the cuts. Hence phase can vary, at most, from 0 to π over each side of a cut, giving each cut a maximum possible strength of $\alpha = +1$. So the sink at infinity has, at most, strength $\alpha = -2$. Or,

$$A(s,t) \geq s^{-2}, \tag{4.5}$$

as $|s| \to \infty$; $0 \leq t < 4m_\pi^2$. This result is identical to Jin and Martin's[18] classic lower bound on the scattering amplitude. However, we have ignored the possibility of logarithmic terms, either at infinity or over the cuts. It is only to be expected that such a simple qualitative approach, which does not employ dispersion relations or the properties of Herglotz functions,[19] cannot be as powerful as Jin and Martin's more detailed analysis. Nevertheless, a graphical picture does give us a more natural insight into the result which can be hidden in the details of analytical calculations.

5. CONCLUSIONS

We have tried in this work to indicate the topological aspects of the phase contour method for the analysis of scattering amplitudes, with some help from the results of Morse theory. Although the difference in the topological nature of ordinary contours and critical value contours is well known from Morse theory, we have been able to show also that this difference in the homotopy types in the simple case of a one-dimensional complex manifold leads to the different components of a unique graph representing all topologically equivalent phase contour diagrams. We believe the phase contour graph is easier to handle mathematically and should provide as much qualitative, particularly combinatorial, information regarding links between zeros and singularities of the amplitude as the full PCD. Observing that such a graph must in general be a maximally connected bichromatic multigraph, we found numerical relations between the number of poles and zeros and the number of critical points of an algebraic function on any sheet. We also saw that periodicity can give us simple predictions about asymptotic behavior, and in the case of the physical scattering amplitude, unitarity can constrain the strengths of the vertices representing the cuts and hence the asymptotic power behavior—a result previously obtained by Jin and Martin using dispersion relations. We have observed how the number of components of the graph can indicate the presence of symmetries of the function in the complex plane.

Although the use of PCG's is intuitive and heuristic and the results obtainable from it are only qualitative, this method probably concerns the more basic aspects of the scattering function and avoids the manifold prob-

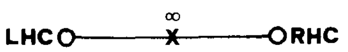

FIG. 5. PCG with two cuts and point at infinity.

lems of detailed quantitative analysis. However, the full power of Morse theory and the topological aspects of PCD's can probably be felt in the multidimensional case, e.g., for production amplitudes with zeros and poles in many complex variables, with higher dimensional simplicial complexes generalizing the concept of phase contour graphs.

ACKNOWLEDGMENTS

The author is indebted to Professor Abdus Salam as well as the International Atomic Energy Agency and UNESCO for the kind hospitality at the International Centre for Theoretical Physics, Trieste. He would also like to thank Professor A.O. Barut for reading the manuscript.

[1]L.È. Èl'sgol'c, *Qualitative Methods in Mathematical Analysis* (American Mathematical Society, Providence, R.I., 1964).
[2]C.B. Chiu, R.J. Eden, and C-I. Tan, Phys. Rev. Lett. **20**, 406 (1968)
[3]C.B. Chiu, R.J. Eden, and C-I. Tan, Phys. Rev. **170**, 1490 (1968).
[4]R.J. Eden and C-I. Tan, Phys. Rev. **170**, 1516 (1968).
[5]R.J. Eden and C-I. Tan, Phys. Rev. **172**, 1583 (1968).
[6]C.B. Chiu, R.J. Eden, M.B. Green, and F. Guerin, Phys. Rev. **182**, 1669 (1969).
[7]C.B. Chiu, R.J. Eden, and M.B. Green, Phys. Rev. **185**, 1734 (1969).
[8]G.D. Kaiser, Phys. Rev. **178**, 2372 (1969).
[9]G.D. Kaiser, Phys. Rev. **183**, 1499 (1969).
[10]S. Jorna and J.A. McClure, Nucl. Phys. B **13**, 68 (1969).
[11]A. Shafee, Nucl. Phys. B **35**, 556 (1971).
[12]J. Milnor, *Morse Theory* (Princeton U.P., Princeton, N.J., 1969).
[13]M. Morse, *Topological Methods in the Theory of Functions of a Complex Variable* (Princeton U.P., Princeton, N.J., 1947).
[14]M. Morse and S.S. Cairns, *Critical Point Theory in Global Analysis and Differential Topology* (Academic, New York, 1969).
[15]O. Ore, *Four-Colour Problem* (Academic, New York, 1967).
[16]R.J. Eden, P.V. Landshoff, D. Olive, and J.C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
[17]M. Sugawara and A. Tubis, Phys. Rev. **130**, 2127 (1963).
[18]Y.S. Jin and A. Martin, Phys. Rev. **135**, B1369 and B1375 (1964).
[19]J.A. Shohat and J.D. Tamarkin, *The Problem of Moments* (American Mathematical Society, Providence, R.I., 1963).

Limits in systems exhibiting a one-dimensional phase transition

William L. Truex and Joseph L. Strecker

Department of Physics, Wichita State University, Wichita, Kansas 67208
(Received 15 September 1972)

Recent spherical Ising models by Strecker and Gersch which exhibit one-dimensional phase transitions have shown that at the critical temperature $\epsilon \to \gamma^{2/3}$, whereas inside the one phase region $\epsilon \sim \gamma^2$. The models were unable to shed light on the behavior of this function as it passed through the vapor dome. Speculation might lead one to believe that a new type of singularity is present. Instead, we show that the exponent continuously changes from 2/3 to 2. A simple cubic polynomial controls the behavior of the ϵ function.

1. INTRODUCTION

Interest in one-dimensional phase transitions is found in a number of fields, namely, superconductivity, polymer chemistry, thermodynamics, and statistical mechanics. Recently there has arisen the possibility that a new type of phase transition may be in the offing.

The topic of one-dimensional phase transitions is important in the theory of superconductivity where recent discussions have taken place as to whether a whisker can become superconducting.[1] Further, there is interest in Little's[2] work on long chain polymers which may become superconducting. These long polymers often are basically one-dimensional systems.

Using purely thermodynamic arguments, Landau and Lifshitz[3] show that long range order is not possible in a one-dimensional system with finite range forces. Their arguments are relatively simple and employ thermodynamic potential functions.

From the point of view of statistical mechanics models which undergo a phase transition in one dimension are of interest especially because proof or evidence for the existence of one-dimensional phase transitions is still inconclusive. Existence or nonexistence depends markedly upon the model employed. We do have Van Hove's[4] famous work wherein he shows that a one-dimensional system cannot exhibit a phase transition if the forces are of finite range. But even when one goes to long range forces the results as to whether there will be a phase transition depends upon the model. The delicate nature of this type of transition is appreciated when one makes a comparison of various similar models; the Kac model[5] gives no phase transition whereas those of Kac, Uhlenbeck, and Hemmer,[6] and Strecker[7] and Gersch[8] do.

Recently Thouless[9] and Dyson[10] have argued—not rigorously—that a phase transition of a completely new kind will possibly arise in a one-dimensional system in which the strength of the interaction behaves as r^{-2}. Some models, notably that of Carpenter and Strecker,[11] which do exhibit one-dimensional phase transitions have interactions which behave as $r^{-\alpha}$, $1 < \alpha < 2$; $\alpha = 2$ remains elusive and generally poses an unsolvable problem.

We ought also to mention recent work by Anderson and Yuval[12] in which a one-dimensional model has analogs in the Kondo effect.

One of the unsolved problems in the work of Gersch[8] and Strecker[7] is the behavior of the limiting function $\epsilon \to \gamma^{2/3}$ and $\epsilon \to \gamma^2$, the behavior of the function as one transverses through the point which marks the onset of the phase transitions. It is the work of this article to examine more closely this function and to prove that the function is basically a simple cubic polynomial (Fig. 1) which displays a continuous behavior as one moves from the one-phase region into the two-phase region.

These works by Strecker[7] and by Gersch[8] show the existence of a one-dimensional phase transition and the manner in which the transition arises due to the peculiar behavior of the saddle point. For both of these models, a normal saddle point exists for all positive values of the temperatures. In both models, the range of the interaction between elements of the system is specified by a parameter $1/\gamma$. The phase transition occurs at a critical temperature T_c as the range of the interaction becomes infinite ($\gamma \to 0$).

If $\gamma \neq 0$ the following applies. As $T \to 0$ the saddle point approaches a limiting value z_0 that is independent of γ. Let ϵ be the difference between the saddle point z_s and the point z_0. Then both Strecker and Gersch found that ϵ was proportional to $\gamma^{2/3}(\epsilon \sim \gamma^{2/3})$ at the critical temperature T_c and ϵ was proportional to γ^2 when the temperature was well below T_c. There was no evaluation to show how the exponent of γ went from 2/3 to 2 as the temperature dropped below T_c.

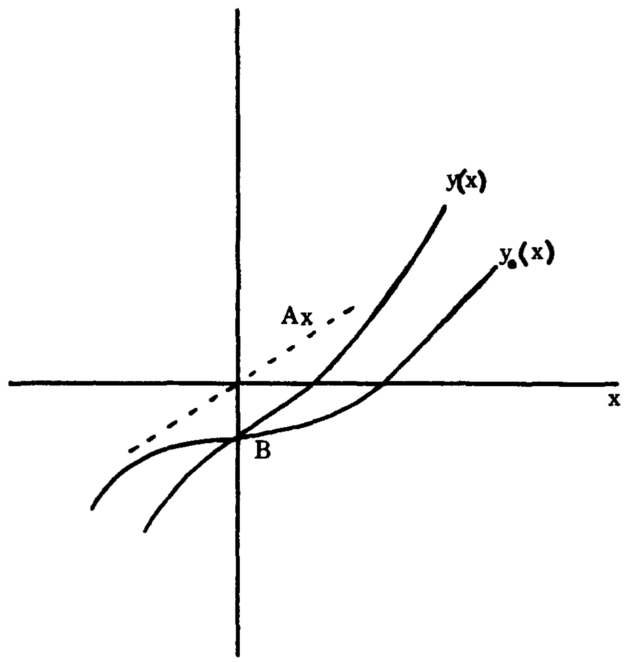

FIG. 1. Behavior of curve $y(x) = x^3 + Ax - B$ where $x \equiv \epsilon^{1/2}$. $y_0(x)$ is $y(x)$ for $A = 0$. $y(x)$ has a single root and $y_0(x)$ has an inflection point at $x = 0$. For $A \neq 0$ the linear term Ax (dotted line) is superimposed on $y_0(x)$ to produce $y(x)$ (broken curve). The different rates at which $\epsilon \to 0$ for both regions $T = T_c$ and $T < T_c$ are continuously produced.

