JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Identity Satisfied by the Racah Coefficients of $U(n)$*

J. D. LOUCK
University of California, Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87544
AND
L. C. BIEDENHARN
Department of Physics, Duke University, Durham, North Carolina 27706

(Received 24 July 1970)

The origin of a well-known relation satisfied by the Racah coefficients of $SU(2)$ is demonstrated to be the associative law of multiplication of the Wigner operators of $SU(2)$. Recognition of this simple fact allows the immediate generalization of this identity to the Racah coefficients of $U(n)$.

I. INTRODUCTION

During a recent investigation[1] of the properties of the multiplication law for symplecton eigenpolynomials,[2] it was observed that a fundamental relation between the triangle coefficients [defined in Eq. (3) below] and the Racah coefficients obtained in consequence of the *associative law of multiplication* of these operators. The symplecton eigenpolynomials are a particular example of a set of irreducible tensor operators

$$\{T^a_\alpha: \alpha = a, a-1, \cdots, -a, a = 0, \tfrac{1}{2}, \cdots\}, \quad (1)$$

with respect to $SU(2)$, which also comprise an operator ring with the multiplication

$$T^a_\alpha T^b_\beta = \sum_c (2c+1)^{-\frac{1}{2}} F(bac) C^{bac}_{\beta\alpha} T^c_{\alpha+\beta}, \quad (2)$$

where $C^{bac}_{\beta\alpha} \equiv C^{bac}_{\beta,\alpha,\alpha+\beta}$ denotes a Wigner coefficient of $SU(2)$ and $F(bac)$ is, in general, an invariant operator with respect to $SU(2)$. Equation (2) is equivalent to the statement that the coupling of two tensor operators from the set (1) is an invariant times an irreducible tensor operator belonging to the set.

The symplecton[2,3] is the most elementary object from which one can construct operator-valued polynomials which satisfy the multiplication rule (2). In this case, the invariant operators $F(bac)$ are explicit *numerical* coefficients, denoted in Ref. 1 as the "triangle coefficients":

$$\Delta(bac) \equiv \left[\frac{(a+b+c+1)!}{(a+b-c)!(a-b+c)!(-a+b+c)!}\right]^{\frac{1}{2}}. \quad (3)$$

Conversely, if the $F(bac)$ are the triangle coefficients, then the multiplication law (2) uniquely defines the algebra of the symplecton.

The associative law of multiplication of symplecton eigenpolynomials,

$$(T^a_\alpha T^b_\beta) T^c_\gamma = T^a_\alpha (T^b_\beta T^c_\gamma), \quad (4)$$

now leads directly to the following fundamental relation between triangle coefficients and Racah coefficients[4,5] [the calculation uses Eq. (18) of Ref. 5]:

$$\Delta(afc)\Delta(bdf) = (2f+1)\sum_e \Delta(abe)\Delta(edc)W(abcd; ef). \quad (5)$$

[The labels a, b, c of Eq. (4) have been renamed in Eq. (5).]

In quantum mechanics, operators are usually represented by the elements of matrices, and the associative law of multiplication always holds in consequence of this property for matrix multiplication. One is thus led into the trap of ignoring the direct consequences of the associativity requirement. The purpose of this paper is to demonstrate that several important relations in unitary group theory owe their origin to the fundamental associativity law.

II. THE RELATION BETWEEN RACAH COEFFICIENTS OF $SU(2)$

There immediately comes to mind a second set of tensor operators which satisfies a multiplication law of the form (2). These are the irreducible tensor operators $Y^L_M(\mathbf{J})$, constructed from the components J_k, $k = 1, 2, 3$, of the generators of an irreducible representation of $SU(2)$ [see Eq. (5.67) of Ref. 6]. However, application of the associativity law to these operators leads to a special case of Eq. (11) derived below.

There is still another set of $SU(2)$ tensor operators which possesses the algebraic property expressed by Eq. (2)—the Wigner operators of $SU(2)$. The notion of a Wigner operator has proved to be extremely valuable in the study of the unitary groups.[7-9] While a Wigner operator of $SU(2)$ can be completely characterized by specifying its algebraic properties without any initial reference to Wigner coefficients, it is not particularly expedient, in the case of $SU(2)$, to follow this course. The properties of $SU(2)$ Wigner

operators can be made more concrete and understandable by giving the initial definition directly in terms of the known $SU(2)$ Wigner coefficients.

A Wigner operator of $SU(2)$ is designated[10] by the notation

$$\begin{pmatrix} a+\rho \\ 2a \qquad 0 \\ a+\alpha \end{pmatrix}, \tag{6}$$

and is, first of all, an irreducible tensor operator of type T^a_α, $\alpha = a, a-1, \cdots, -a$, with respect to its transformation properties. Second, it is an operator which effects the shift ρ on the label j of a generic state vector $|jm\rangle$: More precisely, it is defined by

$$\begin{pmatrix} a+\rho \\ 2a \qquad 0 \\ a+\alpha \end{pmatrix} |jm\rangle = C^{ja\,j+\rho}_{m\alpha}\,|j+\rho, m+\alpha\rangle. \tag{7}$$

In particular, ρ may be any of the values $\rho = a, a-1, \cdots, -a$, so that there are $2a+1$ tensor operators of type T^a_α defined by Eq. (7), i.e., one for each value of ρ. Note that because the Wigner coefficients are defined to be zero if the triangle conditions on j, a, and $j+\rho$ are violated, certain irreducible representation (IR) spaces $\{|jm\rangle\}$ will be annihilated by a given Wigner operator, i.e., *each Wigner operator has an associated null space.*

It is not our purpose here to develop the elegant theory of $SU(2)$ which can be based on Wigner operators, since the essential details have been given previously.[8,11] We will, however, note that the algebraic structure of Wigner operators is *isomorphic* to the set of irreducible operator polynomials which can be built on a *pair* of commuting symplectons, $(a, \bar{a})$ and $(b, \bar{b})$. Thus, Wigner operators present a structure which is richer in details and, at the same time, more complex than that of the single symplecton.

The reader will now be able to verify for himself that the following multiplication law is but an *operator statement* of a well-known relation[5] between Wigner coefficients and Racah coefficients:

$$\begin{pmatrix} a+\rho \\ 2a \qquad 0 \\ a+\alpha \end{pmatrix} \begin{pmatrix} b+\sigma \\ 2b \qquad 0 \\ b+\beta \end{pmatrix} = \sum_c W^{bac}_{\sigma\rho} C^{bac}_{\beta\alpha} \begin{pmatrix} c+\rho+\sigma \\ 2c \qquad 0 \\ c+\alpha+\beta \end{pmatrix}, \tag{8}$$

where $W^{bac}_{\sigma\rho} \equiv W^{bac}_{\sigma\rho\rho+\sigma}$ is a notation for an invariant (Racah) operator of $SU(2)$ which has a Racah coefficient for its eigenvalue on an arbitrary IR space specified by j:

$$W^{bac}_{\sigma\rho}(j) = [(2c+1)(2j-2\rho+1)]^{\frac{1}{2}} \times W(j-\rho-\sigma, b, j, a; j-\rho, c). \tag{9}$$

Note that, in applying Eq. (8) to an arbitrary state $|jm\rangle$, the Racah operator gets evaluated on the shifted label $j+\rho+\sigma$ (the final state vector label).

The notation for the Racah invariants is derived from the remarkable property[12]

$$\lim_{j\to\infty} W^{bac}_{\sigma\rho}(j) = C^{bac}_{\sigma\rho}. \tag{10}$$

We now observe that the product law (8) is precisely of the form (2), the extra labels ρ, σ arising because of the extra labels specifying the particular (shift) properties of the tensor operators.

We can now ask: What are the implications of the associativity law for Wigner operators? The result must have the same general appearance as Eq. (5), with some additional operator labels addended to the quantities. Furthermore, we must be careful to account for the fact that the W's appearing in Eq. (8) are invariant operators, and it makes a difference whether they appear to the left or right of a Wigner operator (but simply by a shift in the state on which the invariant is evaluated).

This straightforward calculation is easily carried out with the following result:

$$W^{afe}_{\alpha\beta+\epsilon}(j) W^{bdf}_{\beta\epsilon}(j) = \sum_e W^{abe}_{\alpha\beta}(j-\epsilon) W^{edc}_{\alpha+\beta\epsilon}(j) [(2e+1)(2f+1)]^{\frac{1}{2}} \times W(abcd; ef). \tag{11}$$

Note the agreement in form with Eq. (5) as one reads across the upper labels.

If we now use Eq. (9) to write Eq. (11) in terms of the usual notation for Racah coefficients, then Eq. (11) is just the expression of the well-known relation, Eq. (25) of Ref. 13. This relation was first derived by considering the couplings of four angular momenta. *It is here demonstrated to be the consequence of a very fundamental property of Wigner operators—the associativity law.*

Observe that, in the limit $j \to \infty$, Eq. (11) reduces to one of the standard relations between Wigner coefficients and Racah coefficients. Thus we see that the Wigner coefficients themselves are still another example of numerical coefficients which satisfy a coupling law of the general form (5).

III. GENERALIZATION TO $U(n)$

Having uncovered a fundamental origin of Eq. (11), we may now easily generalize the result to the general

unitary group $U(n)$. The notation for Wigner operators and Racah invariant operators has been explained in considerable detail elsewhere.[8,9] Here we introduce an abbreviated notation for Gel'fand patterns and operator patterns. The symbols

$$\begin{pmatrix} a \\ \alpha \end{pmatrix}, \begin{pmatrix} b \\ \beta \end{pmatrix}, \cdots \tag{12}$$

designate *Gel'fand* patterns—the triangular set of labels of state vectors of a basis for an IR of $U(n)$—where $a, b, \cdots$ denote IR labels (row vectors of n elements) and $\alpha, \beta, \cdots$ denote the subgroup labels of the canonical Weyl branching law. Similarly, the symbols

$$\begin{pmatrix} a \\ \rho \end{pmatrix}, \begin{pmatrix} b \\ \sigma \end{pmatrix}, \cdots \tag{13}$$

designate the operator patterns of $U(n)$ Wigner operators. The Δ pattern of an operator pattern is written simply as

$$\Delta\begin{pmatrix} a \\ \rho \end{pmatrix} = \Delta(\rho), \tag{14}$$

and $b + \Delta(\rho)$ denotes the set of IR labels obtained by row vector addition of the row vector b and the row vector $\Delta(\rho)$.

The product of two $U(n)$ Wigner operators is a linear combination of Wigner operators, this result being exactly expressed as follows:

$$\left\langle \begin{matrix} \rho \\ a \\ \alpha \end{matrix} \right\rangle \left\langle \begin{matrix} \sigma \\ b \\ \beta \end{matrix} \right\rangle = \sum_{\rho'\gamma\tau} \left\langle \begin{pmatrix} b + \Delta(\rho') \\ \gamma \end{pmatrix} \middle| \begin{matrix} \rho' \\ a \\ \alpha \end{matrix} \middle| \begin{pmatrix} b \\ \beta \end{pmatrix} \right\rangle \times \left\{ \begin{pmatrix} b + \Delta(\rho') \\ \tau \end{pmatrix} \begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix} \begin{pmatrix} b \\ \sigma \end{pmatrix} \right\} \times \left\langle \begin{matrix} \tau \\ b + \Delta(\rho') \\ \gamma \end{matrix} \right\rangle. \tag{15}$$

The symbol $\langle \,|\, \langle \,\rangle \,|\, \rangle$ denotes a $U(n)$ Wigner coefficient, and $\{\cdots\}$ denotes a $U(n)$ Racah invariant operator. The eigenvalue of $\{\cdots\}$ on an arbitrary state vector with IR labels m is denoted by

$$\left\{ \begin{pmatrix} b + \Delta(\rho') \\ \tau \end{pmatrix} \begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix} \begin{pmatrix} b \\ \sigma \end{pmatrix} \right\}(m), \tag{16}$$

and this coefficient is a $U(n)$ Racah coefficient.

We next work out the consequences of the associativity law for the product of three Wigner operators, using Eq. (15). The derivation is quite tedious, but straightforward. The result is the following identity satisfied by the Racah coefficients of $U(n)$ (all patterns in this result are operator patterns):

$$\sum_{\delta} \left\{ \begin{pmatrix} d \\ \delta \end{pmatrix} \begin{pmatrix} \rho'' \\ a \\ \rho \end{pmatrix} \begin{pmatrix} b \\ \sigma \end{pmatrix} \right\}(m) \left\{ \begin{pmatrix} e \\ \epsilon \end{pmatrix} \begin{pmatrix} \delta'' \\ d \\ \delta \end{pmatrix} \begin{pmatrix} c \\ \tau \end{pmatrix} \right\}(m)$$

$$= \sum_{\rho'\sigma'\delta'} \left\{ \begin{pmatrix} d \\ \delta'' \end{pmatrix} \begin{pmatrix} \rho'' \\ a \\ \rho' \end{pmatrix} \begin{pmatrix} b \\ \sigma' \end{pmatrix} \right\}(e) \times \left\{ \begin{pmatrix} c + \Delta(\sigma') \\ \delta' \end{pmatrix} \begin{pmatrix} \sigma' \\ b \\ \sigma \end{pmatrix} \begin{pmatrix} c \\ \tau \end{pmatrix} \right\}(m - \Delta(\rho)) \times \left\{ \begin{pmatrix} e \\ \epsilon \end{pmatrix} \begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix} \begin{pmatrix} c + \Delta(\sigma') \\ \delta' \end{pmatrix} \right\}(m). \tag{17}$$

[To verify that Eq. (17) reduces properly to Eq. (11), it is only necessary to identify the two notations for $SU(2)$. That is, $j = \frac{1}{2}(m_1 - m_2)$, $a = \frac{1}{2}(a_1 - a_2)$, $b = \frac{1}{2}(b_1 - b_2)$, $c = \frac{1}{2}(c_1 - c_2)$, $\sigma = \sigma_1 - \frac{1}{2}(b_1 + b_2)$, $\rho = \rho_1 - \frac{1}{2}(a_1 + a_2)$, and $c_1 + c_2 = a_1 + a_2 + b_1 + b_2$. It follows then that

$$W^{bac}_{\sigma\rho}(j) = \left\{ \begin{pmatrix} c_1 & & c_2 \\ & \rho_1 + \sigma_1 & \end{pmatrix} \begin{pmatrix} c_1 - b_1 & & \\ a_1 & & a_2 \\ & \rho_1 & \end{pmatrix} \begin{pmatrix} b_1 & & b_2 \\ & \sigma_1 & \end{pmatrix} \right\}(m_1 m_2).] \tag{18}$$

An identity of the general type (17) has been derived by other authors.[14] However, we believe our derivation to be the first one which underscores the fundamental origin of this relation. [Observe also that there is no proliferation of the so-called "multiplicity indices" in Eq. (17), only *operator patterns* which are *structural labels*.]