 Copyright © 1974 American Institute of Physics

The behavior of the function in the intermediate region has been left unsolved; in the past the function describing this behavior could not be evaluated as one passes continuously through the transition region.

In this article we show that the behavior of the function is a smoothly continuous function and we show how one passes continuously from the γ^2 behavior well within the transition region to the behavior when the thermodynamic parameters are exactly at the vapordome ($\gamma^{2/3}$). The behavior is not one of the sharp singularities as we might guess, but a smooth function, a simple cubic polynomial (Fig. 1).

In this paper we will use the spherical Ising model of a one-dimensional spin system developed by Berlin and Kac[13] with a square well interaction between spins. The "width" of the square well will be m lattice sites and its "depth" will be J/m. The range parameter corresponding to $1/\gamma$ above is m. The quantity ϵ for this model will vary like $m^{-2/3}$ at $T = T_c$ and m^{-2} for $T \ll T_c$ in exactly the same fashion as in the model of Strecker and Gersch (replace $1/\gamma$ by m). We finally obtain an expression that explicitly shows how γ varies *continuously* from $m^{-2/3}$ to m^{-2} as T decreases from T_c.

2. SYNOPSIS OF THE SPHERICAL MODEL

The partition function for the spherical model[13] is given in Eq. 2.1. (We assume everyone is familiar—at least remotely—with the approximations and assumptions which lead to Eq. 2.1.)

$$Z = \frac{A^{-1}}{2\pi i}\int_{\alpha_0 - i\infty}^{\alpha_0 + i\infty} dS e^{NS} \prod_{j=1}^{N}\left(\frac{\pi}{S - \frac{1}{2}\beta\lambda_j}\right)^{1/2}$$
$$= \frac{A^{-1}}{2\pi i}\pi^{N/2}\int_{\alpha_0 - i\infty}^{\alpha_0 + i\infty} dS \exp\left(NS - \tfrac{1}{2}\sum_j \ln(S - \tfrac{1}{2}\beta\lambda_j)\right). \tag{2.1}$$

We will now define

$$\lambda_j' = 2\lambda_j/\lambda_1 \quad \text{and} \quad S = \tfrac{1}{2}\beta\lambda_1 z. \tag{2.2}$$

Then in terms of these quantities, (2.1) becomes

$$Z = \frac{A^{-1}}{2\pi i}\pi^{N/2}\frac{1}{2}\beta\lambda_1\int_{z_0 - i\infty}^{z_0 + i\infty} dz \exp\left\{\frac{1}{2}N\beta\lambda_1 z - \frac{1}{2}\sum_j\left[\frac{1}{2}\beta\lambda_1 z - \frac{1}{2}\beta\left(\frac{\lambda_1\lambda_j'}{2}\right)\right]\right\}$$
$$= \frac{A^{-1}}{2\pi i}\pi^{N/2}\frac{1}{2}\beta\lambda_1 \exp\left[-\frac{N}{2}\ln\left(\frac{\beta\lambda_1}{2}\right)\right]\int_{z_0 - i\infty}^{z_0 + i\infty} dz \times \exp\left(\frac{1}{2}N\beta\lambda_1 z - \frac{1}{2}\sum_j\left(z - \frac{1}{2}\lambda_j'\right)\right)$$
$$\times z_0 = \frac{2\alpha_0}{\beta\lambda_1} > \frac{2}{\beta\lambda_1}\frac{1}{2}\beta\lambda_1 = 1. \tag{2.3}$$

Next define

$$f(z) = \lim_{N\to\infty}\frac{1}{N}\sum_{j=2}^{N}\ln\left(z - \frac{1}{2}\lambda_j'\right)$$
and
$$g(z) = \tfrac{1}{2}\beta\lambda_1 z - \tfrac{1}{2}f(z) \tag{2.4}$$

For N large, (2.3) becomes approximately

$$Z = \frac{A^{-1}}{2\pi i}\pi^{N/2}\frac{1}{2}\beta\lambda_1 \exp\left[-\frac{N}{2}\ln\left(\frac{\beta\lambda_1}{2}\right)\right] \times \int_{z_0 - i\infty}^{z_0 + i\infty} dz (z-1)^{-1/2} e^{Ng(z)}. \tag{2.5}$$

The evaluation of the function $f(z)$ will be carried out later in the sections devoted to specific models.

The integral in (2.5) can be approximated by the method of steepest descents if a saddle point z_s can be found. The evaluation by the method of steepest descents yields

$$Z = \frac{A^{-1}}{2\pi i}\pi^{N/2}\frac{1}{2}\beta\lambda_1 \exp\left[-\frac{N}{2}\ln\left(\frac{\beta\lambda_1}{2}\right)\right] \times (z_s - 1)^{-1/2} i e^{Ng(z_s)}\left[\frac{2\pi}{Ng''(z_s)}\right]^{1/2}$$
$$= \frac{\beta\lambda_1\Gamma(N/2)}{4N^{(1/2)(N-1)}}\exp\left[-\frac{N}{2}\ln\left(\frac{\beta\lambda_1}{2}\right) + Ng(z_s)\right] \times [2\pi N(z_s - 1)g''(z_s)]^{-1/2} \tag{2.6}$$

where the saddle point z_s, if it exists, satisfies the conditions

$$g'(z_s) = 0 \quad \text{and} \quad g''(z_s) > 0. \tag{2.7}$$

In terms of the function $f(z)$, (2.7) is

$$\beta\lambda_1 = f'(z_s) \quad \text{and} \quad f''(z_s) < 0. \tag{2.8}$$

The saddle point will turn out to be real. It must also be greater than 1 because the real part of z in the integral (2.5) was restricted to be greater than 1.

To calculate the thermodynamic properties of our system in the limit as $N \to \infty$, we will use the free energy per spin ψ defined by

$$-\frac{\psi}{kT} = \lim_{N\to\infty}\frac{1}{N}\ln Z. \tag{2.9}$$

Using (2.6) in (2.9) we obtain

$$-\frac{\psi}{kT} = -\frac{1}{2} - \frac{1}{2}\ln\frac{\lambda_1}{kT} + \frac{\lambda_1 z_s}{2kT} - \frac{1}{2}f(z_s). \tag{2.10}$$

From ψ we may obtain the energy, entropy, and specific heat per particle by differentiating.

$$U = kT^2\frac{d}{dT}\left(-\frac{\psi}{kT}\right), \tag{2.11}$$

$$C = \frac{dU}{dT} \tag{2.12}$$

$$S = -\frac{d\psi}{dT}. \tag{2.13}$$

Using (2.10) in (2.11) gives

$$U = kT^2\left(\frac{1}{2T} - \frac{\lambda_1 z_s}{2kT^2} + \frac{\lambda_1}{2kT}\frac{dz_s}{dT} - \frac{1}{2}f'(z_s)\frac{dz_s}{dT}\right)$$
$$= \tfrac{1}{2}kT - \tfrac{1}{2}\lambda_1 z_s, \tag{2.14}$$

where the two terms containing dz_s/dT in the first line cancel because of (2.8). Then (2.12) shows

$$C = \tfrac{1}{2}k - \tfrac{1}{2}\lambda_1\frac{dz_s}{dT}. \tag{2.15}$$

Equation (2.13) is

$$S = -\tfrac{1}{2}k - \tfrac{1}{2}k \ln\frac{\lambda_1}{kT} + \tfrac{1}{2}k + \tfrac{1}{2}\lambda_1\frac{dz_s}{dT} - \tfrac{1}{2}kf(z_s) - \tfrac{1}{2}kTf'(z_s)\frac{dz_s}{dT} = -\tfrac{1}{2}k \ln\frac{\lambda_1}{kT} - \tfrac{1}{2}kf(z_s). \tag{2.16}$$

The dz_s/dT terms cancelled because of (2.8).

3. SQUARE WELL POTENTIAL

The square well potential is a modification of the nearest neighbor interaction. Each spin site interacts with its nearest m neighbors on either side while the interaction energy between a pair of spins is decreased by the same factor m. On account of this, the total energy per spin with all spins aligned is a constant independent of m and is the same as in the nearest neighbor case ($m = 1$). This energy will be $-J$ with J positive.

The energy of a spin configuration for this potential is given by

$$E = -\frac{1}{2}\sum_{ij} M_{ij}\epsilon_i\epsilon_j$$

with

$$M_{ij} = C_{N+j+1-i},$$

$$C_i = \begin{cases} \dfrac{J}{m}, & i = 2, 3, \ldots, m+1, N, N-1, \ldots, N-m+1 \\ 0, & \text{otherwise} \end{cases}$$

At first sight, this system appears to approach the ideal case of noninteracting spins in the limit of $m \to \infty$ since the interaction of a spin with its neighbors goes to zero. However, our model will be shown to undergo a phase transition at the temperature

$$T_c = 2J/k. \tag{3.1}$$

The reason for this nonideal behavior is the existence of a nonzero potential energy per spin that doesn't vanish even as $m \to \infty$. The critical temperature T_c is the temperature at which the thermal energy per spin $\tfrac{1}{2}kT$ equals the energy J.