One can continue still further with the consequences of the associativity law. The notion of a $U(n): U(n-1)$ projective operator has been developed in some detail [the matrix elements of projective operators are $U(n): U(n-1)$ reduced Wigner coefficients].[8,9] In particular, the product of two such operators is given by[9]

(indices $\alpha, \beta, \cdots$ now designate operator patterns)

$$\begin{bmatrix} \rho \\ a \\ \alpha \end{bmatrix}\begin{bmatrix} \sigma \\ b \\ \beta \end{bmatrix} = \sum_{\rho'\gamma\tau}\left[\begin{pmatrix} b+\Delta(\rho') \\ \gamma \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \alpha \end{pmatrix}\begin{pmatrix} b \\ \beta \end{pmatrix}\right] \times \left\{\begin{pmatrix} b+\Delta(\rho') \\ \tau \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} b \\ \sigma \end{pmatrix}\right\}\begin{bmatrix} \tau \\ b+\Delta(\rho') \\ \gamma \end{bmatrix}, \tag{19}$$

where $[\cdots]$ denotes the square-bracket invariant functions of the $U(n-1)$ IR labels introduced in Ref. 9. Once again one can determine the relation which is implied by the associativity law of multiplication of three $U(n):U(n-1)$ projective operators. The result has the same structural form as Eq. (17) [Eq. (17) is required in the derivation]:

$$\sum_{\delta}\left[\begin{pmatrix} d \\ \delta \end{pmatrix}\begin{pmatrix} \rho'' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} b \\ \sigma \end{pmatrix}\right](m')\left[\begin{pmatrix} e \\ \epsilon \end{pmatrix}\begin{pmatrix} \delta'' \\ d \\ \delta \end{pmatrix}\begin{pmatrix} c \\ \tau \end{pmatrix}\right](m')$$

$$= \sum_{\rho'\sigma'\delta'}\left\{\begin{pmatrix} d \\ \delta'' \end{pmatrix}\begin{pmatrix} \rho'' \\ a \\ \rho' \end{pmatrix}\begin{pmatrix} b \\ \sigma' \end{pmatrix}\right\}(e) \times \left[\begin{pmatrix} c+\Delta(\sigma') \\ \delta' \end{pmatrix}\begin{pmatrix} \sigma' \\ b \\ \sigma \end{pmatrix}\begin{pmatrix} c \\ \tau \end{pmatrix}\right](m'-\Delta'(\rho)) \times \left[\begin{pmatrix} e \\ \epsilon \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} c+\Delta(\sigma') \\ \delta' \end{pmatrix}\right](m'), \tag{20}$$

in which m' designates a set of $U(n-1)$ IR labels (a row vector of $n-1$ elements) and $\Delta'(\rho)$ designates the $(n-1)$-element Δ pattern obtained by deleting the last entry from $\Delta(\rho)$. The notation $[\cdots](m')$ designates an eigenvalue of a square-bracket invariant, and these coefficients relate in a definite manner[9] to $U(n-1)$ Racah coefficients.

It was conjectured in Ref. 9 that there exists a definite limiting relation between Racah coefficients and square-bracket coefficients. If we introduce the more explicit notation

$$(m_n) = (m_{1n}m_{2n}\cdots m_{nn}), \tag{21}$$

$$(m_{n-1}) = (m_{1n}m_{2n}\cdots m_{n-1n}), \tag{22}$$

then this conjectured relation is the property

$$\lim_{m_{nn}\to-\infty}\left\{\begin{pmatrix} c \\ \tau \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} b \\ \sigma \end{pmatrix}\right\}(m_n) = \left[\begin{pmatrix} c \\ \tau \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} b \\ \sigma \end{pmatrix}\right](m_{n-1}). \tag{23}$$

In $SU(2)$, this equation is just the expression of property (10). Equation (23) has also been demonstrated to be valid for a large class of nontrivial cases.[9] Equations (17) and (20) *support*, but do not prove, the general validity of Eq. (23), in that Eq. (20) is an immediate *consequence* of Eq. (17) if property (23) holds.

Indeed, if Eq. (23) is valid generally [for all $U(n)$], then an immediate consequence of it and the explicit expression[9] relating the square-bracket coefficients to $U(n-1)$ Racah coefficients is that, under the successive limits $m_{nn}\to-\infty$, $m_{n-1n}\to-\infty, \cdots, m_{2n}\to-\infty$, a Racah coefficient limits to a Wigner coefficient:

$$\left\{\begin{pmatrix} c \\ \tau \end{pmatrix}\begin{pmatrix} \rho' \\ a \\ \rho \end{pmatrix}\begin{pmatrix} b \\ \sigma \end{pmatrix}\right\}(m_n) \to \left\langle \begin{pmatrix} c \\ \tau \end{pmatrix}\right| \left\langle \begin{matrix} \rho \\ a \\ \rho \end{matrix}\right\rangle \left|\begin{pmatrix} b \\ \sigma \end{pmatrix}\right\rangle. \tag{24}$$

Taking this limit of Eq. (17), we then arrive back at the matrix element expression of the product law (15). Thus, the *compatibility* of the limit properties (23) and (24) with Eqs. (15), (17), and (20) has been demonstrated.

We remark that the identity (17) also assures the validity of the associative law of multiplication for the extended $U(n):U(n-1)$ projective operators introduced in Ref. 9.

A general proof of Eq. (23) would be quite important because from it one can establish the limit properties[9] of $U(n):U(n-1)$ reduced Wigner coefficients which, in turn, can be used to induce uniquely all operator pattern assignments in the canonical resolution of the multiplicity problem.

IV. CONCLUSIONS

One of the more difficult problems encountered in generalizing the Racah–Wigner angular momentum calculus to $U(n)$ is recognizing those features of $SU(2)$ which are generalizable and those which are not, i.e., are particular to $SU(2)$. This generalization is often made quite transparent if a general structural feature underlies the relation, such as the associative law of multiplication which underlies Eq. (11).

Our ultimate goal, of course, is to find those structural features of Wigner operators which uniquely characterize all operator patterns. A significant step in this direction was made in Ref. 9.

The contribution of this paper is not so much in deriving relations which $U(n)$ Racah coefficients must satisfy as in recognizing the simple and general origin of such complicated appearing relations.

* Supported in part by the U.S. Army Research Office (Durham), the National Science Foundation, and the U.S. Atomic Energy Commission.

[1] L. C. Biedenharn and J. D. Louck, to appear in Ann. Phys. (N.Y.) (de Shalit Memorial Volume).

[2] The term "symplecton" was introduced in Ref. 1 to denote the boson pair $(a, \bar{a})$. The conjugation symmetric characteristic polynomials in a and $\bar{a}$ are called "symplecton eigenpolynomials." The name (symplecton) was suggested by the fact that a symplecton having n dimensions (n independent boson pairs) generates a representation space for the group $Sp(2n)$—the symplectic group of rank n.

Since the submission of our paper (Ref. 1) we have found that the word "symplecton" was originally coined by H. Freudenthal, Ref. 3, in his discovery of metasymplectic geometry. Freudenthal used the term symplecton to denote a symplectic geometry; in a metasymplectic geometry the elements are points, lines, planes, and symplecta. Our usage of the term thus accords reasonably with his.

[3] H. Freudenthal, Proc. Ned. Akad. Wetenschap. **A62**, 165 (1959).

[4] G. Racah, Phys. Rev. **62**, 438 (1942).

[5] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Rev. Mod. Phys. **24**, 249 (1952).

[6] J. Schwinger, in *Quantum Theory of Angular Momentum*, L. C. Biedenharn and H. Van Dam, Eds. (Academic, New York, 1965).

[7] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).

[8] J. D. Louck, Am. J. Phys. **38**, 3 (1970).

[9] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).

[10] The notation derives from the general notation for a $U(n)$ Wigner operator (see Refs. 7–9). In particular, note the relation between $U(2)$ Wigner operators and $SU(2)$ Wigner operators given by Eq. (3.54) of Ref. 8.

[11] L. C. Biedenharn, in *Spectroscopic and Group Theoretical Methods in Physics* (Racah Memorial Volume) (North-Holland, Amsterdam, 1968), p. 59.

[12] The authors (and no doubt others) have been aware of this relation for some years, and it was stated in Ref. 9 (in a more general context). Particular cases of it have been noted previously [see Eq. (29) of Ref. 13].

[13] L. C. Biedenharn, J. Math. & Phys. **31**, 287 (1953).

[14] J.-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Boundary Conditions in the Pairwise Point Transformation Method

NORMAN M. WITRIOL
Solid State Physics Branch, Physical Sciences Laboratory, Redstone Arsenal, Alabama 35809

(Received 29 April 1970)

Using boundary conditions, one can eliminate the need of a transformation cutoff function in the pairwise point transformation method applied to the many-body problem.

In previous work[1–3] on the many-body problem using the pairwise point transformation method, it was necessary to introduce a cutoff function into the transformation, such that the transformation approached the identity transformation for interparticle separations greater than an arbitrarily chosen cutoff distance.[1] The purpose of this article is to show that one need not introduce a cutoff function into the transformation, provided that one requires the original wavefunction to approach the free-particle wavefunction within an arbitrarily chosen cutoff distance. If this distance is taken to be less than or equal to the interparticle separation, this requirement is essentially that of "pairwise additivity"; namely, after particle "a" has interacted pairwise with particle "b," it must become a free particle before it can interact pairwise with particle "c." In essence, we show that one can put the cutoff into the wavefunction instead of the transformation. This technique could be useful in simplifying the calculations inherent in applying many-body point transforms[4] to many-body systems.

In the pairwise point transformation method,[2] the system is assumed to be sufficiently dilute such that the Hamiltonian H can be assumed to be the sum of pairwise Hamiltonians H_{ij},

$$H = \sum_{\substack{i,j=1 \\ i<j}}^{N} H_{ij}. \tag{1}$$

We shall therefore only consider the 2-body wavefunction associated with H_{ij}. To illustrate the technique, we shall demonstrate it on the hard-core 2-body problem. This problem was previously studied[1] using a cutoff function in the transformation, and the standard results were obtained.

Our notation for equal-mass particles is that the $X_{i\alpha}$ (the $\mathbf{R}_i$) are the original coordinates, the $x_{i\alpha}$ (the $\mathbf{r}_i$) are the transformed coordinates, the $P_{i\alpha}$ are the original momenta, the $p_{i\alpha}$ are the transformed momenta, the $X_{ij\alpha}$, $x_{ij\alpha}$, $P_{ij\alpha}$, and $p_{ij\alpha}$ represent the

Our ultimate goal, of course, is to find those structural features of Wigner operators which uniquely characterize all operator patterns. A significant step in this direction was made in Ref. 9.

The contribution of this paper is not so much in deriving relations which $U(n)$ Racah coefficients must satisfy as in recognizing the simple and general origin of such complicated appearing relations.

* Supported in part by the U.S. Army Research Office (Durham), the National Science Foundation, and the U.S. Atomic Energy Commission.

[1] L. C. Biedenharn and J. D. Louck, to appear in Ann. Phys. (N.Y.) (de Shalit Memorial Volume).

[2] The term "symplecton" was introduced in Ref. 1 to denote the boson pair $(a, \bar{a})$. The conjugation symmetric characteristic polynomials in a and $\bar{a}$ are called "symplecton eigenpolynomials." The name (symplecton) was suggested by the fact that a symplecton having n dimensions (n independent boson pairs) generates a representation space for the group $Sp(2n)$—the symplectic group of rank n.

Since the submission of our paper (Ref. 1) we have found that the word "symplecton" was originally coined by H. Freudenthal, Ref. 3, in his discovery of metasymplectic geometry. Freudenthal used the term symplecton to denote a symplectic geometry; in a metasymplectic geometry the elements are points, lines, planes, and symplecta. Our usage of the term thus accords reasonably with his.

[3] H. Freudenthal, Proc. Ned. Akad. Wetenschap. **A62**, 165 (1959).

[4] G. Racah, Phys. Rev. **62**, 438 (1942).

[5] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Rev. Mod. Phys. **24**, 249 (1952).

[6] J. Schwinger, in *Quantum Theory of Angular Momentum*, L. C. Biedenharn and H. Van Dam, Eds. (Academic, New York, 1965).

[7] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).

[8] J. D. Louck, Am. J. Phys. **38**, 3 (1970).

[9] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).

[10] The notation derives from the general notation for a $U(n)$ Wigner operator (see Refs. 7–9). In particular, note the relation between $U(2)$ Wigner operators and $SU(2)$ Wigner operators given by Eq. (3.54) of Ref. 8.

[11] L. C. Biedenharn, in *Spectroscopic and Group Theoretical Methods in Physics* (Racah Memorial Volume) (North-Holland, Amsterdam, 1968), p. 59.

[12] The authors (and no doubt others) have been aware of this relation for some years, and it was stated in Ref. 9 (in a more general context). Particular cases of it have been noted previously [see Eq. (29) of Ref. 13].

[13] L. C. Biedenharn, J. Math. & Phys. **31**, 287 (1953).

[14] J.-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Boundary Conditions in the Pairwise Point Transformation Method

NORMAN M. WITRIOL
Solid State Physics Branch, Physical Sciences Laboratory, Redstone Arsenal, Alabama 35809

(Received 29 April 1970)

Using boundary conditions, one can eliminate the need of a transformation cutoff function in the pairwise point transformation method applied to the many-body problem.

In previous work[1–3] on the many-body problem using the pairwise point transformation method, it was necessary to introduce a cutoff function into the transformation, such that the transformation approached the identity transformation for interparticle separations greater than an arbitrarily chosen cutoff distance.[1] The purpose of this article is to show that one need not introduce a cutoff function into the transformation, provided that one requires the original wavefunction to approach the free-particle wavefunction within an arbitrarily chosen cutoff distance. If this distance is taken to be less than or equal to the interparticle separation, this requirement is essentially that of "pairwise additivity"; namely, after particle "a" has interacted pairwise with particle "b," it must become a free particle before it can interact pairwise with particle "c." In essence, we show that one can put the cutoff into the wavefunction instead of the transformation. This technique could be useful in simplifying the calculations inherent in applying many-body point transforms[4] to many-body systems.

In the pairwise point transformation method,[2] the system is assumed to be sufficiently dilute such that the Hamiltonian H can be assumed to be the sum of pairwise Hamiltonians H_{ij},

$$H = \sum_{\substack{i,j=1 \\ i<j}}^{N} H_{ij}. \tag{1}$$

We shall therefore only consider the 2-body wavefunction associated with H_{ij}. To illustrate the technique, we shall demonstrate it on the hard-core 2-body problem. This problem was previously studied[1] using a cutoff function in the transformation, and the standard results were obtained.

Our notation for equal-mass particles is that the $X_{i\alpha}$ (the $\mathbf{R}_i$) are the original coordinates, the $x_{i\alpha}$ (the $\mathbf{r}_i$) are the transformed coordinates, the $P_{i\alpha}$ are the original momenta, the $p_{i\alpha}$ are the transformed momenta, the $X_{ij\alpha}$, $x_{ij\alpha}$, $P_{ij\alpha}$, and $p_{ij\alpha}$ represent the

relative canonical coordinates and momenta in the original and transformed spaces, respectively [$X_{ij\alpha} = X_{i\alpha} - X_{j\alpha}$, $P_{ij\alpha} = \frac{1}{2}(P_{i\alpha} - P_{j\alpha})$, $\cdots$], the $\bar{p}_{ij\alpha}$ represent the center-of-mass momentum in the transformed space ($\bar{p}_{ij\alpha} = p_{i\alpha} + p_{j\alpha}$), Greek indices stand for the Cartesian coordinates 1 to 3, $r = |\mathbf{r}|$, and we use units of $\hbar = 1$. Under the pairwise point transformation

$$X_{ij\alpha} = x_{ij\alpha}[1 + f_{ij}(\mathbf{r}_{ij})], \tag{2}$$

$$P_{ij\alpha} = \tfrac{1}{2}\sum_{\beta=1}^{3}\left(p_{ij\beta}\frac{\partial x_{ij\beta}}{\partial X_{ij\alpha}} + \frac{\partial x_{ij\beta}}{\partial X_{ij\alpha}}p_{ij\beta}\right), \tag{3}$$

the original Hamiltonian

$$H_{ij} = \frac{1}{2m}\sum_{\alpha=1}^{3}(P_{i\alpha}^2 + P_{j\alpha}^2) + V(\mathbf{R}_{ij}, \mathbf{P}_{ij}) \tag{4}$$

is transformed into

$$H'_{ij} = \sum_{\alpha=1}^{3}\left(\frac{1}{4m}\bar{p}_{ij\alpha}^2 + \frac{1}{m}\sum_{\beta=1}^{3}p_{ij\alpha}g_{\alpha\beta}(\mathbf{r}_{ij})p_{ij\beta}\right) + W(\mathbf{r}_{ij}) + V(\mathbf{r}_{ij}, \mathbf{p}_{ij}), \tag{5}$$

where

$$g_{\alpha\beta}(\mathbf{r}_{ij}) = \sum_{\gamma=1}^{3}\frac{\partial x_{ij\alpha}}{\partial X_{ij\gamma}}\frac{\partial x_{ij\beta}}{\partial X_{ij\gamma}}, \tag{6}$$

$$W(\mathbf{r}_{ij}) = \frac{h^2}{m}\sum_{\alpha,\beta=1}^{3}\frac{1}{4}g_{\alpha\beta}(\mathbf{r}_{ij})\frac{\partial \ln B}{\partial x_{ij\alpha}}\frac{\partial \ln B}{\partial x_{ij\beta}} - \frac{1}{2}\frac{\partial}{\partial x_{ij\alpha}}g_{\alpha\beta}(\mathbf{r}_{ij})\frac{\partial \ln B}{\partial x_{ij\beta}}, \tag{7}$$

B is the Jacobian $|\partial x_{i\alpha}/\partial X_{j\alpha}|$ of the inverse transformation ($\mathbf{r} \to \mathbf{R}$), and $V(\mathbf{r}_{ij}, \mathbf{p}_{ij})$ is $V(\mathbf{R}_{ij}, \mathbf{P}_{ij})$ written in terms of $\mathbf{r}_{ij}$ and $\mathbf{p}_{ij}$. In addition, the original normalized wavefunction $\psi_{ij}(\mathbf{R}_{ij})$ is related to the transformed normalized wavefunction $\psi'_{ij}(\mathbf{r}_{ij})$ by

$$\psi_{ij}(\mathbf{R}_{ij}(\mathbf{r}_{ij})) = [B(\mathbf{r}_{ij})]^{\frac{1}{2}}\psi'_{ij}(\mathbf{r}_{ij}). \tag{8}$$

For convenience, we shall henceforth remove the ij subscript from our variables and only consider the relative part of the 2-body Hamiltonian, Eq. (5).

For the hard-core problem

$$V(R) = \infty, \quad R \le c, \qquad = 0, \quad R > c, \tag{9}$$

the transformation

$$f(\mathbf{r}) = c/r \tag{10}$$

removes the singular potential. Substituting Eqs. (2) and (10) into Eqs. (6) and (7), we obtain[1]

$$g_{\alpha\beta}(\mathbf{r}) = \left(\frac{r}{r+c}\right)^2\delta_{\alpha\beta} + \left[1 - \left(\frac{r}{r+c}\right)^2\right]\frac{x_\alpha x_\beta}{r^2},$$

$$W(\mathbf{r}) = 0, \tag{11}$$

$$B(\mathbf{r}) = \left(\frac{r}{r+c}\right)^2. \tag{12}$$

The energy of the system is given by

$$E = \int \psi'^{\dagger}(\mathbf{r})H'\psi'(\mathbf{r})d^3r \Big/ \int \psi'^{\dagger}(\mathbf{r})\psi'(\mathbf{r})d^3r, \tag{13}$$

where H' is the transformed 2-body Hamiltonian and $\psi'(\mathbf{r})$ is the correct transformed wavefunction of the 2-body system. Substituting Eqs. (11) and (5) into Eq. (13), assuming ψ' to be normalized and the center-of-mass motion to be zero, we obtain

$$E = \frac{1}{m}\int \psi'^{\dagger}(\mathbf{r})\sum_{\alpha,\beta=1}^{3}p_\alpha\left\{\left(\frac{r}{r+c}\right)^2\delta_{\alpha\beta} + \left[1 - \left(\frac{r}{r+c}\right)^2\right]\frac{x_\alpha x_\beta}{r^2}\right\}p_\beta\psi'(\mathbf{r})d^3r. \tag{14}$$

Using $p_\alpha \to -i\partial/\partial x_\alpha$, assuming[5] $\psi'(\mathbf{r}) = \psi'(r)$, taking the system to be confined in a spherical box of radius M in the original space, and integrating by parts, we find that Eq. (14) becomes

$$E = -\frac{4\pi}{m}\left[\psi'^{\dagger}(r)r^2\frac{\partial\psi'(r)}{\partial r}\right]_0^{M-c} + \frac{4\pi}{m}\int\frac{\partial\psi'^{\dagger}(r)}{\partial r}r^2\frac{\partial\psi'(r)}{\partial r}dr. \tag{15}$$

We assume the original wavefunction $\psi(\mathbf{R})$ to be a free-body wavefunction whenever the interparticle distance is greater than an arbitrarily chosen distance λ', where λ' is less than the radius of the box M. This assumption is equivalent to the assumption of "pairwise additivity," if λ' is of the order of the interparticle separation R_0. Therefore, for the ground state we require[6]

$$\psi(\mathbf{R}) = K^{-\frac{1}{2}}, \quad \text{for} \quad R \ge \lambda', \tag{16}$$

where K is a constant. From Eqs. (8) and (12) we obtain

$$\psi'(r) = \left(1 + \frac{c}{r}\right)K^{-\frac{1}{2}}, \quad \frac{\partial\psi'(r)}{\partial r} = -\left(\frac{c}{r^2}\right)K^{-\frac{1}{2}}, \quad \text{for} \quad r \ge \lambda, \tag{17}$$

where $\lambda = \lambda' - c$. We assume $\psi'(r)$ is well behaved around the origin; i.e., the contribution of the first term in Eq. (15), at $r = 0$, is zero. Therefore, if our system is confined to a very large box, $c \ll M$,

$$\left[\psi'^{\dagger}(r)r^2\frac{\partial\psi'(r)}{\partial r}\right]_0^{M-c} \simeq -\frac{c}{K}. \tag{18}$$

Therefore, substituting Eq. (18) into Eq. (15), we obtain

$$E = \frac{4\pi c}{mK} + \frac{4\pi}{m}\int\frac{\partial\psi'^{\dagger}(r)}{\partial r}r^2\frac{\partial\psi'(r)}{\partial r}dr. \tag{19}$$

For the ground state we might try $\psi'(r)$ to be the free-body zero-momentum wavefunction in r space for small r, $r \ll \lambda$, i.e.,

$$\psi'(r) = \bar{K}^{-\frac{1}{2}}, \quad \text{for} \quad r \ll \lambda, \tag{20}$$

where $\bar{K}$ is a constant. Therefore, we might try

$$\psi'(r) = \bar{K}^{-\frac{1}{2}}\varphi(r/\lambda) + K^{-\frac{1}{2}}(1 + c/r)[1 - \varphi(r/\lambda)], \tag{21}$$

where $\varphi(r/\lambda) \equiv \varphi(\xi)$ is the cutoff function,

$$\begin{aligned} \varphi(\xi) &= 1, \quad \xi = 0, \\ &\to 0, \quad \xi \to 1, \\ &= 0, \quad \xi = 1, \end{aligned} \tag{22}$$

$\partial\varphi(\xi)/\partial\xi$ is well behaved, and K and $\bar{K}$ are determined by the normalization of $\psi'(r)$ and the minimization of E. In the dilute system limit we let $\lambda \gg c$. Substituting Eq. (21) into Eq. (19) and neglecting contributions proportional to c/M and c/λ, we obtain $K \simeq \Omega$ and

$$E \simeq 4\pi c/m\Omega, \tag{23}$$

where Ω is the volume of the box containing the system, $\Omega = \frac{4}{3}\pi M^3$. For the N-body problem we have $N(N-1)/2$ pairs and thus, for large N,

$$E \to 2\pi c N^2/m\Omega. \tag{24}$$

If we had not required $\psi(\mathbf{R})$ to be the free-body wavefunction for $R \geq \lambda'$ but had tried $\psi'(r) = \text{const}$, we would have obtained $E = 0$ in Eq. (23), the correct 2-body scattering result. Thus the requirement that $\psi(\mathbf{R})$ be a free-body wavefunction for $R \geq \lambda'$ gives us the proper energy, to first order in c, for the dilute many-body hard-core system taken as a sum of 2-body systems. This requirement with $\lambda' \leq R_0$ is actually equivalent to the assumption of "pairwise additivity."

In conclusion, this method is equivalent to taking a cutoff function in the transformation. For many-body systems with short-range potentials, the taking of the cutoff function in the transformation results in the need of performing very messy algebraic computations.[4] The method of not using a cutoff function in the transformation but of requiring the wavefunction to satisfy certain boundary conditions could significantly simplify these calculations.

ACKNOWLEDGMENT

The author would like to thank Dr. J. D. Stettler for his helpful suggestions.

[1] F. M. Eger and E. P. Gross, Nuovo Cimento **34**, 1225 (1964).
[2] F. M. Eger and E. P. Gross, J. Math. Phys. **7**, 578 (1966).
[3] N. M. Witriol, Phys. Rev. A **1**, 1775 (1970).
[4] N. M. Witriol, J. Math. Phys. **11**, 669 (1970).

[5] For simplicity, we shall only consider "s" states.
[6] We note that, ideally, one would use $\psi(\mathbf{R}) = (\text{const})\sin(kR)/kR$ and take $\lim k \to 0$. The results are the same.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

New Generalized Bessel Transform and Its Relationship to the Fourier, Watson, and Kontorowich–Lebedev Transforms

E. BAHAR
Department of Electrical Engineering, University of Nebraska
Lincoln, Nebraska 68508

(Received 28 May 1970)

A generalized Bessel transform is postulated, and the corresponding inversion formula is deduced. This transform is used to show the direct relationship between the Fourier, Watson, and Kontorowich–Lebedev transforms. Thus, solutions of a certain class of boundary value problems can be expressed in terms of a single Bessel transform. Furthermore, the curvature of the boundary may range from zero to infinity.

1. INTRODUCTION

A large variety of boundary value problems have been solved by using transform techniques based on the inversion of definite integrals. Thus, instead of seeking directly the required solution, say $E_z(r, \phi)$, that satisfies a certain partial differential equation and specified boundary conditions, we seek the solution of the transform of the desired solution, $E(\nu, \phi)$, which satisfies an ordinary differential equation with the corresponding boundary conditions. When the transform $E(\nu, \phi)$ is determined, the desired solution is found by using the inversion formula. For simplicity,

For the ground state we might try $\psi'(r)$ to be the free-body zero-momentum wavefunction in r space for small r, $r \ll \lambda$, i.e.,

$$\psi'(r) = \bar{K}^{-\frac{1}{2}}, \quad \text{for} \quad r \ll \lambda, \tag{20}$$

where $\bar{K}$ is a constant. Therefore, we might try

$$\psi'(r) = \bar{K}^{-\frac{1}{2}}\varphi(r/\lambda) + K^{-\frac{1}{2}}(1 + c/r)[1 - \varphi(r/\lambda)], \tag{21}$$

where $\varphi(r/\lambda) \equiv \varphi(\xi)$ is the cutoff function,

$$\begin{aligned} \varphi(\xi) &= 1, \quad \xi = 0, \\ &\to 0, \quad \xi \to 1, \\ &= 0, \quad \xi = 1, \end{aligned} \tag{22}$$

$\partial\varphi(\xi)/\partial\xi$ is well behaved, and K and $\bar{K}$ are determined by the normalization of $\psi'(r)$ and the minimization of E. In the dilute system limit we let $\lambda \gg c$. Substituting Eq. (21) into Eq. (19) and neglecting contributions proportional to c/M and c/λ, we obtain $K \simeq \Omega$ and

$$E \simeq 4\pi c/m\Omega, \tag{23}$$

where Ω is the volume of the box containing the system, $\Omega = \frac{4}{3}\pi M^3$. For the N-body problem we have $N(N-1)/2$ pairs and thus, for large N,

$$E \to 2\pi c N^2/m\Omega. \tag{24}$$

If we had not required $\psi(\mathbf{R})$ to be the free-body wavefunction for $R \geq \lambda'$ but had tried $\psi'(r) = \text{const}$, we would have obtained $E = 0$ in Eq. (23), the correct 2-body scattering result. Thus the requirement that $\psi(\mathbf{R})$ be a free-body wavefunction for $R \geq \lambda'$ gives us the proper energy, to first order in c, for the dilute many-body hard-core system taken as a sum of 2-body systems. This requirement with $\lambda' \leq R_0$ is actually equivalent to the assumption of "pairwise additivity."

In conclusion, this method is equivalent to taking a cutoff function in the transformation. For many-body systems with short-range potentials, the taking of the cutoff function in the transformation results in the need of performing very messy algebraic computations.[4] The method of not using a cutoff function in the transformation but of requiring the wavefunction to satisfy certain boundary conditions could significantly simplify these calculations.

ACKNOWLEDGMENT

The author would like to thank Dr. J. D. Stettler for his helpful suggestions.