Due to the development in Sec. 2, the problem of deducing the thermodynamic behavior of our model has been reduced to finding the saddle point z_s satisfying (2.8) and the function $f(z)$ and its derivatives. The renormalized eigenvalues of the interaction matrix M are

$$\lambda'_k = \frac{2}{m}\sum_{j=1}^{m} \cos\frac{2\pi}{Nj}(k-1) \tag{3.2}$$

From (2.6) the function $f(z)$ is

$$f(z) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=2}^{N} \ln\left(z - \frac{1}{m}\sum_{j=1}^{m}\cos\frac{2\pi}{Nj}(k-1)\right)$$

which becomes

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} d\omega \ln\left(z - \frac{1}{m}\sum_{j=1}^{m}\cos jw\right) = \frac{1}{\pi}\int_0^{\pi} d\omega \ln\left(z - \frac{1}{m}\sum_{j=1}^{m}\cos jw\right). \tag{3.3}$$

This integral cannot be evaluated in closed form so it will be approximated in Appendix B. The results show that

$$f(z) = \ln z + \frac{\ln m}{m}\zeta(z), \tag{3.4a}$$

where $\zeta(z)$ is a function of z and m and can be bounded independent of z and m. When we speak of bounding a function independently of z, we are speaking of z real and greater than or equal to one. Also,

$$\left|\frac{\ln m}{m}\zeta(z)\right| < \frac{1}{z} \tag{3.4b}$$

for z sufficiently large.

The saddle point Eq. (2.8) for this model is

$$\frac{1}{\pi}\int_0^{\pi} dw^{-1}\left(z_s - \frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-1} = \frac{2J}{kT}, \tag{3.5}$$

where we obtained $f'(z)$ by differentiating (3.3). In (3.5), the integral cannot be evaluated in closed form. The left side of (3.5) must be approximated. The details of this approximation are carried out in Appendix A. The results are

$$\frac{1}{z_s} + \frac{1}{\pi\sqrt{\alpha(z_s-1)}}\tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z_s-1}\right)^{1/2}\right] + \frac{\eta(z_s)}{m} = \frac{2J}{kT}, \tag{3.6a}$$

where $\eta(z)$ is a function of z and m that can be bounded independent of z or m. Also,

$$\left|\frac{1}{\pi\sqrt{\alpha(z-1)}}\tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\right] + \frac{\eta(z)}{m}\right| < \frac{1}{z^2} \tag{3.6b}$$

for z sufficiently large. For future ease of discussion, we will give each of the three terms on the left side of (3.6) its own number.

$$\frac{1}{z_s} \tag{3.7a}$$

$$\frac{\eta(z_s)}{m} \tag{3.7b}$$

$$\frac{1}{\pi\sqrt{\alpha(z_s-1)}}\tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z_s-1}\right)^{1/2}\right]. \tag{3.7c}$$

By (3.6b), we see that a first approximation to z_s is

$$z_s = kT/2J \tag{3.8}$$

for $2J/kT \ll 1$. This approximation is actually quite good for values of T very close to the critical temperature T_c. We will prove this statement in the following paragraphs.

We define the temperature T' by

$$2J/kT' = 1 - m^{-2/3}. \tag{3.9}$$

The exponent $-2/3$ on the m is chosen for later convenience. The approximations we will develop for $T > T'$ and $T < T'$ will have the same order of magnitude at T' when T' is defined by (3.9). We now contend that (3.8) is a first approximation for all $T \geq T'$. This is proved by the use of the following self-consistent argument.

Assume for a moment that (3.7a) is the only sizable term in the saddle point equation (3.6a) for all $T \geq T'$.

If this is true then (3.8) really is a first approximation to z_s. The saddle point equation is approximately

$$1/z_s = 2J/kT.$$

For $T \geq T'$ we have by using (3.9) in the last equation

$$\frac{1}{z_s} = \frac{2J}{kT} < \frac{2J}{kT'} = 1 - m^{-2/3}$$

or

$$z_s > 1/(1 - m^{-2/3}) \cong 1 + m^{-2/3}.$$

Then $z_s - 1 > m^{-2/3}$ and we have for (3.7c)

$$\frac{1}{\pi\sqrt{\alpha(z_s - 1)}} \tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z_s - 1}\right)^{1/2}\right] < \frac{1}{2\sqrt{(\frac{1}{6}m^2)(m^{-2/3})}} = \left(\frac{3}{2}\right)^{1/2} m^{-2/3},$$

where we replaced the arctangent by its maximum $\pi/2$ and α by its approximate value $\frac{1}{6}m^2$. For large m, this is clearly negligible compared to (3.7a). Term (3.7b) is also small since $\eta(z_s)$ is bounded. Hence, for $T \geq T'$, the neglect of (3.7b) and (3.7c) introduces an error of at most order $m^{-2/3}$.

If we assume for $T \gg T'$ that (3.7b) and (3.7c) behave like c/z^2 in accordance with (3.6b) where C may depend on m but is of order $m^0 = 1$, then a slightly better approximation to the saddle point z_s than (3.8) may be obtained.

For these large values of T, the saddle point equation is nearly

$$\frac{1}{z_s} + \frac{C}{z_s^2} = \frac{2J}{kT}.$$

Multiplying through by z_s^2 and solving the resulting quadratic equation gives $z_s = \{1 + [1 + 2c(2J/kT)]^{1/2}\}/[2(2J/kT)]$. Since $2J/kT \ll 1$, the radical may be expanded by the binomial expansion and the approximation $z_s = (kT/2J) + C$ obtained. This result may be obtained more easily by writing the saddle point equation in the form

$$z_s = (kT/2J)[1 + (C/z_s)]$$

and reiterating the first approximation (3.8):

$$z_s = \frac{kT}{2J}\left[1 + \left(C\Big/\frac{kT}{2J}\right)\right] = \frac{kT}{2J} + C. \tag{3.10}$$

In the region where T is close to T' but is larger than T' we can get another improvement on (3.8). In this region of temperature (3.7c) is much larger than (3.7b) even though it is still much smaller than (3.7a). The argument of the arctangent is large so we may set the value of the arctangent to be $\pi/2$. The saddle point equation is approximately

$$\frac{1}{z_s} + \frac{1}{2\sqrt{\alpha(z_s - 1)}} = \frac{2J}{kT}.$$

Since the second term on the left is much smaller than the first one, we will rewrite the equation in the form

$$z_s = \frac{kT}{2J}\left(1 + \frac{z_s}{2\sqrt{\alpha(z_s - 1)}}\right)$$

and use (3.8) and the method of successive approximations to obtain

$$z_s = \frac{kT}{2J}\left(1 + \left\{\frac{kT}{4J}\left[\alpha\left(\frac{kT}{2J} - 1\right)\right]^{-1/2}\right\}\right) \tag{3.11}$$

valid for $T \geq T'$ and near T'.

Now we are ready to consider the region $T < T'$. Since z_s will be within a range of about $m^{-2/3}$ of 1 and $\eta(z_s)/m$ is a relatively slowly varying function of z_s, we will replace this term by its value at $z = 1$. The argument of the arctangent will be much larger than 1 so that we will use the series

$$\tan^{-1}x = \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \cdots, \quad x > 1 \tag{3.12}$$

in (3.7c). The saddle point equation (3.6a) becomes

$$\frac{1}{z_s} + \frac{\eta(1)}{m} + \frac{1}{\pi\sqrt{\alpha(z_s - 1)}}\left[\frac{\pi}{2} - \frac{m}{\pi}\left(\frac{z_s - 1}{\alpha}\right)^{1/2} + \frac{1}{3}\left(\frac{m}{\pi}\right)^3\left(\frac{z_s - 1}{\alpha}\right)^{3/2} - \cdots\right] = \frac{2J}{kT} \tag{3.13}$$

Define ϵ by the equation

$$z_s = 1 + \epsilon \tag{3.14}$$

and substitute this in (3.13). We already know that the largest value ϵ we may obtain is something on the order of $m^{-2/3}$. Therefore, we may use the binomial expansion to write $1/z_s$ in terms of ϵ. Then (3.13) becomes

$$1 - \epsilon + \epsilon^2 - \cdots + \frac{\eta(1)}{m} + \frac{1}{\pi\sqrt{\alpha}\,\epsilon}\left[\frac{\pi}{2} - \frac{m}{\pi}\left(\frac{\epsilon}{\alpha}\right)^{1/2} + \frac{1}{3}\left(\frac{m}{\pi}\right)^3\left(\frac{\epsilon}{\alpha}\right)^{3/2} - \cdots\right] = \frac{2J}{kT}.$$

If we multiply through by $\epsilon^{1/2}$ and collect like powers of ϵ together, we have

$$\frac{1}{2\sqrt{\alpha}} + \epsilon^{1/2}\left(1 - \frac{2J}{kT} + \frac{\eta(1)}{m} - \frac{m}{\alpha\pi^2}\right) - \epsilon^{3/2}\left(1 - \frac{m^3}{3\alpha^2\pi^4}\right) + \cdots + (-1)^n\epsilon^{(2n+1)/2}\left[1 - \frac{1}{(2n+1)\pi\sqrt{\alpha}}\left(\frac{m}{\pi\sqrt{\alpha}}\right)^{2n+1}\right] + \cdots = 0. \tag{3.15}$$

First look at the general term

$$(-1)^n\epsilon^{(2n+1)/2}\left[1 - \frac{1}{(2n+1)\pi\sqrt{\alpha}}\left(\frac{m}{\pi\sqrt{\alpha}}\right)^{2n+1}\right] \tag{3.16}$$

for $n \geq 2$. The second term in the brackets is seen to always be of order $1/m$ since α is approximately $\frac{1}{6}m^2$. Therefore, the 1 in the brackets is the largest term by far. The magnitude of the entire term is then determined by the factor $\epsilon^{(2n+1)/2}$. We already know that for $T < T'$ that ϵ will be smaller than a quantity about the size of $m^{-2/3}$. The higher order terms form an alternating series and since the magnitude of the terms in this series is decreasing, then the total contribution has an absolute value smaller than that of the first term. The first term is $m^{(5/2)(-2/3)} = m^{-5/3}$ or smaller in size. We will see shortly that this is $m^{-1/3}$ smaller than the smallest possible value of any lower power of ϵ. Therefore, we will neglect the terms in $\epsilon^{5/2}$ and higher in the saddle point equation. The dependence of ϵ on the temperature will be essentially determined by the terms up to order $\epsilon^{3/2}$.

The $\epsilon^{3/2}$ term has the same form as the general term. Approximating α by $\frac{1}{6}m^2$ gives for this term

$$-\epsilon^{3/2}\left(1-\frac{12}{m\pi^4}\right). \tag{3.17}$$

This term will be larger than the neglected terms by a factor of $1/\epsilon$ which will be at least $m^{2/3}$.

The term in $\epsilon^{1/2}$ is

$$\epsilon^{1/2}\left(1-\frac{2J}{kT}+\frac{\eta(1)}{m}-\frac{6}{m\pi^2}\right), \tag{3.18}$$

α was replaced by $\frac{1}{6}m^2$. This term is larger than the neglected terms by a factor of about

$$\frac{1}{\epsilon^2}\left(1-\frac{2J}{kT}+\frac{\eta(1)}{m}-\frac{6}{m\pi^2}\right).$$

The expression $1-2J/kT$ can have a magnitude anywhere from 0 to ∞. The largest value of ϵ is about $m^{-2/3}$ so that we can see the smallest possible size of this term is larger than the neglected terms (3.16) by a factor of $m^{1/3}$ if we assume that $\eta(1)/m-6/m\pi^2$ is of order $m^0=1$.

We will show shortly that the results are not much changed even if $\eta(1)/m-6/m\pi^2=0$.

The constant term $1/2\sqrt{\alpha}\cong(\frac{3}{2})^{1/2}(1/m)$ is clearly much larger than the neglected terms.

The saddle point equation is essentially

$$\epsilon^{3/2}+\epsilon^{1/2}\left(\frac{2J}{kT}-1+\frac{6}{m\pi^2}-\frac{\eta(1)}{m}\right)-\frac{1}{m}\sqrt{\frac{3}{2}}=0 \tag{3.19}$$

when all the negligible terms are discarded. This is a cubic equation in $\epsilon^{1/2}$ whose real solution is

$$\epsilon^{1/2}=\left\{\frac{1}{2m}\left(\frac{3}{2}\right)^{1/2}+\left[\frac{3}{8m^2}+\frac{1}{27}\left(\frac{2J}{kT}-1+\frac{6}{m\pi^2}-\frac{\eta(1)}{m}\right)^3\right]^{1/2}\right\}^{1/3}$$
$$+\left\{\frac{1}{2m}\left(\frac{3}{2}\right)^{1/2}-\left[\frac{3}{8m^2}+\frac{1}{27}\left(\frac{2J}{kT}-1+\frac{6}{m\pi^2}-\frac{\eta(1)}{m}\right)^3\right]^{1/2}\right\}^{1/3} \tag{3.20}$$

This equation show the dependence of the saddle point on the interaction range m and the temperature T for $T<T'$.

At the temperature T_c defined by (3.1), $2J/kT-1=0$, and by Eq. (3.20),

$$\epsilon^{1/2}\cong\left(\frac{1}{m}\sqrt{\frac{3}{2}}\right)^{1/3}$$

or

$$\epsilon\cong\left(\frac{3}{2}\right)^{1/3}m^{-2/3}. \tag{3.21}$$

ϵ is proportional to the $-2/3$ power of m. Even if $\eta(1)/m-6/m\pi^2=0$, these results are not changed.

On the other hand, if $2J/kT-1\gg m^{-2/3}$ then this term becomes the significant one in (3.20). In this case ϵ will be much smaller than $m^{-2/3}$ and an approximate value for the saddle point can be obtained directly from the saddle point equation (3.19) by dropping the 3/2 power of ϵ from the equation. This term is negligible in this case.

The approximation yields

$$\epsilon\cong\left[\frac{3}{2}\Big/\left(\frac{2J}{kT}-1\right)^2\right]m^{-2} \tag{3.22}$$

which shows that ϵ is proportional to m^{-2}.

These results, that $\epsilon\sim m^{-2/3}$ at the critical temperature and $\epsilon\sim m^{-2}$ sufficiently below the critical temperature, were previously obtained by Strecker.[7] In his paper, it was shown that $\epsilon\sim\gamma^{2/3}$ at the critical temperature and $\epsilon\sim\gamma^2$ below the critical temperature (where $1/\gamma$ corresponds to m in the present paper) but the means by which the exponent changed from 2/3 to 2 could not be obtained. Equation (3.20) shows how this transformation takes place in terms of m.

To summarize: The width of the $m^{-2/3}$ behavior of ϵ can be estimated from (3.20). ϵ will go like $m^{-2/3}$ as long as the term containing $2J/kT-1$ is very close to zero. If $2J/kT-1$ becomes of order $m^{-2/3}$, then the cube of this term compares favorably with the term $3/8m^2$ inside the square root. If $2J/kT-1$ increases beyond this point, then this term becomes the largest one and the exponent of m moves toward 2. The width of the $m^{-2/3}$ behavior is therefore the range of temperatures for which $2J/kT-1\lesssim m^{-2/3}$.

The thermodynamic functions can now be computed using the various approximations for the saddle point in the different temperature regions. Care must be exercised in using these approximations so that one does not go beyond their domain of validity. The procedure is relatively straightforward. We can use the approximation (3.10) in (2.22) to obtain

$$U=\tfrac{1}{2}kT-J(kT/2J+C)=-CJ$$

for $T\gg T'$. This shows that our assumption that C is constant in (3.10) is incorrect. It actually approaches 0 as $T\to\infty$. This is true because when the spins are completely randomized at high temperatures, the energy per spin will be zero.

For $T\ll T_c$, we would use (3.22) in (2.22) and obtain

$$U=\tfrac{1}{2}kT-J\,[1+(3/2m^2)(2J/kT-1)^{-2}]$$
$$=\tfrac{1}{2}kT-J+\text{orders }(1/m^2).$$

This actually holds up to a temperature very close to T_c when m is large. Similar considerations for other temperature ranges shows that the energy in the limit as $m\to\infty$ is

$$U=\begin{cases}\tfrac{1}{2}kT-J, & T<T_c\\ 0, & T>T_c\end{cases}.$$

The specific heat for this limit is

$$C=\begin{cases}\frac{1}{2}k, & T<T_c\\ 0, & T>T_c\end{cases}.$$

The discontinuity in the specific heat occurs at the transition temperature T_c.

4. CONCLUSION

The spherical model of a one-dimensional spin system with a square well interaction clearly shows a phase transition. The behavior of the saddle point as a function of the interaction range m is very similar to the saddle point behavior in the models of Strecker and Gersch. These behaviors are $\epsilon \sim m^{-2/3}$ at T_c and $\epsilon \sim m^{-2}$ for $T \ll T_c$. The change from $m^{-2/3}$ to m^{-2} could not be observed in the models of Strecker and Gersch. In the present model, an expression is obtained which shows this transition.

APPENDIX A: THE SADDLE POINT EQUATION

In this appendix we approximate the function

$$f'(z)=\frac{1}{\pi}\int_0^{\pi}d\omega\left[z-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right]^{-1} \tag{A1}$$

defined for all real $z>1$ and derive some elementary properties of the saddle point which is the solution of

$$f'(z_s)=\frac{1}{\pi}\int_0^{\pi}d\omega\left(z_s-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-1}=\frac{2J}{kT}, \tag{A2}$$

which is equation (2.16) with $\lambda_1=2J$.

First of all, the saddle point is a monotone increasing function of the temperature. This is seen by differentiating both sides of (A2) with respect to T and solving the resulting equation for dz_s/dT.

$$\frac{dz_s}{dT}=\frac{2J}{kT^2}\;\frac{1}{\pi}\int_0^{\pi}d\omega\left(z_s-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-2}. \tag{A3}$$

Since $z_s>1$, the right side is positive. We will show shortly that the integral (A1) increases without bound as $z\to 1^+$ ($z\to 1$ and $z>1$). Also, the right side of the saddle point equation (A2) blows up as T approaches absolute zero. From the comments in this paragraph we then see that as the temperature rises from absolute zero, the saddle point, z_s, increases monotonically from the value 1.

For z much greater than 1, the function $f'(z)$ approaches the function $1/z$. In fact, from (A1) we have

$$\begin{aligned}&\left|\frac{1}{\pi}\int_0^{\pi}d\omega\left(z-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-1}-\frac{1}{z}\right|\\&=\frac{1}{\pi z}\left|\int_0^{\pi}d\omega\left\{\left(1-\frac{1}{mz}\sum_{j=1}^{m}\cos j\omega\right)^{-1}-1\right\}\right|\\&<\frac{1}{\pi z}\int_0^{\pi}d\omega\left[\frac{1}{1-1/z}-1\right]=\frac{1}{z}\left(\frac{1}{1-1/z}-1\right),\end{aligned} \tag{A4}$$

where we have used the fact that

$$\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\le 1. \tag{A5}$$

Equation (A4) places an upper limit on the difference between $f'(z)$ and $1/z$. It is valid for any $z>1$ and shows that as $z\to 1$, $f'(z)$ doesn't blow up faster than $1/(z-1)$. For $z\gg 1$, (A4) shows that the difference between $f'(z)$ and $1/z$ is less than a quantity of order $1/z^2$.

We will be primarily interested in the behavior of $f'(z)$ in the neighborhood of $z=1$. This is because for m large, $f'(z)$ approaches $1/z$ very closely when z is only a small distance from 1. The unboundedness of $f'(z)$ as $z\to 1$ is squeezed into a very narrow band about $z=1$ as $m\to\infty$. This behavior, as we shall see, is responsible for the phase transition in the limit as $m\to\infty$. This type of result has already been obtained by Strecker.[7] To simplify the determination of $f'(z)$ in the neighborhood of $z=1$, we will define

$$f_1'(z)=\frac{1}{\pi}\int_0^{\pi/m}d\omega\left(z-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-1}. \tag{A6}$$

and

$$f_2'(z)=\frac{1}{\pi}\int_{\pi/m}^{\pi}d\omega\left(z-\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\right)^{-1}. \tag{A7}$$

Then

$$f'(z)=f_1'(z)+f_2'(z). \tag{A8}$$

First consider $f_1'(z)$. In the interval $[0,\pi/m]$ we will represent $1/m\sum_{j=1}^{m}\cos j\omega$ by a power series,

$$\frac{1}{m}\sum_{j=1}^{m}\cos j\omega=\frac{1}{m}\sum_{j=1}^{m}\sum_{K=0}^{\infty}(-1)^K\frac{(j\omega)^{2K}}{(2K)!}=\sum_{K=0}^{\infty}(-1)^K A_K \tag{A9}$$

where

$$A_K=\frac{1}{m}\frac{\omega^{2K}}{(2K)!}\sum_{j=1}^{m}j^{2K}.$$

For each k, $A_K>0$, $1/m\sum_{j=1}^{m}\cos j\omega$ is represented by an alternating series.