[1] F. M. Eger and E. P. Gross, Nuovo Cimento **34**, 1225 (1964).
[2] F. M. Eger and E. P. Gross, J. Math. Phys. **7**, 578 (1966).
[3] N. M. Witriol, Phys. Rev. A **1**, 1775 (1970).
[4] N. M. Witriol, J. Math. Phys. **11**, 669 (1970).

[5] For simplicity, we shall only consider "s" states.
[6] We note that, ideally, one would use $\psi(\mathbf{R}) = (\text{const})\sin(kR)/kR$ and take $\lim k \to 0$. The results are the same.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

New Generalized Bessel Transform and Its Relationship to the Fourier, Watson, and Kontorowich–Lebedev Transforms

E. BAHAR
Department of Electrical Engineering, University of Nebraska
Lincoln, Nebraska 68508

(Received 28 May 1970)

A generalized Bessel transform is postulated, and the corresponding inversion formula is deduced. This transform is used to show the direct relationship between the Fourier, Watson, and Kontorowich–Lebedev transforms. Thus, solutions of a certain class of boundary value problems can be expressed in terms of a single Bessel transform. Furthermore, the curvature of the boundary may range from zero to infinity.

1. INTRODUCTION

A large variety of boundary value problems have been solved by using transform techniques based on the inversion of definite integrals. Thus, instead of seeking directly the required solution, say $E_z(r, \phi)$, that satisfies a certain partial differential equation and specified boundary conditions, we seek the solution of the transform of the desired solution, $E(\nu, \phi)$, which satisfies an ordinary differential equation with the corresponding boundary conditions. When the transform $E(\nu, \phi)$ is determined, the desired solution is found by using the inversion formula. For simplicity,

we shall restrict our attention initially to cylindrical structures and assume the solutions are independent of the z axis.

In Sec. 2 we postulate a generalized Bessel transform and deduce the corresponding inversion formula. This transform pair is shown to merge with the familiar Fourier integral transform when the parameter corresponding to the curvature of the boundary vanishes. It is identified with the discrete Watson transform when the parameter related to the curvature is finite, and it is shown to reduce on the Kontorowich–Lebedev transform when the curvature parameter approaches infinity.

In Sec. 3 we express the solution to the problem of radiation by a line source in the presence of a cylindrical structure, in terms of the generalized Bessel transform. The limiting forms of this solution, as the curvature of the diffracting structure is made to range from zero (a plane surface) to infinity, are then derived.

This transform representation is of special interest for solving problems in which the local curvature of the diffracting structure is varying.

2. THE GENERALIZED BESSEL TRANSFORM AND THE DEDUCTION OF THE CORRESPONDING INVERSION FORMULA

Consider the problem of diffraction of horizontally polarized waves by an infinite circular cylindrical structure of radius $r = R$. (See Fig. 1.) For simplicity of presentation, we shall assume initially that the cylindrical surface is perfectly conducting. Subsequently, the boundary of the cylinder is characterized by a finite surface impedance, and the presentation is generalized. Thus, problems involving diffraction of vertically polarized waves can be analyzed in a similar manner.

We postulate that a general solution for the z-independent electric field (z component only) may be expressed as

$$E_z(\xi, \phi) = \int_{-\infty}^{\infty} E(\nu, \phi)\psi_\nu(\xi)\, d\nu, \tag{2.1}$$

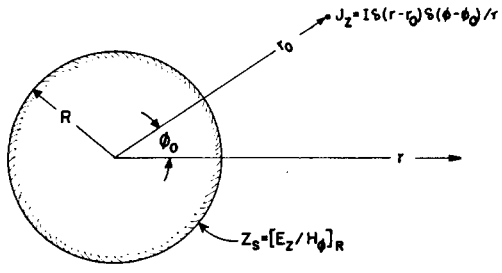

FIG. 1. Radiation by an electric line source parallel to a circular cylinder characterized by a surface impedance boundary.

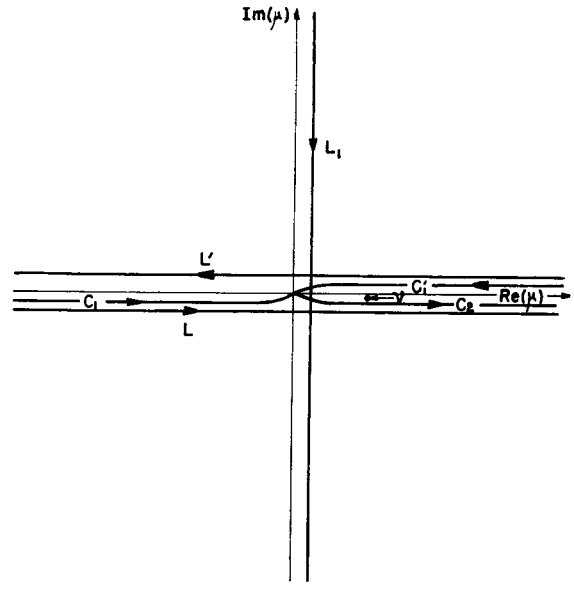

FIG. 2. Integration paths in the complex μ plane.

in which $\xi = kr$ and k is the wavenumber. The transform of the function $E_z(\xi, \phi)$ is $E(\nu, \phi)$. The basis function $\psi_\nu(\xi)$ satisfies the homogeneous scalar wave equation $L_\rho[\psi_\nu] = (\nu/\xi)^2\psi_\nu$, where the operator L_ρ is defined by

$$L_\rho[\psi_\nu] - \left(\frac{\nu}{\xi}\right)^2\psi_\nu \equiv \frac{1}{\xi}\frac{d}{d\xi}\left(\xi\frac{d}{d\xi}\psi_\nu(\xi)\right) + \left[1 - \left(\frac{\nu}{\xi}\right)^2\right]\psi_\nu(\xi) = 0, \tag{2.2a}$$

and the boundary condition

$$\psi_\nu(\xi_R) = 0 \tag{2.2b}$$

in which $\xi_R = kR$. Thus,

$$\psi_\nu(\xi) = H_\nu^{(1)}(\xi) + R_\nu H_\nu^{(2)}(\xi), \tag{2.3a}$$

where

$$R_\nu = -H_\nu^{(1)}(\xi_R)/H_\nu^{(2)}(\xi_R) \tag{2.3b}$$

and $H_\nu^{(1,2)}(\xi)$ are the Hankel functions of order ν of the first and second kind, respectively. In this paper, we assume an exp ($i\omega t$) time excitation.

In order to justify the representation of the field $E_z(\xi, \phi)$ in terms of a continuous spectrum of basis functions, $\psi_\nu(\xi)$ (ν varies along the entire real axis, path L in Fig. 2), it is necessary to demonstrate that the transform function $E(\nu, \phi)$ may be derived uniquely from the expression for $E_z(\xi, \phi)$. We shall show that this is achieved by the inverse transform

$$E(\nu, \phi) = \frac{1}{4}\int_{\xi_R}^{\infty} E_z(\xi, \phi)H_\nu^{(2)}(\xi)\frac{\nu}{\xi}\, d\xi. \tag{2.4}$$

To this end, we must show that

$$E(\nu, \phi) = \frac{1}{4}\int_{\xi_R}^{\infty}\left(\int_{-\infty}^{\infty} E(\mu, \phi)\psi_\mu(\xi)\, d\mu\right) H_\nu^{(2)}(\xi)\frac{\nu}{\xi}\, d\xi. \tag{2.5a}$$

The above relationship can also be expressed as follows:

$$\delta(\nu, \mu) = \frac{1}{4}\int_{\xi_R}^{\infty}\psi_\mu(\xi)H_\nu^{(2)}(\xi)\frac{\nu}{\xi}\, d\xi, \tag{2.5b}$$

where $\delta(\nu, \mu)$ is the Dirac δ function.

For the purpose of the proof of the above theorem, it is necessary to show that, in the integral (2.5a), the order of integration can be changed. To justify this step, it is necessary to study certain estimations of the Bessel functions which can be found from their integral representations. A rigorous study of this nature has been carried out in detail by Kontorowich and Lebedev[1] and is therefore omitted from this paper. Thus, on interchanging the orders of integration and integrating with respect to ξ and on making use of the Debye–Watson expansions for the Hankel functions,[2,3] we obtain, from (2.5a),

$$\begin{aligned} E(\nu, \phi) &= \frac{\nu}{4}\int_{-\infty}^{\infty} E(\mu, \phi)\, d\mu \int_{\xi_R}^{\infty} H_\nu^{(2)}(\xi)\psi_\mu(\xi)\frac{d\xi}{\xi} \\ &= -\frac{\nu}{4}\int_{-\infty}^{\infty}\frac{E(\mu, \phi)}{\mu^2 - \nu^2}[\xi\{H_\nu^{(2)}(\xi)\psi_\mu'(\xi) \\ &\quad - H_\nu^{(2)\prime}(\xi)\psi_\mu(\xi)\}]_{\xi_R}\, d\mu, \end{aligned} \tag{2.6a}$$

for $0 > \text{Im}\,(\nu) > \text{Im}\,(\mu)$. To complete the derivation, we express (2.6a) as

$$\begin{aligned} E(\nu, \phi) &= -\frac{1}{8}\int_L\left(\frac{1}{\mu - \nu} - \frac{1}{\mu + \nu}\right)E(\mu, \phi)\xi_R \\ &\quad \times [H_\nu^{(2)}(\xi)\psi_\mu'(\xi) - H_\nu^{(2)\prime}(\xi)\psi_\mu(\xi)]_{\xi_R}\, d\mu \\ &= -\frac{1}{8}\int_{L+L'}\frac{1}{\mu - \nu}E(\mu, \phi)\xi_R \\ &\quad \times [H_\nu^{(2)}(\xi)\psi_\mu'(\xi) - H_\nu^{(2)\prime}(\xi)\psi_\mu(\xi)]_{\xi_R}\, d\mu, \end{aligned} \tag{2.6b}$$

where the path L' is shown in Fig. 2. In (2.6b) we have also made use of the relationships between the Hankel functions of positive and negative orders, and from (2.4) it follows that

$$E(-\mu, \phi) = -E(\mu, \phi)\exp(-i\nu\pi). \tag{2.7a}$$

Thus, applying Cauchy's residue theorem, we reduce the contour integral (2.6b) to

$$\begin{aligned} E(\nu, \phi) &= E(\nu, \phi)(-\tfrac{1}{4}\pi i)\xi_R W[H_\nu^{(2)}(\xi), H_\nu^{(1)}(\xi)]_{\xi_R} \\ &= E(\nu, \phi). \end{aligned} \tag{2.7b}$$

In (2.7a) we have used the value of the Wronskian, W, which is

$$W[H_\nu^{(2)}(\xi), H_\nu^{(1)}(\xi)] = 4i/\pi\xi. \tag{2.7c}$$

We now present the relationship between the transform pair (2.1), (2.4) and the Watson transform. Note that in (2.1) the poles of the integrand in the lower half-plane [i.e., for Im $(\mu) < 0$] are located at the roots of the equation

$$1/R_\nu = 0, \quad \text{i.e.,} \quad H_{\nu_m}^{(2)}(\xi_R) = 0, \quad m = 1, 2, 3, \cdots. \tag{2.8a}$$

Thus, by closing the contour in (2.1) with an infinite semicircle in the lower half-plane and using Cauchy's theorem, we can express (2.1) as

$$E_z(\xi, \phi) = (2\pi i)\sum_{m=1}^{\infty}\frac{E(\nu_m, \phi)H_{\nu_m}^{(2)}(\xi)H_{\nu_m}^{(1)}(\xi_R)}{[\partial H_\nu^{(2)}(\xi_R)/\partial\nu]_{\nu=\nu_m}}. \tag{2.8b}$$

Multiply both sides of (2.8b) by $\nu_n H_{\nu_n}^{(2)}(\xi)/4\xi$ and integrate with respect to ξ over the range $kR \leq \xi < \infty$. Using the orthogonal properties of the Hankel functions $H_{\nu_m}^{(2)}(\xi)$, we obtain directly the expression for $E(\nu_n, \phi)$ given by (2.4). Thus, the Watson transform pair (2.8), (2.4) and the generalized Bessel transform pair (2.1), (2.4) are equivalent for finite and nonzero values of R.

To show that the Bessel transform pair reduces to the Kontorowich–Lebedev transform for the case $R = 0$, we note that there are no roots to the equation $R_\nu = 0$ since

$$\lim_{R\to 0}[R_\nu] \to 1. \tag{2.9a}$$

Furthermore, since

$$\lim_{R\to 0}\psi_\nu(\xi) \to 2J_\nu(\xi) \tag{2.9b}$$

and

$$H_{-\nu}^{(1,2)}(\xi) = \exp(\pm i\nu\pi)H_\nu^{(1,2)}(\xi), \tag{2.9c}$$

it follows from (2.5a) that the original path of integration L in the ν plane consisting of sections C_1 and C_2 (see Fig. 2) may be replaced by the path consisting of the sections C_1' and C_2.[4] Finally, using Jordan's lemma, we may transform this path to the one L_1 along the imaginary ν axis. Thus, for $R \to 0$, the appropriate form of the generalized Bessel transform is the Kontorowich–Lebedev transform[1]

$$E_z(\xi, \phi) = \int_{-i\infty}^{i\infty} E(\nu, \phi)J_\nu(\xi)\, d\nu \tag{2.10a}$$

and

$$E(\nu, \phi) = -\frac{1}{2}\int_0^{\infty} E_z(\xi, \phi)H_\nu(\xi)\frac{\nu}{\xi}\, d\nu. \tag{2.10b}$$

For the case $R \to \infty$ the generalized Bessel transform pair can be shown to merge with the Fourier transform pair. To this end, we renormalize the transform function $E(\nu, \phi)$ by defining a new transform function $E'(\nu, \phi)$ as follows:

$$E(\nu, \phi) = (\nu/\xi_R)E'(\nu, \phi)H_\nu^{(2)}(\xi_R). \tag{2.11}$$

Thus, the pair of equations (2.1), (2.4) may also be written as

$$E_z(\xi, \phi) = \int_{-\infty}^{\infty} E(\nu, \phi)H_\nu^{(2)}(\xi_R)\psi_\nu^{(2)}(\xi)\frac{\nu}{\xi_R}\,d\nu \tag{2.12a}$$

and

$$E(\nu, \phi) = \frac{1}{4}\int_{\xi_R}^{\infty} E_z(\xi, \phi)\left(\frac{H_\nu^{(2)}(\xi)}{H_\nu^{(2)}(\xi_R)}\right)\frac{\xi_R}{\xi}\,d\xi. \tag{2.12b}$$