For $1\le j\le m$ we have $j^{2(K+1)}\le m^2 j^{2K}$. Therefore,

$$\sum_{j=1}^{m}j^{2(K+1)}<m^2\sum_{j=1}^{m}j^{2K}$$

or

$$\sum_{j=1}^{m}j^{2(K+1)}\Big/\sum_{j=1}^{m}j^{2K}<m^2.$$

With this result we see that

$$\begin{aligned}\frac{A_{K+1}}{A_K}&=\frac{1}{m}\frac{\omega^{2(K+1)}}{[2(K+1)]!}\sum_{j=1}^{m}j^{2(K+1)}\Big/\frac{1}{m}\frac{\omega^{2K}}{(2K)!}\sum_{j=1}^{m}j^{2K}\\&<\frac{m^2\omega^2}{(2K+2)(2K+1)}<\frac{\pi^2}{(2K+2)(2K+1)}<1.\end{aligned}$$

Then $(1/m)\sum_{j=1}^{m}\cos j\omega$ is represented by an alternating series, the magnitude of whose terms is decreasing. On account of this, it is bounded above and below by any two consecutive terms in the power series.

In particular,

$$1-\alpha\omega^2\le\frac{1}{m}\sum_{j=1}^{m}\cos j\omega\le 1-\alpha\omega^2+\beta\omega^4$$

$$\alpha=\frac{1}{6}\left(m^2+\frac{3}{2}m+\frac{1}{2}\right),$$

$$\beta=\frac{1}{120}\left(m^4+\frac{5}{2}m^3+\frac{5}{3}m^2-\frac{1}{6}\right) \tag{A10}$$

With the bounds (A10) we can deduce

$$\frac{1}{\pi}\int_0^{\pi/m} d\omega \frac{1}{z-1+\alpha\omega^2} \le f_1'(z) \le \frac{1}{\pi}\int_0^{\pi/m} d\omega \frac{1}{z-1+\alpha\omega^2-\beta\omega^4} \quad \text{(A11)}$$

By integrating we get

$$[1/\pi\sqrt{\alpha(z-1)}]\tan^{-1}[\omega\sqrt{\alpha/(z-1)}]|_0^{\pi/m} \le f_1'(z) \le \{\pi[\alpha^2+4\beta(z-1)]^{1/2}\}^{-1}$$
$$\times\Bigg[\Bigg\{1/2\Bigg[\frac{\alpha}{2\beta}+\left(\frac{\alpha^2}{4\beta^2}+\frac{z-1}{\beta}\right)^{-1/2}\Bigg]^{1/2}\Bigg\}$$
$$\times\ln\Bigg(\Bigg\{\Bigg[\frac{\alpha}{2\beta}+\left(\frac{\alpha^2}{4\beta^2}+\frac{z-1}{\beta}\right)^{1/2}\Bigg]^{1/2}+\omega\Bigg\}$$
$$\times\Bigg\{\Bigg[\frac{\alpha}{2\beta}+\left(\frac{\alpha^2}{4\beta^2}+\frac{z-1}{\beta}\right)^{1/2}\Bigg]^{1/2}-\omega\Bigg\}\Bigg]$$
$$+\Bigg\{1/\Bigg[\left(\frac{\alpha^2}{4\beta^2}+\frac{z-1}{\beta}\right)^{1/2}-\frac{\alpha}{2\beta}\Bigg]^{1/2}\Bigg\}$$
$$\times\tan^{-1}\Bigg\{\omega/\Bigg[\left(\frac{\alpha^2}{4\beta^2}+\frac{z-1}{\beta}\right)^{1/2}-\frac{\alpha}{2\beta}\Bigg]^{1/2}\Bigg\}\Bigg]\Bigg|_0^{\pi/m}$$

and after inserting the limits of integration and rearranging,

$$\frac{1}{m}\Bigg\{\frac{m}{\pi[\alpha(z-1)]^{1/2}}\tan^{-1}\Bigg[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\Bigg]\Bigg\} \le f_1'(z) \le \frac{1}{m}$$
$$\times\Bigg[m\left(\frac{2\beta}{\alpha^3}\right)^{1/2}\quad \pi\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\Bigg]$$
$$\times\Bigg[\Bigg\{1\quad\Bigg[\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}-1\Bigg]^{1/2}$$
$$\times\tan^{-1}\Bigg\{\frac{\pi}{m}\left(\frac{2\beta}{\alpha}\right)^{1/2}\quad\Bigg[\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}-1\Bigg]^{1/2}\Bigg\}$$
$$\times\Bigg\{1/2\Bigg[1+\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\quad 1/2$$
$$\times\ln\Bigg(\Bigg\{\left(\frac{\alpha}{2\beta}\right)^{1/2}\Bigg[1+\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\Bigg]^{1/2}+\frac{\pi}{m}\Bigg\}$$
$$\times\Bigg\{\left(\frac{\alpha}{2\beta}\right)^{1/2}\Bigg[1+\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\Bigg]^{1/2}-\frac{\pi}{m}\Bigg\}\Bigg)\Bigg] \quad \text{(A12)}$$

Consider the expression

$$\Bigg[m\left(\frac{2\beta}{\alpha^3}\right)^{1/2}\Big/\pi\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\Bigg]. \quad \text{(A13)}$$

For any given value of m, this expression attains its maximum value at $z=1$ (the maximum value for the allowed values of z, namely, $z \ge 1$.). Since for large m, α is approximately $\frac{1}{6}m^2$ and β is approximately $(1/120)m^4$, then (A13) is approximately $\sqrt{18/5}\ (\pi\sqrt{1+(6/5)(z-1)})^{-1}$. We can bound (A13) independently of m.

The same type of considerations apply to both the argument and coefficient in front of the logarithm term on the right side of (A12). The argument of the logarithm term is also bounded away from zero independent of m.

Next consider the arctangent term on the right side of (A12),

$$\Bigg[\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}-1\Bigg]^{-1/2}$$
$$\times\tan^{-1}\Bigg\{\frac{\pi}{m}\left(\frac{2\beta}{\alpha}\right)^{1/2}\Big/\Bigg[\left(1+\frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}-1\Bigg]^{1/2}\Bigg\}. \quad \text{(A14)}$$

It is easy to see that for $z\neq 1$, (A14) may be bounded by a quantity depending on z but independent of m. If we multiply (A14) by (A13) and retain just the leading terms for $z-1 \ll 1$, we have

$$[\text{Eq. (A13)}][\text{Eq. (A14)}] \cong \frac{m}{\pi[\alpha(z-1)]^{1/2}}\tan^{-1}\Bigg[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\Bigg], \quad z-1 \ll 1.$$

By comparing this with Eq. (A12) and using the comments following (A13), we see that

$$f_1'(z) = \frac{1}{\pi[\alpha(z-1)]^{1/2}}\tan^{-1}\Bigg[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\Bigg]+\frac{1}{m}\eta_1(z) \quad \text{(A15)}$$

where $\eta_1(z)$ is a function of z and m that can be bounded independent of m.

Next consider $f_2'(z)$. In the interval $[\pi/m, \pi]$ we will use the trigonometric identity

$$\frac{1}{m}\sum_{j=1}^{m}\cos j\omega = \frac{1}{2m}\sum_{j=1}^{m}(e^{ij\omega}+e^{-ij\omega})$$
$$= \frac{1}{2m}\Bigg[e^{i\omega}\frac{1-e^{im\omega}}{1-e^{i\omega}}+e^{-i\omega}\frac{1-e^{-im\omega}}{1-e^{-i\omega}}\Bigg]$$
$$= \frac{\sin(m+\frac{1}{2})\omega}{2m\,\sin(\omega/2)}-\frac{1}{2m} \quad \text{(A16)}$$

Since

$$2x/\pi \le \sin x \le x \quad \text{for } 0 \le x \le \pi/2, \quad \text{(A17)}$$

then

$$\frac{\sin(m+\frac{1}{2})\omega}{2m\,\sin(\omega/2)}-\frac{1}{2m} < \frac{1}{2m\,\sin(\omega/2)} \le \frac{1}{2m(2/\pi)(\omega/2)} \le \frac{1}{2m(2/\pi)(\pi/2m)} = \frac{1}{2}.$$

Since $z>1$, there will be no singularity in the integrand of $f_2'(z)$. From (A7) and (A16) we have

$$f_2'(z) = \frac{1}{\pi}\int_{\pi/m}^{\pi} d\omega\left(z+\frac{1}{2m}-\frac{\sin(m+\frac{1}{2})\omega}{2m\,\sin(\omega/2)}\right)^{-1}$$
$$= \frac{1}{\pi}\frac{1}{(z+1/2m)} \times \int_{\pi/m}^{\pi} d\omega\Bigg[\sin(\omega/2)\Big/\left(\sin(\omega/2)-\frac{\sin(m+\frac{1}{2})\omega}{2mz+1}\right)\Bigg]$$
$$= \frac{1-1/m}{z+1/2m}+\frac{1}{\pi(z+1/2m}\,\frac{1}{(2mz+1)} \times \int_{\pi/m}^{\pi} d\omega\Bigg[\sin(m+\tfrac{1}{2})\omega\Big/\left(\sin(\omega/2)-\frac{\sin(m+\frac{1}{2})\omega}{2mz+1}\right) \quad \text{(A18)}$$

A very important result is that this last integral is a function of z and m that is bounded independent of m.