Now using the WKB-type Debye–Watson expansion for the Hankel function, we can show that[3,4]

$$H_\nu^{(2)}(\xi)/H_\nu^{(2)}(\xi_R) \to e^{-ikCy} \tag{2.13a}$$

and

$$H_\nu^{(2)}(\xi_R)\psi_\nu^{(2)}(\xi) \to (4i/\pi\xi C)\sin kCy, \tag{2.13b}$$

where, as $R \to \infty$, the variable y and the new integration parameter C are defined as

$$(\xi - \xi_R) = k(r - R) \to ky \tag{2.13c}$$

and

$$[1 - (\nu/\xi)^2]^{\frac{1}{2}} \to (1 - S)^{\frac{1}{2}} \equiv C, \quad \text{Im}\,(C) \leq 0. \tag{2.13d}$$

Thus it follows that the appropriate (renormalized) form for (2.6), for the case $R \to \infty$, is

$$E(S, \phi) = \frac{i}{\pi}\int_0^{\infty}\int_{-\infty}^{\infty} E(S', \phi) \times e^{-ikCy}\sin(kC'y)\frac{S'\,dS'}{C'}\,d(ky). \tag{2.14}$$

Noting that

$$\frac{d\nu}{\xi} \to dS = \frac{-C\,dC}{S}, \tag{2.15}$$

we may now express (2.14) in the form of the familiar Fourier sine transform pair

$$E_z(y, \phi) = \int_0^{\infty} E(C, \phi)\sin(kCy)\,dC \tag{2.16a}$$

and

$$E(C, \phi) = \frac{2}{\pi}\int_0^{\infty} E_z(y, \phi)\sin(kCy)d(ky). \tag{2.16b}$$

The above results may be generalized for diffractive boundaries characterized by finite surface impedances. Thus at $r = R$ we write, for the surface impedance,

$$Z_S = [E_z/H_\phi]_R, \tag{2.17a}$$

where H_ϕ is the azimuthal component of the magnetic field. Using Maxwell's electromagnetic field equations, we express the boundary condition (2.17a) as

$$\frac{\partial E_z}{\partial \xi} = \frac{i\eta_0}{Z_S}E_z \equiv \frac{iE_z}{z_s}, \tag{2.17b}$$

where z_s is the normalized surface impedance and η_0 is the free-space wave impedance. In this case, the coefficient R_ν in (2.3a) must be given by the expression

$$R_\nu = -\frac{H_\nu^{(1)'}(\xi_R) - iy_sH_\nu^{(1)}(\xi_R)}{H_\nu^{(2)'}(\xi_R) - iy_sH_\nu^{(2)}(\xi_R)}, \tag{2.17c}$$

where

$$H_\nu^{(1,2)'}(\xi_R) = \left[\frac{\partial}{\partial\xi}H_\nu^{(1,2)}(\xi)\right]_{\xi_R} \quad \text{and} \quad y_s = \frac{1}{z_s}. \tag{2.17d}$$

Thus, for the surface impedance case, the corresponding form for (2.14) can be shown to be

$$E(S, \phi) = \frac{1}{2\pi}\int_0^{\infty}\int_{-\infty}^{\infty} E(S', \phi)\left(\frac{C'}{y_s} + 1\right) \times [e^{ikC'y} + R(C')e^{-ikC'y}] \times \frac{e^{-ikCy}}{[(C/y_s) + 1]}\frac{S'\,dS'}{C'}\,d(ky), \tag{2.18a}$$

where

$$R(C) = \lim_{\xi_R \to \infty}\frac{H_\nu^{(2)}(\xi_R)}{H_\nu^{(1)}(\xi_R)}R_\nu \to \frac{[(C/y_s) - 1]}{[(C/y_s) + 1]} \tag{2.18b}$$

is the reflection coefficient for horizontally polarized waves over a plane surface. It is interesting to note that for the surface impedance case it is necessary to take into consideration the singularity at

$$1/R(C) = 0, \quad \text{i.e.,} \quad C = C_0 = -y_s. \tag{2.18c}$$

Thus, (2.18c), together with the condition (2.13d), indicates that, for C_0 in the third quadrant (i.e., y_s in the first quadrant), a discrete modal solution exists. This term, $\exp(ikyy_s)$, is recognized to be the surface wave in radio propagation.[5] In this case, therefore, a continuous spectrum of propagating and evanescent waves, together with the surface wave, must be used in a complete representation of the solution.[6] We transform the path L in the S plane to the one consisting of $D_1 + D_2$ along the branch cut (see Fig. 3) and account for the residue at the surface wave pole ($C = C_0$) to evaluate the integral (2.18a). Thus, using the relationships

$$\psi(C, y) \equiv [(C/y_s) + 1]e^{ikCy} + [(C/y_s) - 1]e^{ikCy} = -\psi(-C, y) \tag{2.19a}$$

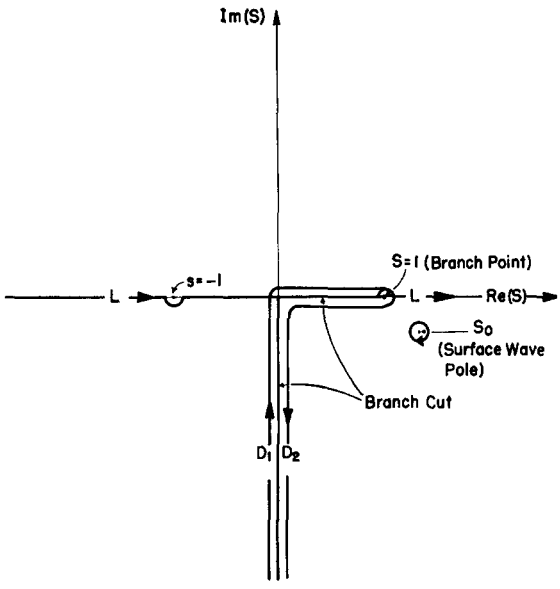

FIG. 3. Integration paths in the complex S plane.

and

$$\sigma(C, y) - \sigma(-C, y) = \frac{\psi(C, y)}{(C/y_s + 1)(C/y_s - 1)}, \tag{2.19b}$$

where

$$\sigma(C, y) = e^{-ikCy}/(C + y_s), \tag{2.19c}$$

we can write Eq. (2.18a) as the transform pair

$$E_z(y, \phi) = (2i)^{-1}\left(\int_0^\infty E(C, \phi)\psi(C, y)\, dC + E(C_0, \phi)\psi(C_0, y)\right) \tag{2.20a}$$

and

$$E(C, \phi) = \frac{i}{\pi}\int_0^\infty \frac{E_z(\xi, \phi)\psi(\xi, \phi)d(ky)}{[(C/y_s) + 1][(C/y_s) - 1]}, \tag{2.20b}$$

for the continuous spectrum along the positive real axis (the radiation term) and for the discrete mode (the surface wave)

$$E(C_0, \phi) = y_s\int_0^\infty E_z(y, \phi)\psi(C_0, y)d(ky). \tag{2.20c}$$

For $y_s \to \infty$, the surface-wave basis function $\psi(C_0, y)$ vanishes and the transform pair (2.20a), (2.20b) reduces to (2.16a), (2.16b). Several other forms for the generalized Bessel transform pair can be established. The transverse component of the magnetic field H_ϕ, for example, must be expressed in terms of the basis functions

$$Y_\nu(\xi)\psi_\nu(\xi) = (\nu/\eta\xi)\psi_\nu(\xi), \tag{2.21a}$$

where η is the intrinsic impedance of the medium of propagation.

Thus, the appropriate transform pair for this case is

$$H_\phi(\xi, \phi) = \int_{-\infty}^\infty H(\nu, \phi)Y_\nu(\xi)\psi_\nu(\xi)\, d\nu \tag{2.21b}$$

and

$$H(\nu, \phi) = \eta\int_{\xi_R}^\infty H_\phi(\xi, \phi)H_\nu^{(2)}(\xi)\, d\xi. \tag{2.21c}$$

Here $Y_\nu(\xi)$ is identified as the transverse wave admittance for cylindrical waves.

For problems with spherical boundaries, it is necessary to express the solution in terms of the spherical Hankel functions

$$h_\nu^{(1,2)}(\xi) = (\pi/2\xi)^{\frac{1}{2}}H_\nu^{(1,2)}(\xi). \tag{2.22a}$$

Thus, for these problems, we use the following transform pair:

$$E(\xi, \theta) = \int_{-\infty}^\infty E(\mu, \theta)\psi_\mu(\xi)\, d\mu \tag{2.22b}$$

and

$$E(\mu, \theta) = (2\pi)^{-1}\int_{\xi_R}^\infty E(\xi, \theta)h_\mu^{(2)}(\xi)\mu\, d\xi. \tag{2.22c}$$

Here the basis function is

$$\psi_\mu(\xi) = h_\mu^{(1)}(\xi) + R_\mu h_\mu^{(2)}(\xi), \tag{2.22d}$$

and R_μ is chosen such that $\psi_\mu(\xi)$ satisfies the appropriate boundary conditions at $\xi = \xi_R$. A similar transform pair may be derived in terms of the functions $\hat{h}(\xi)$,

$$\hat{h}^{(1,2)}(\xi) = \xi h^{(1,2)}(\xi), \tag{2.23a}$$

which satisfy the differential equation

$$\frac{\partial^2 E}{\partial \xi^2} + [1 - \nu(\nu + 1)]E = 0. \tag{2.23b}$$

The expressions for the transform pair (2.19) can be verified with no difficulty by following the steps outlined in this section.

In the following section, we use the generalized Bessel transform pair to derive the well-known solutions to the problem of radiation from an infinite electric line source in the presence of a conducting cylinder. This will provide a simple illustration of the technique and show, in particular, that the same form for the expansion of the fields may be used for all values of the radius of curvature, R, including cases in which $R \to 0$ and $R \to \infty$.

3. RADIATION FROM A LINE SOURCE IN THE PRESENCE OF A CYLINDRICAL STRUCTURE

As an illustrative example, we shall solve the problem of radiation by an electric line source in the presence of a perfectly conducting cylindrical structure,

using the generalized cylindrical transform discussed in the preceding sections (see Fig. 1).

The electromagnetic fields may be derived from the z-directed vector potential $\bar{A} = A(\xi, \phi)\bar{a}_z$ which satisfies the inhomogeneous wave equation

$$\frac{1}{\xi}\frac{\partial}{\partial \xi}\left(\xi \frac{\partial A}{\partial \xi}\right) + \frac{1}{\xi^2}\frac{\partial^2 A}{\partial \phi^2} + A = -\frac{\mu_0}{k^2} J_z = -\frac{\mu_0 I \delta(\xi - \xi_0)\delta(\phi - \phi_0)}{\xi}, \quad (3.1)$$

where $\xi_0 = kr_0$, μ_0 is the permeability, and I is the intensity of the current filament located at $r = r_0$ and $\phi = \phi_0$. Following (2.4), we express $A(\xi, \phi)$ as follows in terms of its transform $a(\mu, \phi)$:

$$A(\xi, \phi) = \int_{-\infty}^{\infty} a(\mu, \phi)\psi_\mu(\xi)\, d\mu. \quad (3.2)$$

Similarly, the Dirac δ function $\delta(\xi - \xi_0)$ is represented in terms of its transform. Thus, using (2.1) and (2.4), we obtain

$$\xi\delta(\xi - \xi_0) = \tfrac{1}{4}\int_{-\infty}^{\infty} \psi_\mu(\xi)H_\mu^{(2)}(\xi_0)\mu\, d\mu, \quad \xi < \xi_0. \quad (3.3)$$

For $\xi > \xi_0$ the appropriate expression for $\delta(\xi - \xi_0)$ is obtained by interchanging ξ and ξ_0 in (3.3). Substituting (3.2) and (3.3) into (3.1) and noting the uniqueness of the transform, we obtain, on employing (2.2a), the ordinary differential equation for $a(\mu, \phi)$,

$$\left(\frac{d^2}{d\phi^2} + \nu^2\right)a(\mu, \phi) = -\tfrac{1}{4}\mu_0 I \mu H_\mu^{(2)}(\xi_0)\delta(\phi - \phi_0). \quad (3.4a)$$

To solve for $a(\mu, \phi)$, we note that[7]

$$a(\mu, \phi_0) = a(\mu, \phi_0 + 2\pi)$$

and

$$\left[\frac{d}{d\phi}a(\mu, \phi)\right]_{2\pi+\phi_0}^{\phi_0} = -\tfrac{1}{4}\mu_0 I \mu H_\mu^{(2)}(\xi_0), \quad 0 \le \phi - \phi_0 \le 2\pi. \quad (3.4b)$$

Hence,

$$a(\mu, \phi) = -\tfrac{1}{8}\mu_0 I H_\mu^{(2)}(\xi_0)\cos\mu(\phi - \phi_0 - \pi)/\sin\mu\pi = -\tfrac{1}{8}\mu_0 I H_\mu^{(2)}(\xi_0)[\cot\mu\pi\cos\mu(\phi - \phi_0) + \sin\mu(\phi - \phi_0)]. \quad (3.5)$$

Since $H_\mu^{(2)}(\xi_0)\psi_\mu(\xi)\sin\mu(\phi - \phi_0)$ is an odd function of μ, we get on substituting (3.5) into (3.2) the solution to the problem:

$$A(\xi, \phi) = -\tfrac{1}{8}\mu_0 I \int_{-\infty}^{\infty} H_\mu^{(2)}(\xi_0)\psi_\mu(\xi) \times \cot\mu\pi\cos\mu(\phi - \phi_0)\, d\mu. \quad (3.6)$$

This solution can be reduced to the familiar harmonic form by noting, as in Sec. 2, that the contour of integration $C_1' + C_2$ can be substituted for the path L (see Fig. 2). Thus, using Cauchy's integral theorem, we obtain, from the residue at the poles along the positive real axis, $\nu = n = 0, 1, 2, 3, \cdots$,

$$A(\xi, \phi) = -i\tfrac{1}{8}\mu_0 I \sum_n \epsilon_n H_n(\xi_0)\psi_n(\xi)\cos n(\phi - \phi_0), \quad (3.7a)$$

where

$$\epsilon_n = \begin{cases} 1, & n = 0 \\ 2, & n = 1, 2, 3, \cdots \end{cases}, \quad (3.7b)$$

since we take only half the value of the residue at $n = 0$. The Watson expansion for $A(\xi, \phi)$ may be derived from (3.6) by closing the path L with the infinite semicircle in the lower half-plane. Thus, noting that the poles in the lower half-plane are at the values of ν_n that satisfy the modal equation (2.8a), we see that

$$A(\xi, \phi) = -\tfrac{1}{4}i\pi\mu_0 I \times \sum_n \left[H_\mu^{(2)}(\xi_0)\left(\frac{H_\mu^{(1)}(\xi_R)}{\partial[H_\mu^{(2)}(\xi_R)]_\mu/\partial\mu}\right)H_\mu^{(2)}(\xi) \times \cot\mu\pi\cos\mu(\phi - \phi_0)\right]_{\mu=\mu_n}. \quad (3.8)$$

Now we derive the forms for $A(\xi, \phi)$ as $\xi_R \to 0$. Obviously, the Watson transform does not exist, and (3.6) reduces to

$$A(\xi, \phi) = -\tfrac{1}{4}\mu_0 I \int_{-\infty}^{\infty} H_\mu^{(2)}(\xi_0)J_\mu(\xi) \times \cot\mu\pi\cos\mu(\phi - \phi_0)\, d\mu. \quad (3.9)$$

Similarly, using the addition theorem for Bessel functions, we get from (3.7a)

$$A(\xi, \phi) = -\tfrac{1}{4}i\mu_0 I \sum_n \epsilon_n H_n^{(2)}(\xi_0)J_n(\xi)\cos n(\phi - \phi_0) = -\tfrac{1}{4}i\mu_0 I H_0^{(2)}([\xi^2 + \xi_0^2 - 2\xi\xi_0\cos(\phi - \phi_0)]^{\frac{1}{2}}). \quad (3.10)$$

For the case $\xi_R \to \infty$ ($\xi - \xi_R \to y$) the cylinder becomes a flat surface, and again the discrete Watson transform does not exist. We express the solution in terms of the Cartesian coordinates x, y. Thus, making use of the WKB-type expansions for the Hankel functions (2.13), we get

$$kR\phi \to kx \quad \text{and} \quad \nu \to kRS.$$

Hence, $\cot\nu\pi \to i\coth(ikRS\pi) \to i$ along the path L, and (3.6) reduces to the Fourier expansion

$$A(x, y) = \frac{\mu I}{2\pi}\int_{-\infty}^{\infty} \exp(-ikCy_0) \times \sin kCy\cos kS(x - x_0)\frac{dS}{C}. \quad (3.11)$$

It is obvious that, in this case, it is not possible to express $A(x, y)$ in terms of a (discrete) harmonic expansion similar to (3.7).

Thus, we have derived in this section the connection between the continuous spectral representation of the solution, the discrete modal expansion (for finite values of R), and the discrete harmonic expansion (for bounded values of R). The various forms for the solution $A(\xi, \phi)$ may be verified directly. To obtain the harmonic expansions (R bounded), we begin by expanding the function $\delta(\phi - \phi_0)$ in (3.1) in terms of the periodic functions $\cos n(\phi - \phi_0)$. The Watson expansions may be derived by expanding the function $\delta(\xi - \xi_0)$ directly in terms of the complete set of basis functions $H_{\nu_n}^{(2)}(\xi)$ that satisfy the appropriate boundary conditions at $\xi = \xi_R$ [(2.8a)]. To obtain the Kontorowich–Lebedev expansion, we begin by obtaining an expression for $\delta(\xi - \xi_0)$ using the transform pair (2.10). Finally, to obtain the Fourier expansion for the solution when $R \to \infty$, we begin the solution of (3.1) [with $J_z = I\delta(x - x_0)\delta(y - y_0)$] by substituting the Fourier transform of the function $\delta(x - x_0)$:

$$\delta(x - x_0) = (2\pi)^{-1} \int_{-\infty}^{\infty} \cos \alpha(x - x_0)\, d\alpha. \quad (3.12)$$

4. CONCLUDING REMARKS

We have demonstrated that it is possible to derive the solution to the diffraction problem (Sec. 3) for all values of the radius of curvature R (including both zero and infinity) through the use of a single transform pair.

It is, therefore, possible to express the solution to the problem of diffraction by convex objects [including corners of continuously varying radius of curvature (see Fig. 4)] in terms of the generalized Bessel transform. In this case, however, the basis function $\psi_\mu(\xi)$ is also a continuous function of the second variable x ($dx = R\, d\phi$), since it satisfies the varying, local, boundary condition. The solution of this problem which follows a similar analysis for closed structures[8] is beyond the scope of this presentation.[9] The continuous spectral representation of the solution, which is determined here directly (without recourse to the harmonic representation), is also useful in reducing the problem of multipath propagation to a set of integral equations. Recently, Wait has derived exact solutions to these integral equations using the Wiener–Hopf technique.[10]

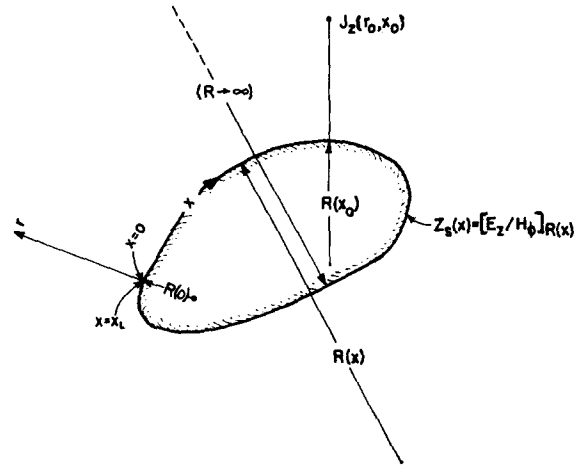

FIG. 4. Radiation by an electric line source parallel to a convex cylinder of arbitrary cross section and variable surface impedance boundary.

ACKNOWLEDGMENTS

The author wishes to thank K. Shere and F. Ullman for their comments and J. R. Wait for stimulating discussions. The manuscript was prepared by Mrs. M. Alles. This work has been supported in part by a grant from the University of Nebraska Engineering Research Center.

[1] M. J. Kontorowich and N. N. Lebedev, J. Phys. **1**, 229 (1939).
[2] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1952).
[3] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[4] J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, Oxford, 1962).
[5] H. M. Barlow and J. Brown, *Radio Surface Waves*, International Monograms on Radio (Clarendon, Oxford, 1962).
[6] R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960).
[7] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, London, 1960).
[8] E. Bahar, IEEE Trans. Microwave Theory Techniques **17**, 210 (1969).
[9] E. Bahar, J. Math. Phys. **12**, 186 (1971).
[10] J. R. Wait, J. Math. Phys. **11**, 2851 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Diffraction of Electromagnetic Waves by Cylindrical Structures Characterized by Variable Curvature and Surface Impedance

E. BAHAR
Department of Electrical Engineering, University of Nebraska
Lincoln, Nebraska 68508

(Received 8 May 1970)

The electromagnetic fields around a cylindrical boundary of variable curvature and surface impedance are expressed in terms of a complete set of local cylindrical modes, and Maxwell's equations are converted into coupled differential equations for the forward and backward wave amplitudes. The local cylindrical modes are shown to merge with the Fourier type expansion of the fields over plane surfaces. As an illustrative example, the launching of a surface wave over an infinite wedge with a rounded corner is considered in detail.

1. INTRODUCTION

In this paper, we derive an analytic solution to the problem of diffraction of electromagnetic waves by a convex cylindrical boundary, characterized by a varying radius of curvature $R(x)$ and surface impedance $Z_s(x)$. The diffracting structure may be of either finite or infinite cross section (Figs. 1 and 2).

The solutions are expressed in terms of a complete set of local cylindrical modes. Using a generalized Bessel transform, we show that the discrete Watson expansions of the fields merge with the Fourier-type (plane wave) expansions above plane boundaries ($R \to \infty$) and with the Kontorowich–Lebedev transform[1] for $R \to 0$. Thus, in our analysis, no restrictions are made on the local radius of curvature $R(x)$; however, the surface impedance variations must be compatible with the surface impedance concept.[2]

The expansion of the solution in terms of the local cylindrical modes provides the basis for the transformation of Maxwell's equations into a set of coupled first-order differential equations for the forward and backward wave amplitudes. An alternative expansion of the solution in terms of plane waves that are reflected by the local "tangent" planes results in stronger coupling between the component waves.

Iterative solutions of the coupled wave equations are considered in detail. The first-order iterations (coupling neglected) are identified as WKB-type solutions, and these are used to generate higher-order iterations which account for mode scattering.

As an illustrative example, we have considered the launching of a surface wave over an infinite wedge with a rounded corner in Sec. 4. Expressions are derived for both the transmitted surface wave and the scattered mode amplitudes.

2. FIELDS AROUND A CONVEX CYLINDRICAL BOUNDARY OF ARBITRARILY VARYING SURFACE IMPEDANCE AND CURVATURE EXPRESSED IN TERMS OF LOCAL CIRCULAR CYLINDRICAL MODES

In this section, we shall consider the problem of radiation from an electric line-source $\bar{J} = J(x, y)\bar{a}_z$ parallel to a convex, cylindrical boundary of finite, cross-sectional area. The boundaries of the diffracting object are characterized by an arbitrarily varying surface impedance $Z_s(x)$ and radius of curvature $R(x)$, where the variable x is the distance measured along the surface of the cross section (see Fig. 1). The special case of plane wave excitation can be obtained by allowing the electric line source to recede to infinity (Sec. 4). Thus, in this paper, we are restricting our attention to horizontally polarized waves (the electric field having only an E_z component). The solutions for vertically polarized waves may be derived in a similar manner by substituting a magnetic line-source excitation for the electric line-source excitation and by interchanging the electric and magnetic properties of the diffracting body.

In view of the geometry of the diffracting object, we construct a natural coordinate system (x, y, z) around the surface of the straight convex cylinder (see Fig. 1). The surfaces, $x = \text{const}$, are planes normal to the convex cylinder, and x is the distance measured along the boundary of the cylinder from the plane through the origin ($x = 0$) to the plane $x = \text{const}$. The orthogonal surfaces, $y = \text{const}$, are convex, cylindrical surfaces around the diffracting boundary, and y is the (normal) distance from the boundary ($y = 0$) to the $y = \text{const}$ surface. The $z = \text{const}$ surfaces are planes normal to the axis of the convex cylinder, and z is the distance from the plane through the origin

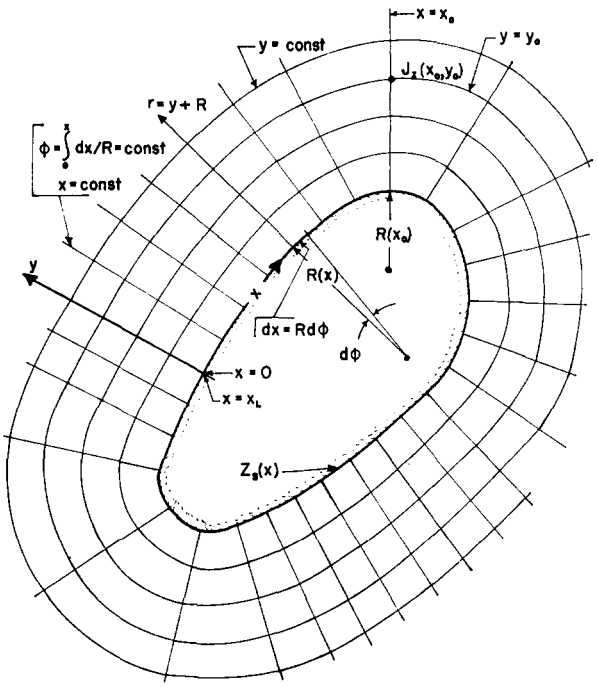

FIG. 1. Radiation by an electric line source parallel to a convex cylinder of arbitrary cross-section and variable surface impedance boundary.

$(z = 0)$ and the $z = \text{const}$ plane. Obviously, if the cylinder is of circular cross section of radius $R = \text{const}$, (x, y, z) is related to the cylindrical coordinates (r, ϕ, z') with $r = y + R$, $\phi = x/R$, and $z = z'$. In the general case, however, x and y are related to the "local," circular, cylindrical coordinates, r and ϕ,

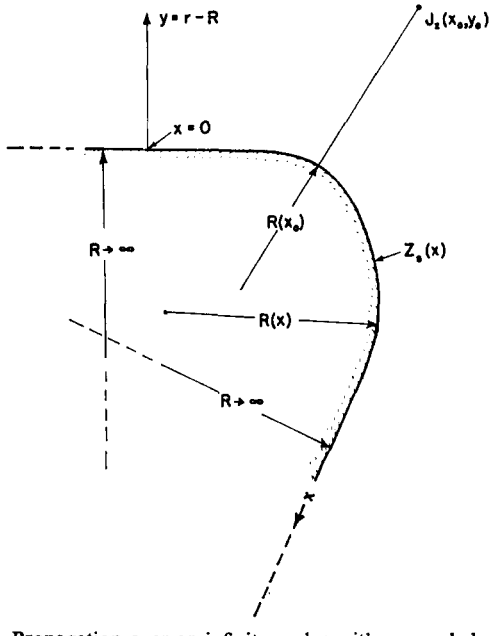

FIG. 2. Propagation over an infinite wedge with a rounded corner.

through the Jacobian of the transformation, J_T, where

$$J_T^{-1} = \begin{Vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{Vmatrix} = \begin{Vmatrix} \frac{dR}{dx} & 1 \\ \frac{1}{R} & 0 \end{Vmatrix} \tag{2.1a}$$

and

$$J_T = \begin{Vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{Vmatrix} = \begin{Vmatrix} 0 & R \\ 1 & -\frac{R\,dR}{dx} \end{Vmatrix} \tag{2.1b}$$

Thus, we have

$$r(x, y) = y + R(x), \tag{2.2a}$$

$$\phi(x) = \int_0^x \frac{dx}{R(x)}. \tag{2.2b}$$

Note, therefore, that the differential elements of length along the three coordinate lines through an arbitrary point (r, ϕ, z) are

$$dl_1 = h_1\,dr, \quad dl_2 = h_2\,d\phi, \quad \text{and} \quad dl_3 = h_3\,dz, \tag{2.3a}$$

where the metric parameters h_1, h_2, and h_3 are

$$h_1 = 1, \quad h_2 = r, \quad \text{and} \quad h_3 = 1. \tag{2.3b}$$

We now express Maxwell's curl equations in terms of the natural coordinate variables (r, ϕ, z) for an $\exp(i\omega t)$ time excitation. Assuming no field variations along the z axis, we see that these are

$$\frac{1}{r}\frac{\partial E_z}{\partial \phi} = -i\omega\mu H_r, \tag{2.4a}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) - \frac{1}{r}\frac{\partial H_r}{\partial \phi} = i\omega\epsilon E_z + J_z, \tag{2.4b}$$

and

$$\frac{-\partial E_z}{\partial r} = -\omega\mu H_\phi, \tag{2.4c}$$

in which H_r and H_ϕ are the nonvanishing components of the magnetic field and μ and ϵ are the permeability and permittivity of the medium of propagation, respectively. Equations (2.4) may be expressed in terms of the coordinate variables (x, r) through the transformations (2.1) and (2.2). Thus, on eliminating H_ϕ from (2.4b), using (2.4c), we obtain,

$$-\frac{\partial E_z}{\partial x} = i\omega\mu\frac{r}{R}H_r \tag{2.5a}$$

and

$$-\frac{\partial H_r}{\partial x} = i\omega\epsilon\left[\frac{r}{R}E_z + \frac{1}{k^2}\frac{\partial}{\partial r}\left(\frac{r}{R}\frac{\partial E_z}{\partial r}\right)\right] + \frac{r}{R}J_z, \tag{2.5b}$$

where $k = \omega(\mu\epsilon)^{\frac{1}{2}}$ is the wavenumber of the medium and

$$J_z = I\delta(r - r_0)\delta(\phi - \phi_0)/r = I\delta(r - r_0)\delta(x - x_0)R/r, \quad (2.5c)$$

in which $\delta(x - x_0)$, $\delta(r - r_0)$, and $\delta(\phi - \phi_0)$ are Dirac δ functions. Equations (2.5a) and (2.5b) involve only components of the electromagnetic field transverse to the x axis. For the special case $R = \text{const}$ and $Z_s = \text{const}$, E_z can be shown to satisfy the scalar wave equation

$$\nabla^2 E_z = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2} + k^2 E_z = i\omega\mu J_z. \quad (2.6)$$