The proof is involved and will be carried out in a separate appendix (see Appendix C). Equation (A18) then shows that

$$f_2'(z) = \frac{1}{z} + \frac{\eta_2(z)}{m}, \tag{A19}$$

where $\eta_2(z)$ is a function of z and m that is bounded independent of m.

By combining (A15) and (A19) we see that

$$f'(z) = \frac{1}{\pi[\alpha(z-1)]^{1/2}} \tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\right] + \frac{1}{z} + \frac{\eta(z)}{m}, \tag{A20}$$

where $\eta(z)$ is a regular function of z and m is a neighborhood of the real line greater than 1 and is bounded independent of m. From the paragraph following (A5) we see that

$$\left|\frac{1}{\pi[\alpha(z-1)]^{1/2}} \tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\right] + \frac{1}{m}\eta(z)\right| < \frac{1}{z^2} \quad \text{for } z \gg 1. \tag{A21}$$

APPENDIX B: THE FUNCTION $f(z)$

Herein, we will develop the function

$$f(z) = \frac{1}{\pi}\int_0^\pi d\omega \ln\left[z - \frac{1}{m}\sum_{j=1}^m \cos j\omega\right] \tag{B1}$$

defined for all real $z > 1$. Some of the properties of $f(z)$ will also be derived.

For $z \gg 1, f(z)$ asymptotically approaches $\ln z$. We have that

$$\begin{aligned}\left|\frac{1}{\pi}\int_0^\pi d\omega \ln\left[z - \frac{1}{m}\sum_{j=1}^m \cos j\omega\right] - \ln z\right| &= \frac{1}{\pi}\left|\int_0^\pi d\omega \ln\left[1 - \frac{1}{mz}\sum_{j=1}^m \cos j\omega\right]\right| \\ &< \frac{1}{m}\left|\int_0^\pi d\omega \ln\left[1 - \frac{1}{z}\right]\right| = -\ln\left[1 - \frac{1}{z}\right].\end{aligned} \tag{B2}$$

Equation (B2) shows that for large $z, f(z)$ can differ from $\ln z$ by a quantity at most of magnitude $1/z$ since $\ln(1+x) \cong x$ when $|x| \ll 1$. Equation (B2) provides a bound for $f(z)$ that is independent of m for all real $z > 1$. For $z \to 1$, however, this bound is bad since $f(z)$ may have a logarithmic singularity at $z = 1$.

To determine the behavior of $f(z)$ in the neighborhood of $z = 1$ we follow the following route. Define the functions

$$f_1(z) = \frac{1}{\pi}\int_0^{\pi/m} d\omega \ln\left[z - \frac{1}{m}\sum_{j=1}^m \cos j\omega\right] \tag{B3}$$

and

$$f_2(z) = \frac{1}{\pi}\int_{\pi/m}^{\pi} d\omega \ln\left[z - \frac{1}{m}\sum_{j=1}^m \cos j\omega\right]. \tag{B4}$$

Consider $f_1(z)$ first. In this interval we will use the bounds (A10) to obtain

$$\frac{1}{\pi}\int_0^{\pi/m} d\omega \ln[z - 1 + \alpha\omega^2 - \beta\omega^4] \le f_1(z) \le \frac{1}{\pi}\int_0^{\pi/m} d\omega \ln[z - 1 + \alpha\omega^2]. \tag{B5}$$

After integrating, (B5) becomes

$$\begin{aligned}&\frac{1}{m}\ln\beta + \frac{1}{\pi}\Bigg[\omega \ln\left(\frac{z-1}{\beta} + \frac{\alpha}{\beta}\omega^2 - \omega^4\right) \\ &+ \left[\left(\frac{\alpha^2}{4\beta^2} + \frac{z-1}{\beta}\right)^{1/2} - \frac{\alpha}{2\beta}\right]^{1/2} \times \tan^{-1}\left\{\omega\Big/\left[\left(\frac{\alpha^2}{4\beta^2} + \frac{z-1}{\beta}\right)^{1/2} - \frac{\alpha}{2\beta}\right]^{1/2}\right\} \\ &+ \left[\frac{\alpha}{2\beta} + \left(\frac{\alpha^2}{4\beta^2} + \frac{z-1}{\beta}\right)^{1/2}\right]^{1/2} \\ &\times \ln\left(\left\{\left[\frac{\alpha}{2\beta} + \left(\frac{\alpha^2}{4\beta^2} + \frac{z-1}{\beta}\right)^{1/2}\right]^{1/2} + \omega\right\}\Big/ \left\{\left[\frac{\alpha}{2\beta} + \left(\frac{\alpha^2}{4\beta^2} + \frac{z-1}{\beta}\right)^{1/2}\right]^{1/2} - \omega\right\}\right) - 4\omega\Bigg]\Bigg|_0^{\pi/m} \\ &\le f_1(z) \le \frac{1}{m}\ln\alpha + \frac{1}{\pi}\left\{\omega \ln\left(\omega^2 + \frac{z-1}{\alpha}\right) + 2\left(\frac{z-1}{\alpha}\right)^{1/2}\tan^{-1}\left[\omega\left(\frac{\alpha}{z-1}\right)^{1/2}\right] - 2\omega\right\}\Bigg|_0^{\pi/m}\end{aligned}$$

and after substituting in the limits of integration this simplifies to

$$\begin{aligned}&\frac{1}{m}\Bigg[\ln\left(z - 1 + \alpha\frac{\pi^2}{m^2} - \beta\frac{\pi^4}{m^4}\right) \\ &+ \frac{2m}{\pi}\left(\frac{\alpha}{2\beta}\right)^{1/2}\left[\left(1 + \frac{4\beta(z-1)}{\alpha^2}\right)^{1/2} - 1\right]^{1/2} \times \tan^{-1}\left\{\pi\left(\frac{2\beta}{\alpha}\right)^{1/2}\Big/ m\left[\left(1 + \frac{4\beta(z-1)}{\alpha^2}\right)^{1/2} - 1\right]^{1/2}\right\} \\ &+ \frac{m}{\pi}\left(\frac{\alpha}{2\beta}\right)^{1/2}\left[1 + \left(1 + \frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\right]^{1/2} \\ &\times \ln\left(\left\{\left(\frac{\alpha}{2\beta}\right)^{1/2}\left[1 + \left(1 + \frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\right]^{1/2} + \frac{\pi}{m}\right\}\Big/ \left\{\left(\frac{\alpha}{2\beta}\right)^{1/2}\left[1 + \left(1 + \frac{4\beta(z-1)}{\alpha^2}\right)^{1/2}\right]^{1/2} - \frac{\pi}{m}\right\}\right) - 4\Bigg] \\ &\le f_1(z) \le \frac{1}{m}\left\{\ln\left(z - 1 + \alpha\frac{\pi^2}{m^2}\right) + \frac{2m}{\pi}\left(\frac{z-1}{\alpha}\right)^{1/2} \times \tan^{-1}\left[\frac{\pi}{m}\left(\frac{\alpha}{z-1}\right)^{1/2}\right] - 2\right\}.\end{aligned} \tag{B6}$$

Both the upper and lower bounds of $f(z)$ in (A11) contain a singularity at $z = 1$. This singularity occurs in the argument of the arctangent function on both sides though and since $\tan^{-1}x \to \pi/2$ as $x \to \infty$, this singularity poses no problem. Remembering that α is of order m^2 and β is of order m^4, we see from (B6) that $f_1(z)$ is of the form

$$f_1(z) = \frac{1}{m}\zeta_1(z), \quad z \ge 1, \tag{B7}$$

where $\zeta_1(z)$ is a function of z and m that can be bounded independent of m.

Next consider $f_2(z)$. By using (A16) in (B4) we have the following equation:

$$f_2(z) = \frac{1}{\pi}\int_{\pi/m}^{\pi} d\omega \ln\left(z + \frac{1}{2m} - \frac{\sin(m+\frac{1}{2})\omega}{2m\sin(\omega/2)}\right)$$
$$= \left(1 - \frac{1}{m}\right)\ln\left[z + \frac{1}{2m}\right]$$
$$+ \frac{1}{\pi}\int_{\pi/m}^{\pi} d\omega \ln\left(1 - \frac{\sin(m+\frac{1}{2})\omega}{(2mz+1)\sin(\omega/2)}\right). \quad \text{(B8)}$$

By the use of (A17), the last integral can be estimated. We shall define this last integral as $K(z)$. Then

$$|K(z)| < \frac{1}{\pi}\left|\int_{\pi/m}^{\pi} d\omega \ln\left\{1 - \left[1\Big/(2mz+1)\left(\frac{2}{\pi}\right)\left(\frac{\omega}{2}\right)\right]\right\}\right|$$
$$= \frac{2}{\pi}\left|\left[\left(\frac{\omega}{2} - \frac{\pi}{2(2mz+1)}\right)\ln\left(\frac{\omega}{2} - \frac{\pi}{2(2mz+1)}\right)\right.\right.$$
$$\left.\left.- \left(\frac{\omega}{2} - \frac{\pi}{(2mz+1)}\right) - \frac{\omega}{2}\ln\frac{\omega}{2} + \frac{\omega}{2}\right]\right|_0^{\pi/m}$$
$$= \left|\left(1 - \frac{1}{2mz+1}\right)\ln\left(1 - \frac{1}{2mz+1}\right)\right.$$
$$- \frac{1}{m}\left(1 - \frac{m}{2mz+1}\right)\ln\left(1 - \frac{m}{2mz+1}\right)$$
$$\left.- \frac{1}{2mz+1}\ln m\right|. \quad \text{(B9)}$$

The term with the largest magnitude in this expression is the last one. But

$$\frac{1}{2mz+1}\ln m < \frac{1}{m}\ln m$$

since $z \geqslant 1$ and this approaches zero as $m \to \infty$ almost as fast as $1/m$. Equation (B9) can clearly be bounded independent of m and (B8) can be written in the form

$$f_2(z) = \ln z + (\ln m/m)\zeta_2(z), \qquad z \geqslant 1 \quad \text{(B10)}$$

in which $\zeta_2(z)$ is a function of z and m that can be bounded independent of m.

Equations (B7) and (B10) combine to give

$$f(z) = \ln z + (\ln m/m)\zeta(z). \quad \text{(B11)}$$

$\zeta(z)$ is a function of z and m and is bounded independent of m. Equations (B2) and (B11) together show that $(\ln m/m)\zeta(z) \to 0$ at least as fast as $1/z$ as $z \to \infty$.

APPENDIX C

Proof that

$$J \equiv \frac{1}{2}\int_{\pi/m}^{\pi} d\omega\left[\sin(m+\tfrac{1}{2})\omega\right]\left[\sin\frac{\omega}{2} - \frac{\sin(m+\frac{1}{2})\omega}{2mz+1}\right]^{-1} \quad \text{(C1)}$$

is bounded. This is the definition of J. This function arises from Eq. (A18).

We are concerned about the magnitude of the integrand since at the lower limit, $\omega = \pi/m$, the denominator take the value

$$\sin\frac{\pi}{m} - \frac{\sin(m+\frac{1}{2})(\pi/m)}{2mz+1} = \sin\frac{\pi}{2m} + \frac{\sin(\pi/2m)}{2mz+1}$$
$$= \frac{\pi}{2m}\left[1 + \frac{1}{2mz+1}\right].$$

We want to assure ourselves that the integral does not make any contribution of order m.

First make the change of variable $\omega = 2x$. Then J becomes,

$$J = \int_{\pi/2m}^{\pi/2} dx\left[\sin(2m+1)x\right]\left[\sin x - \frac{\sin(2m+1)x}{2mz+1}\right] \quad \text{(C2)}$$

Define the positive term sequence A_q by

$$A_q = \left|\int_{q\pi/(2m+1)}^{(q+1)\pi/(2m+1)} dx\left[\sin(2m+1)x\right]\left[\sin x - \frac{\sin(2m+1)x}{2mz+1}\right]^{-1}\right|,$$
$$q = 1, 2, \ldots, m. \quad \text{(C3)}$$

Then (C1) is equal to

$$J = \sum_{q=1}^{m}(-1)^q A_q + C_1 + C_m, \quad \text{(C4)}$$

where

$$C_1 \equiv -\int_{\pi/(2m+1)}^{\pi/2m} dx \times\left[\sin(2m+1)x\right]\left[\sin x - \frac{\sin(2m+1)x}{2mz+1}\right]^{-1},$$

$$C_m \equiv -\int_{\pi/2}^{[(m+1)\pi/(2m+1)]} dx \times\left[\sin(2m+1)x\right]\left[\sin x - \frac{\sin(2m+1)x}{2mz+1}\right]^{-1}.$$

By letting $x = [\pi/(2m+1)] + \epsilon$ we can easily estimate C_1. For C_m, the change of variable $\pi/2 + \epsilon$ is useful. We obtain

$$C_1 \cong \frac{\pi}{(2m)^2}(1 - \ln z), \qquad C_m \cong -\frac{(-1)^m}{2m+1}. \quad \text{(C5)}$$

From (C4) and (C5) we see that the convergence of the integral in (C1) depends on the convergence of the series

$$S_m = \sum_{q=1}^{m}(-1)^q A_q. \quad \text{(C6)}$$

The first terms of the series are the largest since the denominator in the integral is the largest. Consequently we will have to show that their sum is bounded. What is meant by first terms will be made precise.

From (C3) we can write

$$A_q = \left|\int_{q\pi/(2m+1)}^{[(q+1/3)/(2m+1)]\pi} dx\left[\left[\sin(2m+1)x\right]\left(\sin x - \frac{\sin(2m+1)x}{2mz+1}\right)^{-1}\right]\right| + \left|\int_{[(q+1/3)/(2m+1)]\pi}^{[(q+2/3)/(2m+1)]\pi} dx\right.$$
$$\left.\times\left[\left[\sin(2m+1)x\right]\left(\sin x - \frac{\sin(2m+1)x}{2mz+1}\right)^{-1}\right]\right| + \left|\int_{[(q+2/3)/(2m+1)]\pi}^{[(q+1)/(2m+1)]\pi} dx\left[\left[\sin(2m+1)x\right]\left(\sin x - \frac{\sin(2m+1)x}{2mz+1}\right)^{-1}\right]\right|, \quad \text{(C7)}$$

$$
\begin{aligned}
A_q < & \left| \int_{q\pi/(2m+1)}^{[(q+1/3)/(2m+1)]\pi} dx \left[\left(\sin(2m+1)x \right) \left(\sin\frac{q\pi}{2m+1} - \frac{1}{2mz+1} \right)^{-1} \right] \right| \\
& + \left| \int_{[(q+1/3)/(2m+1)]\pi}^{[(q+2/3)/(2m+1)]\pi} dx \left[\left(\sin(2m+1)x \right) \left(\sin\frac{q+\frac{1}{3}}{2m+1}\pi - \frac{1}{2mz+1} \right)^{-1} \right] \right| \\
& + \left| \int_{[(q+2/3)/(2m+1)]\pi}^{[(q+1)/(2m+1)]\pi} dx \left[\left(\sin(2m+1)x \right) \left(\sin\frac{q+\frac{2}{3}}{2m+1}\pi - \frac{1}{2mz+1} \right)^{-1} \right] \right| \\
\leqslant & \frac{1}{2m+1} \left\{ \left[\frac{1}{2} \left(\sin\frac{q\pi}{2m+1} - \frac{1}{2mz+1} \right)^{-1} \right] + \left[\left(\sin\frac{q+\frac{1}{3}}{2m+1}\pi - \frac{1}{2mz+1} \right)^{-1} \right] + \left[\frac{1}{2} \left(\sin\frac{q+\frac{2}{3}}{2m+1}\pi - \frac{1}{2mz+1} \right)^{-1} \right] \right\} \\
= & \frac{1}{2} \left[\left((2m+1) \sin\frac{q\pi}{2m+1} - \beta \right)^{-1} + 2 \left((2m+1) \sin\frac{q+\frac{1}{3}}{2m+1}\pi - \beta \right)^{-1} + \left((2m+1) \sin\frac{q+\frac{2}{3}}{2m+1}\pi - \beta \right)^{-1} \right]
\end{aligned}
\tag{C8}
$$

where $\beta = (2m+1)/(2mz+1)$. β is less than or equal to 1 for all real $z \geq 1$.

If we use

$$
\begin{aligned}
\sin\frac{q+\frac{1}{3}}{2m+1}\pi & = \sin\frac{q\pi}{2m+1} \cos\frac{\frac{1}{3}\pi}{2m+1} + \cos\frac{q\pi}{2m+1} \sin\frac{\frac{1}{3}\pi}{2m+1} \\
& \cong \sin\frac{q\pi}{2m+1} + \frac{\frac{1}{3}\pi}{2m+1} \cos\frac{q\pi}{2m+1}, \\
\sin\frac{q+\frac{2}{3}}{2m+1}\pi & \cong \sin\frac{q\pi}{2m+1} + \frac{\frac{2}{3}\pi}{2m+1} \cos\frac{q\pi}{2m+1}
\end{aligned}
$$

in (C8), we have

$$
\begin{aligned}
& \frac{1}{2} \left\{ \left((2m+1) \sin\frac{q\pi}{2m+1} - \beta \right)^{-1} \right. \\
& \quad + 2 \left[(2m+1) \sin\frac{q\pi}{2m+1} - \left(\beta - \frac{\pi}{3} \cos\frac{q\pi}{2m+1} \right) \right]^{-1} \\
& \quad \left. + \left[(2m+1) \sin\frac{q\pi}{2m+1} - \left(\beta - \frac{2\pi}{3} \cos\frac{q\pi}{2m+1} \right) \right]^{-1} \right\} \\
& > A_q .
\end{aligned}
\tag{C9}
$$

We will call the left-hand side of this expression U_q.

The error introduced into the denominators in (C9) by the approximations

$$
\sin\frac{\frac{1}{3}\pi}{2m+1} \cong \frac{\frac{1}{3}\pi}{2m+1}, \quad \sin\frac{\frac{2}{3}\pi}{2m+1} \cong \frac{\frac{2}{3}\pi}{2m+1}, \quad \cos\frac{\frac{1}{3}\pi}{2m+1} \cong 1, \quad \cos\frac{\frac{2}{3}\pi}{2m+1} \cong 1,
$$

is of order $(1/m)^3$ in the first case and of order $(1/m)^2$ in the second case.

From Eq. (C7) we can proceed in the opposite direction as (C8) and obtain

$$
\begin{aligned}
& \left| \int_{q\pi/(2m+1)}^{[(q+1/3)/(2m+1)]\pi} dx \left[[\sin(2m+1)x] \left(\sin\frac{q+\frac{1}{3}}{2m+1}\pi + \frac{1}{2mz+1} \right)^{-1} \right] \right| + \left| \int_{[(q+1/3)/(2m+1)]\pi}^{[(q+2/3)/(2m+1)]\pi} dx \right. \\
& \times \left. \left[[\sin(2m+1)x] \left(\sin\frac{q+\frac{2}{3}}{2m+1}\pi + \frac{1}{2mz+1} \right)^{-1} \right] \right| + \left| \int_{[(q+2/3)/(2m+1)]\pi}^{[(q+1)/(2m+1)]\pi} dx \left[[\sin(2m+1)x] \left(\sin\frac{q+1}{2m+1}\pi + \frac{1}{2mz+1} \right)^{-1} \right] \right| .
\end{aligned}
$$

We can perform a series of steps analogous to those leading to (C9) to obtain

$$
\begin{aligned}
A_q > & \frac{1}{2} \left\{ \left[(2m+1) \sin\frac{q+1}{2m+1}\pi + \left(\beta - \tfrac{2}{3}\pi \cos\frac{q+1}{2m+1}\pi \right) \right]^{-1} \right. \\
& + 2 \left[(2m+1) \sin\frac{q+1}{2m+1}\pi + \left(\beta - \tfrac{1}{3}\pi \cos\frac{q+1}{2m+1}\pi \right) \right]^{-1} \\
& \left. + \left((2m+1) \sin\frac{q+1}{2m+1}\pi + \beta \right)^{-1} \right\} = L_q
\end{aligned}
\tag{C10}
$$