A solution for E_z satisfying the above equation can be expressed in terms of a complete modal expansion. Thus,for $Z_s = 0$,[3,4]

$$E_z(\xi, \phi) = -\frac{\omega\mu\pi I}{4}\sum H_{\nu_n}^{(2)}(\xi_0)H_{\nu_n}^{(2)}(\xi) \times \left(\frac{H_\nu^{(1)}(\xi_R)}{\partial[H_\nu^{(2)}(\xi_R)]/\partial\nu}\right)_{\nu_n}\frac{\cos\nu_n(\phi - \phi_0 - \pi)}{\sin\nu_n\pi}, \quad (2.7a)$$

in which $H_{\nu_n}^{(1,2)}(\xi)$ are the Hankel functions of the first and second kind, respectively, and ξ and ξ_R are the dimensionless quantities

$$\xi = kr \quad \text{and} \quad \xi_R = kR. \quad (2.7b)$$

The electromagnetic fields satisfy the impedance boundary condition at $r = R$,

$$\frac{\partial E_z}{\partial(kr)} = \frac{i\eta E_z}{Z_s}, \quad (2.7c)$$

where η is the intrinsic impedance of the medium ($\eta = (\mu/\epsilon)^{\frac{1}{2}}$). Thus, the orders ν_n of the basis functions $H_{\nu_n}^{(2)}(\xi)$ satisfy the modal equation

$$H_{\nu_n}^{(2)'}(\xi_R) - iy_s H_{\nu_n}^{(2)}(\xi_R) = 0, \quad (2.7d)$$

in which y_s is the normalized surface admittance,

$$y_s = \eta/Z_s. \quad (2.7e)$$

We can also write the azimuthal dependence of the solution (2.7a) in terms of forward and backward wave amplitudes $a_n(\phi)$ and $b_n(\phi)$, respectively, as

$$a_n(\phi) + b_n(\phi) \propto \frac{\cos\nu_n(\phi - \phi_0 - \pi)}{\sin\nu_n\pi} = 2i(e^{-i\nu_n(\phi-\phi_0)} + e^{i\nu_n(\phi-\phi_0-2\pi)})\sum_{p=0}^{\infty} e^{-ip2\pi\nu_n}. \quad (2.8)$$

For $p = 0$, the first term of the sum is the forward traveling or direct wave from the source to the observation point while the second term is the contribution from the backward traveling wave (propagating in the direction of decreasing ϕ). For $p = 1, 2, 3, \cdots$, the terms constitute forward and backward creeping waves which propagate around the cylinder p times in the forward and backward directions.[5]

For the general case in which the local radius of curvature and surface impedance are functions of x (see Fig. 1), it is not possible to derive a solution in the separable form given in (27). However, in view of the above discussion, we express the solution in terms of the complete set of basis functions $H_{\nu_n}^{(2)}(\xi)$ whose terms individually satisfy the local boundary conditions.[4] Thus, these local basis functions which satisfy the modal equation (2.7d) for all values of x are not only functions of r but also functions of x [through $R(x)$ and $y_s(x)$]. Therefore, we express the desired solution for E_z as

$$E_z(\xi, x) = \sum_{n=1}^{\infty} e_n(x)\frac{H_{\nu_n}^{(2)}(\xi)}{N_n} \equiv \sum_n [a_n(x) + b_n(x)]\frac{H_{\nu_n}^{(2)}(\xi)}{N_n}. \quad (2.9a)$$

Using the orthogonal properties of the basis functions, we recognize the electric field nth-mode amplitude $e_n(x)$ to be the transform of the function $E_z(\xi, x)$, i.e.,

$$e_n(x) = \int_{\xi_R}^{\infty} E_z(\xi, x)\left(\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}\right)\nu_n\frac{d\xi}{\xi}. \quad (2.9b)$$

In the expressions above, which are recognized to be the Watson transform pair,[4] the normalization coefficients N_n and M_n must be chosen such that

$$M_n N_n = \frac{\xi_R}{2}\left[H_\nu^{(2)'}(\xi)\frac{\partial}{\partial\nu}[H_\nu^{(2)}(\xi)] - H_\nu^{(2)}(\xi)\frac{\partial}{\partial\nu}[H_\nu^{(2)'}(\xi)]\right]_{\substack{\xi=\xi_R \\ \nu=\nu_n}} = -\frac{\xi_R}{2}H_{\nu_n}^{(2)}(\xi_R)\frac{\partial}{\partial\nu}[H_\nu^{(2)'}(\xi_R) - iy_s H_\nu^{(2)}(\xi_R)]_{\nu=\nu_n}. \quad (2.9c)$$

It is useful to retain this freedom in the choice of the normalization coefficients at this phase of our analysis.

In a similar manner, we define the Watson transform pair related to the magnetic field component H_r. We note that, for the case $R = \text{const}$ and $Z_s = \text{const}$, the appropriate basis function for H_r is

$Y_n(\xi)H^{(2)}_{\nu_n}(\xi)$, where $Y_{\nu_n}(\xi)$ is the nth-mode transverse wave admittance

$$Y_n(\xi) = \nu_n/\xi\eta. \tag{2.10a}$$

Thus, we can write, for the magnetic field component,

$$H_r(\xi, x) = \sum_{n=1}^{\infty} h_n(x) \frac{Y_n(\xi)H^{(2)}_{\nu_n}(\xi)}{N_n} \equiv \sum_{n=1}^{\infty} [a_n(x) - b_n(x)] \frac{Y_n(\xi)H^{(2)}_{\nu_n}(\xi)}{N_n}, \tag{2.10b}$$

in which the magnetic field mode amplitude $h_n(x)$ [the transform function of $H_r(\xi, x)$] is

$$h_n(x) = \eta \int_{\xi_R}^{\infty} H_r(\xi, x)H^{(2)}_{\nu_n}(\xi) \frac{d\xi}{M_n}. \tag{2.10c}$$

We are now left with the problem of determining the solutions of the electric and magnetic field mode amplitudes $e_n(x)$ and $h_n(x)$ which are defined, respectively, as the sum and difference of the forward and backward wave amplitudes $a_n(x)$ and $b_n(x)$. To this end, we substitute the complete modal expansions (2.9) and (2.10) for E_z and H_r into the differential equations (2.5). Multiply (2.5a) by

$$H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n} \frac{d\xi}{\xi}$$

and integrate with respect to ξ between the limits $[\xi_R, \infty]$. Thus, using the orthogonal properties of the basis functions, we get

$$-\int_{\xi_R}^{\infty} \frac{\partial E_z}{\partial x} H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n} \frac{d\xi}{\xi} = i \frac{\nu_n}{R} \eta \int_{\xi_R}^{\infty} H_r H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n} = i \frac{\nu_n}{R} h_n(x). \tag{2.11a}$$

Instead of term-by-term differentiation of E_z [(2.9)] with respect to x, we express the first term in (2.11a) as

$$-\int_{\xi_R}^{\infty} \frac{\partial E_z}{\partial x} H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n} \frac{d\xi}{\xi} = -\frac{d}{dx} \int_{\xi_R}^{\infty} E_z H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n} \frac{d\xi}{\xi} + \int_{\xi_R}^{\infty} E_z \frac{\partial}{\partial x} \left(H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n \xi} \right) d\xi = -\frac{de_n(x)}{dx} + \sum_m C_{nm} e_m, \tag{2.11b}$$

where the coupling coefficient C_{nm} is defined as

$$C_{nm} = \int_{\xi_R}^{\infty} \frac{\partial}{\partial x} \left(H^{(2)}_{\nu_n}(\xi) \frac{\nu_n}{M_n \xi} \right) H^{(2)}_{\nu_m}(\xi) \frac{d\xi}{N_m} \tag{2.11c}$$

and we have taken note that both ξ and ξ_R are functions of x. Hence, (2.11a) reduces to

$$\frac{-de_n(x)}{dx} = i \frac{\nu_n}{R} h_n(x) - \sum_m C_{nm} e_m(x). \tag{2.11d}$$

Multiplying (2.5b) by $\eta H^{(2)}_{\nu_n}(\xi)\, d\xi/M_n$ and integrating with respect to ξ, we obtain

$$-\eta \int_{\xi_R}^{\infty} \frac{\partial H_r}{\partial x} \frac{H^{(2)}_{\nu_n}}{M_n} d\xi = ik \int_{\xi_R}^{\infty} \left[\frac{\xi}{\xi_R} E_z + \frac{\partial}{\partial \xi} \left(\frac{\xi}{\xi_R} \frac{\partial E_z}{\partial \xi} \right) \right] H^{(2)}_{\nu_n}(\xi) \frac{d\xi}{M_n} + \frac{k\eta}{M_n} \int_{\xi_R}^{\infty} I\delta(x - x_0)\delta(\xi - \xi_0) H^{(2)}_{\nu_n}(\xi)\, d\xi. \tag{2.12a}$$

We treat the first term in (2.12a) as the first term in (2.11a). Thus, it follows that

$$-\eta \int_{\xi_R}^{\infty} \frac{\partial H_r}{\partial x} \frac{H^{(2)}_{\nu_n}(\xi)}{M_n} d\xi = -\frac{dh_n(x)}{dx} + \sum_m D_{nm} h_m(x), \tag{2.12b}$$

where

$$D_{nm} \equiv \int_{\xi_R}^{\infty} \frac{\partial}{\partial x} \left(\frac{H^{(2)}_{\nu_n}(\xi)}{M_n} \right) H^{(2)}_{\nu_m}(\xi) \frac{\nu_m}{N_m} \frac{d\xi}{\xi}. \tag{2.12c}$$

Instead of term-by-term differentiation of E_z with respect to ξ in (2.12a), we integrate by parts (the application of Green's theorem in one dimension). Thus, it can be shown that

$$\int_{\xi_R}^{\infty} \frac{\partial}{\partial \xi} \left(\frac{\xi}{\xi_R} \frac{\partial E_z}{\partial \xi} \right) H^{(2)}_{\nu_n}(\xi) \frac{d\xi}{M_n} = \int_{\xi_R}^{\infty} E_z \frac{\partial}{\partial \xi} \left[\frac{\xi}{\xi_R} \frac{\partial}{\partial \xi} \left(\frac{H^{(2)}_{\nu_n}(\xi)}{M_n} \right) \right] d\xi - \left[\frac{\partial E_z}{\partial \xi} \frac{H^{(2)}_{\nu_n}(\xi)}{M_n} - E_z \frac{\partial}{\partial \xi} \left(\frac{H^{(2)}_{\nu_n}(\xi)}{M_n} \right) \right]_{\xi_R}. \tag{2.12d}$$

Using the boundary condition (2.7c) and the modal equation (2.7d), we can show that the last term in (2.12d) vanishes. Noting that $H^{(2)}_{\nu_n}(\xi)$ satisfies Bessel's differential equation, we reduce the right-hand term in (2.12a) to

$$ik \int_{\xi_R}^{\infty} E_z \frac{\xi}{\xi_R} \left(\frac{\nu_n}{\xi} \right)^2 H^{(2)}_{\nu_n}(\xi) \frac{d\xi}{M_n} = i \frac{\nu_n}{R} e_n(x). \tag{2.12e}$$

Thus (2.12a) can now be written as

$$\frac{-dh_n(x)}{dx} = i \frac{\nu_n}{R} e_n(x) - \sum_m D_{nm} h_n(x) + 2J_0 \delta(x - x_0), \tag{2.12f}$$

where

$$J_0 \equiv k\eta I H^{(2)}_{\nu_n}(\xi_0)/2M_n. \tag{2.12g}$$

The relationship between the coupling coefficients C_{nm} and D_{nm} is now derived. From the normalization condition, it follows that

$$\frac{\partial}{\partial x}\int_{\xi_R}^{\infty}\left(\frac{H_{\nu_m}^{(2)}(\xi)}{N_m}\right)\left(\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}\right)\frac{\nu_n}{\xi}\,d\xi = 0$$
$$= \int_{\xi_R}^{\infty}\frac{\partial}{\partial x}\left[\left(\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}\right)\frac{\nu_n}{\xi}\right]\left(\frac{H_{\nu_m}^{(2)}(\xi)}{N_m}\right)d\xi$$
$$+\int_{\xi_R}^{\infty}\frac{\partial}{\partial x}\left(\frac{H_{\nu_m}^{(2)}(\xi)}{N_m}\right)\left(\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}\right)\frac{\nu_n}{\xi}\,d\xi. \quad (2.13a)$$

Thus, after some manipulation, it can be shown that

$$C_{nm} + \frac{M_m N_n}{M_n N_m} D_{mn} + \delta_{nm}\frac{d}{dx}\ln\left(\frac{M_m}{N_m}\right) = 0, \quad (2.13b)$$

where δ_{nm} is the Kronecker δ function. For the particular choice of the normalization constants $N_m = M_m$, (2.13b) reduces to

$$C_{nm} = -D_{mn}. \quad (2.13c)$$

While we have avoided term-by-term differentiation of the infinite series expansions for E_z and H_r, it is obvious from the above analysis that this would be permissible here since the local basis functions individually satisfy the varying boundary conditions. Using the relationships between the electric and magnetic field mode amplitudes, $e_n(x)$ and $h_n(x)$, respectively, and the forward and backward wave amplitudes $a_n(x)$ and $b_n(x)$ [(2.9a) and (2.10b)], respectively, we obtain the following, coupled, first-order, differential equations:

$$\frac{-da_n(x)}{dx} - i\frac{\nu_n}{R}a_n(x)$$
$$= \sum_{n=1}^{\infty}\frac{dT_{nm}}{dx}a_m(x) + \frac{dR_{nm}}{dx}b_m(x) + J_0\delta(x - x_0) \quad (2.14a)$$

and

$$\frac{-db_n(x)}{dx} + i\frac{\nu_n}{R}b_n(x)$$
$$= \sum_{n=1}^{\infty}\frac{dT_{nm}}{dx}b_m(x) + \frac{dR_{nm}}{dx}a_m(x) - J_0\delta(x - x_0), \quad (2.14b)$$

where the transmission and reflection scattering coefficients dT_{nm}/dx and dR_{nm}/dx are defined as

$$\frac{dT_{nm}}{dx} = -\tfrac{1}{2}(C_{nm} + D_{nm})$$

and

$$\frac{dR_{nm}}{dx} = -\tfrac{1}{2}(C_{nm} - D_{nm}). \quad (2.14c)$$

Thus, for the particular choice $N_m = M_n$, (2.14c) reduces to

$$\frac{dT_{nm}}{dx} = \tfrac{1}{2}(D_{mn} - D_{nm})$$

and

$$\frac{dR_{nm}}{dx} = \tfrac{1}{2}(D_{mn} + D_{nm}). \quad (2.14d)$$

From the above expression, it is clear that $T_{nn} = 0$. On considering solutions to (2.14b), we shall discuss the effects of the above choice of the normalization coefficients. In the following section, we derive explicit forms for the coupling coefficients.