Equations (C9) and (C10) give upper and lower bounds for A_q. We wish to find for which values of q, if any, the upper bound on the $(q+1)$th term is smaller than the qth term. The series formed by a sum over this set of q's then forms an alternating series with the magnitude of each term decreasing. In symbols, we wish to find a range of q's satisfying

$$U_{q+1} < L_q$$

or (by introducing $\psi = \dfrac{\pi}{2m+1}$

$$
\begin{aligned}
& \frac{1}{2} \left(\frac{1}{(2m+1)\sin(q+1)\psi - \beta} \right. \\
& \quad + \frac{2}{(2m+1)\sin(q+1)\psi - [\beta - \frac{1}{3}\pi \cos(q+1)\psi]} \\
& \quad \left. + \frac{1}{(2m+1)\sin(q+1)\psi - (\beta - \frac{2}{3}\pi \cos(q+1)\psi} \right) \\
< & \frac{1}{2} \left(\frac{1}{(2m+1)\sin(q+1)\psi + [\beta - \frac{2}{3}\pi \cos(q+1)\psi} \right. \\
& \quad + \frac{2}{(2m+1)\sin(q+1)\psi + [\beta - \frac{1}{3}\pi \cos(q+1)\psi]} \\
& \quad \left. + \frac{1}{(2m+1)\sin(q+1)\psi + \beta} \right)
\end{aligned}
$$

or upon rearranging and collecting terms

$$\left(\frac{1}{(2m+1)\sin(q+1)\psi-\beta}-\frac{1}{(2m+1)\sin(q+1)\psi+\beta}\right)$$
$$<\left(\frac{1}{(2m+1)\sin(q+1)\psi+[\beta-\frac{2}{3}\pi\cos(q+1)\pi]}-\frac{1}{(2m+1)\sin(q+1)\psi-[\beta-\frac{2}{3}\pi\cos(q+1)\psi]}\right)$$
$$+2\left(\frac{1}{(2m+1)\sin(q+1)\psi+(\beta-\frac{1}{3}\pi\cos(q+1)\psi)}-\frac{1}{(2m+1)\sin(q+1)\psi-[\beta-\frac{1}{3}\pi\cos(q+1)\psi]}\right).$$

A simple calculation then gives

$$\frac{2\beta}{(2m+1)^2\sin^2(q+1)\psi-\beta^2}$$
$$<\frac{2[\frac{2}{3}\pi\cos(q+1)\psi-\beta]}{(2m+1)^2\sin^2(q+1)\psi-[\beta-\frac{2}{3}\pi\cos(q+1)\psi]^2}+2\frac{2[\frac{1}{3}\pi\cos(q+1)\psi-\beta]}{(2m+1)^2\sin^2(q+1)\psi-[\beta-\frac{1}{3}\pi\cos(q+1)\psi]^2}.$$

This equation will be satisfied if the following two conditions are simultaneously satisfied:

$$\frac{2\beta}{(2m+1)^2\sin^2(q+1)\psi-\beta^2}<\frac{2[\frac{2}{3}\pi\cos(q+1)\psi-\beta]}{(2m+1)^2\sin^2(q+1)\psi-[\frac{2}{3}\pi\cos(q+1)\psi-\beta]^2}$$

and

$$2\frac{2[\frac{1}{3}\pi\cos(q+1)\psi-\beta]}{(2m+1)^2\sin^2(q+1)\psi-[\frac{1}{3}\pi\cos(q+1)\psi-\beta]^2}>0.$$

The first is satisfied if $2\beta<2[\frac{2}{3}\pi\cos(q+1)\psi-\beta]$ or $\beta<\frac{1}{3}\pi\cos(q+1)\psi$.

The second is satisfied by the same condition. Since $\beta\le 1$, then if $\frac{1}{3}\pi\cos(q+1)\psi>1$, the condition will always be satisfied irregardless of the value of z.

The condition

$$\cos[(q+1)/(2m+1)]\pi>\pi/3$$

implies that the argument of the cosine will be less than about .30.

$$[(q+1)/(2m+1)]\pi\lesssim .30$$

or

$$q\lesssim .19m.$$

This is a significant portion of the terms of the series. From previous remarks we see that

$$-A_1<\sum_{q=1}^{.19m}<-A_1+A_2<0.$$

From (C9) we have that

$$-\frac{1}{2}\left\{\left((2m+1)\sin\frac{\pi}{2m+1}-\beta\right)^{-1}+2\left[(2m+1)\sin\frac{\pi}{2m+1}-\left(\beta-\frac{\pi}{3}\cos\frac{\pi}{2m+1}\right)\right]^{-1}+\left[(2m+1)\sin\frac{\pi}{2m+1}-\left(\beta-\frac{2\pi}{3}\cos\frac{\pi}{2m+1}\right)\right]^{-1}\right\}$$
$$\cong-\frac{1}{2}\left\{\frac{1}{\pi-\beta}+\frac{2}{(4\pi/3)-\beta}+\frac{1}{(5\pi/3)-\beta}\right\}<-A_1.$$

This is bounded irrespective of m even when β attains its maximum value 1.

The rest of the terms in the sum (C6) are represented by the integral from about $x=.30$ on up to $x=\pi/2$. It is easy to show that this integral is bounded for all m since

$$\left|\int_{.3}^{\pi/2}dx\left[[\sin(2m+1)x]\left(\sin x-\frac{\sin(2m+1)x}{2mz+1}\right)^{-1}\right]\right|$$
$$<\left(\sin(.3)-\frac{1}{2mz+1}\right)^{-1}\int_{.3}^{\pi/2}dx$$
$$\times\left[\frac{\pi}{2}\left(\sin(.3)-\frac{1}{2mz+1}\right)^{-1}\right]$$
$$\le\left[\frac{\pi}{2}\left(\sin(.3)-\frac{1}{2m+1}\right)^{-1}\right].$$

Therefore, the integral (C1) has a magnitude on the order of $m^0=1$ and can be bounded independent of m.

[1]R. A. Ferrel, Bull. Am. Phys. Soc. **2**, 315 (1966).
[2]W. A. Little, Proc. R. Soc. A **134**, 416 (1964).
[3]L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958), p. 482.
[4]L. Van Hove, Physica (Utr.) **16**, 137 (1950).
[5]M. Kac, Phys. Fluids **2**, 8 (1959).
[6]M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216 (1963); J. Math. Phys. **4**, 229 (1963); J. Math. Phys. **5**, 60 (1964).
[7]Joseph L. Strecker, J. Math. Phys. **10**, 1541 (1969).
[8]H. A. Gersch, Phys. Fluids **6**, 599 (1963).
[9]D. J. Thouless, Phys. Rev. **187**, 732 (1969).
[10]F. J. Dyson, "Phase Transitions in Ferromagnets," in *Statistical Mechanics at the Turn of the Decade*, edited by E. G. D. Cohen (Dekker, New York, 1971).
[11]K. H. Carpenter and J. L. Strecker, J. Math. Phys. **11**, 2063 (1970).
[12]P. W. Anderson and G. Yuval, Phys. Rev. Lett. **23**, 89 (1969).
[13]T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).

Erratum: Two-magnon bound states in Heisenberg ferromagnets [J. Math. Phys. 14, 1837 (1973)]

A. W. Saénz and W. W. Zachary

Naval Research Laboratory, Washington, D.C. 20375
(Received 24 May 1974)

We wish to correct a number of typographical errors which appeared in the above paper.

(1) In the ninth line of the second paragraph of the first column on p. 1837, Silbergliti should read Silberglitt.

(2) A negative sign should be inserted in front of the first term on the right side of Eq. (2.1).

(3) In the third line following the definition of the set P_N in the second column on p. 1838, the expression $\lambda + 2\pi\lambda(\gamma)$ should read $\gamma + 2\pi\lambda(\gamma)$.

(4) In the fourth line following Eq. (2.8) the comma should be deleted from the expression $Z_N, (\Gamma)$.

(5) In Eq. (2.14) the argument of the first cosine factor should read $\frac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_l$.

(6) In the fourth line following Eq. (2.16) the negative sign should be replaced by an equality sign.

(7) In the first term on the right side of Eq. (2.21) ξ' should have a zero subscript.

(8) The second line of Eq. (2.22) should read

$$\xi_l' = -\sqrt{2}\ \eta \cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_l)\,\xi_0 + \xi_l.$$

(9) In the seventh line of the statement of Property (1) in the second column of p. 1840, the symbol $E_{\max}(\Gamma_0)$ should read $E_{\max}^{(N)}(\Gamma_0)$.

(10) In Eq. (2.26) the symbol $\mathscr{H}$ should be replaced by $\mathscr{K}$.

(11) In Eq. (2.30) the upper limit of the summation in the denominator should read n instead of u.

(12) In Eq. (2.34) the lower limit of the summation in the denominator should read l instead of i.

(13) In the second line of the last paragraph in the second column on p. 1842, the first "of" should read "or".

(14) The equation appearing in the fourth line following Eq. (2.40) should read

$$K(E,\Gamma)\,\psi_\mu(E,\Gamma) = k_\mu(E,\Gamma)\,\psi_\mu(E,\Gamma).$$

(15) In the statement of Theorem 2.1 on p. 1843, the reference to Eq. (2.25) should refer instead to Eq. (2.15).

(16) The inequality following the reference to Jensen's inequality on p. 1844 should read

$$[\mathrm{Tr}K^q(E_0,\Gamma_0)]^{p/q} = \left(\sum_{\mu=0}^{n} k_\mu{}^q(E_0,\Gamma_0)\right)^{p/q} \geqslant \sum_{\mu=0}^{n} |k_\mu{}^p(E_0,\Gamma_0)|.$$

(17) The exponents p/q occurring in the proof of Theorem 3.5 should be $p-q$.

(18) The summation in Eq. (3.10) should have n as an upper limit.

(19) In the last line of the second column on p. 1847 the inequality should read

$$\mathrm{Tr}K^3(\Gamma) > \mathrm{Tr}K^2(\Gamma).$$

 Copyright © 1974 American Institute of Physics