3. EVALUATION OF THE SCATTERING COEFFICIENTS AND THEIR DEPENDENCE UPON THE DERIVATIVES OF THE LOCAL RADIUS OF CURVATURE AND SURFACE IMPEDANCE

It is clear that the scattering coefficients dT_{nm}/dx and dR_{nm}/dx are related to variations in both the local radius of curvature $R(x)$ and the local surface impedance $Z_s(x)$. It is convenient to separate the effects of the curvature variations from the effects of variations in the surface impedance. Thus we write

$$\frac{dT_{nm}}{dx} = \frac{\partial T_{nm}}{\partial y_s}\frac{dy_s}{dx} + \frac{\partial T_{nm}}{\partial R}\frac{dR}{dx} \quad (3.1a)$$

and

$$\frac{dR_{nm}}{dx} = \frac{\partial R_{nm}}{\partial y_s}\frac{dy_s}{dx} + \frac{\partial R_{nm}}{\partial R}\frac{dR}{dx}. \quad (3.1b)$$

To determine $(\partial T_{nm}/\partial y_s)(dy_s/dx)$ and $(\partial R_{nm}/\partial y_s) \times (dy_s/dx)$ [(3.1)], it is sufficient to evaluate D_{nm} [(2.12c)] for the case $R = \text{const}$ since the coefficients C_{nm} are shown to be related to D_{nm} [(2.13)]. Thus, noting the orthogonality relationship between the basis functions,

$$\delta_{nm} = \int_{\xi_R}^{\infty}\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}H_{\nu_m}^{(2)}(\xi)\frac{\nu_m}{N_m}\frac{d\xi}{\xi}$$
$$= \left[\frac{\nu_m\xi}{M_nN_m(\nu_m^2 - \nu_n^2)}(H_{\nu_n}^{(2)}(\xi)H_{\nu_m}^{(2)'}(\xi) - H_{\nu_m}^{(2)}(\xi)H_{\nu_n}^{(2)'}(\xi))\right]_{\xi_R}^{\infty}, \quad (3.2)$$

we obtain, for $n \neq m$,

$$D_{nm} = \left[\frac{\nu_m\xi}{M_nN_m(\nu_m^2 - \nu_n^2)}\left(\frac{\partial}{\partial\nu}H_{\nu}^{(2)'}(\xi)H_{\nu_m}^{(2)}(\xi) - H_{\nu_m}^{(2)'}(\xi)\frac{\partial}{\partial\nu}(H_{\nu}^{(2)}(\xi))\right)\frac{d\nu}{dx}\right]_{\nu_n}^{\xi_R}$$
$$= \left[\frac{\nu_m\xi H_{\nu_m}^{(2)}(\xi)}{M_nN_m(\nu_m^2 - \nu_n^2)}\frac{\partial}{\partial\nu} \times [H_{\nu}^{(2)'}(\xi) - iy_sH_{\nu}^{(2)}(\xi)]\frac{d\nu}{dx}\right]_{\nu_n}^{\xi_R}. \quad (3.3a)$$

In (3.3) we have used the modal equation (2.7d) for the nth mode. Furthermore, we note that, since the modal equation is satisfied for all x,

$$\left[\frac{\partial}{\partial \nu}\left(H_\nu^{(2)'}(\xi_R) - iy_s H_\nu^{(2)}(\xi_R)\right)\frac{d\nu}{dx}\right]_{\nu=\nu_n} - iH_{\nu_n}^{(2)}(\xi_R)\frac{dy_s}{dx} = 0. \quad (3.3b)$$

Using the relationship (3.3b), between $d\nu_n/dx$ and dy_s/dx, we reduce (3.3a) to

$$D_{nm} = \frac{i\nu_m \xi_R}{\nu_m^2 - \nu_n^2} H_{\nu_m}^{(2)}(\xi_R) H_{\nu_n}^{(2)}(\xi_R)\frac{dy_s}{dx}(M_n N_m)^{-1}. \quad (3.3c)$$

Thus, using (2.13b) for $n \neq m$, we obtain, for the transmission scattering coefficient,

$$\left.\frac{dT_{nm}}{dx}\right|_{R=\text{const}} = \frac{\frac{1}{2}i}{\nu_n - \nu_m}[\xi H_{\nu_m}^{(2)}(\xi) H_{\nu_n}^{(2)}(\xi)]_{\xi_R}\frac{dy_s}{dx}(M_n N_m)^{-1}, \quad n \neq m. \quad (3.4a)$$

For the choice $N_m = 1$, $m = 1, 2, \cdots$, (3.4a) reduces to

$$\left.\frac{dT_{nm}}{dx}\right|_{R=\text{const}} = \frac{-i\, dy_s/dx}{\nu_n - \nu_m} H_{\nu_m}^{(2)}(\xi)\left(\frac{\partial}{\partial \nu}[H_\nu^{(2)'}(\xi_R) - iy_s H_\nu^{(2)}(\xi_R)]_{\nu_n}\right)^{-1} = -\frac{d\nu_n/dx}{\nu_n - \nu_m}\frac{H_{\nu_m}^{(2)}(\xi_R)}{H_{\nu_n}^{(2)}(\xi_R)}, \quad n \neq m. \quad (3.4b)$$

In a recent paper, Wait[5] uses the Wiener–Hopf technique to derive an exact solution to the problem of propagation across a circular cylindrical surface (R = const) with an abrupt, step-type change in the value of the surface impedance. Setting $(Z_{s1} - Z_{s2})/\eta = dz_s$ in the solution by Wait and noting that he has considered the dual problem of propagation by a magnetic line source (vertically polarized waves), we see that the above expression for dT_{nm} [(3.4b)] corresponds precisely to the value for the transmission coefficient T_{nm} derived by Wait. Similarly, it can be shown that

$$\left.\frac{dR_{nm}}{dx}\right|_{R=\text{const}} = \frac{\nu_n - \nu_m}{\nu_n + \nu_m}\left.\frac{dT_{nm}}{dx}\right|_{R=\text{const}}, \quad n \neq m. \quad (3.5)$$

The above expression, for the reflection scattering coefficient, also corresponds to the value of the reflection coefficient R_{nm} derived by Wait.[5] It also follows directly from (3.5) that, for the case

$$|\nu_n + \nu_m| \gg |\nu_n - \nu_m|, \quad (3.6)$$

we may neglect the reflected waves. Using (2.13b) and (2.14c), we can show that

$$\left.\frac{dT_{nn}}{dx}\right|_{R=\text{const}} = \frac{1}{2}\frac{d}{dx}\ln\left(\frac{M_n}{N_n}\right) \quad (3.7a)$$

and

$$\left.\frac{dR_{nn}}{dx}\right|_{R=\text{const}} = -\frac{1}{2\nu_n}\frac{d\nu_n}{dx} = -\frac{1}{2Y_n}\frac{d}{dx}Y_n. \quad (3.7b)$$

Thus, $dT_{nn}/dx = 0$, if the normalization coefficients are chosen such that $M_n = N_n$. In the next section it will be shown that the term dT_{nn}/dx constitutes a WKB-type, wave amplitude, modification factor. The reflection scattering coefficient dR_{nn}/dx corresponds to the (dominant-mode) reflection coefficient in transmission line theory.

We now consider, in detail, the scattering due to variations in the curvature only. Thus, we need to evaluate D_{nm} (2.12c) for the case y_s = const. We note, in this case, that, since both ν_n and ξ are functions of x,

$$\frac{d}{dx}H_{\nu_n}^{(2)}(\xi) = \left[\frac{\partial}{\partial \nu}H_\nu^{(2)}(\xi)\frac{d\nu}{dx}\right]_{\nu_n} + \frac{\partial}{\partial \xi}H_{\nu_n}^{(2)}(\xi)k\frac{dR}{dx}. \quad (3.8)$$

On substituting (3.8) into (2.12c), D_{nm}, $n \neq m$, can be expressed as a sum of two terms Q_{nm} and S_{nm}. Thus, using (3.3a), we obtain

$$\begin{aligned} Q_{nm} &= \int_{\xi_R}^{\infty}\left[\frac{\partial}{\partial \nu}[H_\nu^{(2)}(\xi)]\frac{d\nu}{dx}\right]_{\nu_n} H_{\nu_m}^{(2)}(\xi)\frac{\nu_m}{M_n N_m}\frac{d\xi}{\xi} \\ &= \left[\frac{\nu_m \xi H_{\nu_m}^{(2)}(\xi)}{M_n N_m(\nu_m^2 - \nu_n^2)}\frac{\partial}{\partial \nu}[H_\nu^{(2)'}(\xi) - iy_s H_\nu^{(2)}(\xi)]\frac{d\nu}{dx}\right]_{\nu_n}^{\xi_R} \\ &= \frac{-\nu_m \xi_R H_{\nu_m}^{(2)}(\xi_R)}{M_n N_m(\nu_m^2 - \nu_n^2)}\left[\frac{\partial}{\partial \xi}[H_{\nu_n}^{(2)'}(\xi) - y_s H_{\nu_n}^{(2)}(\xi)]k\frac{dR}{dx}\right]_{\xi_R}. \end{aligned} \quad (3.9)$$

In (3.9), we have noted that since the total x-derivative of the modal equation vanishes,

$$\left[\frac{\partial}{\partial \nu}[H_\nu^{(2)'}(\xi_R) - iy_s H_\nu^{(2)}(\xi_R)]\frac{d\nu}{dx}\right]_{\nu=\nu_n} + \left[\frac{\partial}{\partial \xi}[H_{\nu_n}^{(2)'}(\xi) - iy_s H_{\nu_n}^{(2)}(\xi)]k\frac{dR}{dx}\right]_{\xi_R} = 0. \quad (3.10)$$

To evaluate S_{nm}, we make use of the relationships between the Bessel functions and their derivatives.[6] Thus,

$$S_{nm} = \int_{\xi_R}^{\infty} \frac{\partial}{\partial \xi}[H_{\nu_n}^{(2)}(\xi)]k\frac{dR}{dx}H_{\nu_m}^{(2)}(\xi)\frac{\nu_m}{M_nN_m}\frac{d\xi}{\xi}$$
$$= \frac{\nu_m \xi_R H_{\nu_m}^{(2)}(\xi_R)}{2M_nN_m}\left[\frac{\partial}{\partial\xi}[H_{\nu_n}^{(2)'}(\xi) - iy_sH_{\nu_n}^{(2)}(\xi)]k\frac{dR}{dx}\right]_{\xi_R}$$
$$\times\left(\frac{1}{\nu_m^2-(\nu_n+1)^2}+\frac{1}{\nu_m^2-(\nu_n-1)^2}\right)$$
$$+\frac{\nu_m\nu_nH_{\nu_m}^{(2)}(\xi_R)H_{\nu_n}^{(2)}(\xi_R)}{2M_nN_m\xi_R}$$
$$\times\left(\frac{1}{\nu_m^2-(\nu_n+1)^2}-\frac{1}{\nu_m^2-(\nu_n-1)^2}\right)k\frac{dR}{dx}. \tag{3.11}$$

The explicit expression for D_{nm} is, therefore,

$$D_{nm} = \frac{\nu_m\xi_RH_{\nu_m}^{(2)}(\xi_R)}{M_nN_m}\left[\frac{\partial}{\partial\xi}[H_{\nu}^{(2)'}(\xi) - iy_sH_{\nu_n}^{(2)}(\xi)]k\frac{dR}{dx}\right]_{\xi_R}$$
$$\times\left(\frac{\nu_m^2+3\nu_n^2-1}{(\nu_m^2-\nu_n^2)[\nu_m^2-(\nu_n+1)^2][\nu_m^2-(\nu_n-1)^2]}\right)$$
$$+\frac{H_{\nu_m}^{(2)}(\xi_R)H_{\nu_n}^{(2)}(\xi_R)}{M_nN_m\xi_R}$$
$$\times\frac{2\nu_m\nu_n^2k\,dR/dx}{[\nu_m^2-(\nu_n+1)^2][\nu_m^2-(\nu_n-1)^2]},$$
$$n\neq m. \tag{3.12}$$

The WKB-type Debye–Watson expansions for the Hankel functions of large argument are[7,3]

$$H_{\nu_n}^{(1,2)}(\xi) = \left(\frac{2}{\pi\xi C_n(\xi)}\right)^{\frac{1}{2}}\exp(\mp\tfrac{1}{4}i\pi)$$
$$\times\exp\left(\pm i\int_{\xi_0}^{\xi}C_n(u)\,du\right), \tag{3.13a}$$

where

$$C_n(\xi) = [1-(\xi_0/\xi)^2]^{\frac{1}{2}}$$

and

$$\xi_0 = [\nu^2-\tfrac{1}{4}]^{\frac{1}{2}} \equiv \xi_RS_n(\xi_R),\quad \operatorname{Im}C_n \leq 0. \tag{3.13b}$$

Thus, the expressions for the scattering coefficients can be shown to be proportional to $(dR/dx)/R^3$ for $kR \gg 1$. This factor clearly indicates the advantages of using the Bessel function expansion for the electromagnetic fields, (2.9) and (2.10), around the diffracting boundary. A plane wave expansion of the fields around the diffracting boundary will result in relatively stronger coupling between the component waves. Note that, even for $R = \text{const}$, these plane waves will be coupled. A related comparison of the solution, expressed in terms of Bessel function and plane wave expansions, has been carried out in detail for propagation in waveguide bends.[8]

For the case $n = m$, we evaluate the expression for D_{nn} by noting that

$$\frac{d}{dx}\int_{\xi_R}^{\infty}\frac{H_{\nu_n}^{(2)}(\xi)}{M_n}\frac{H_{\nu_n}^{(2)}(\xi)}{N_n}\frac{\nu_n}{\xi}d\xi = 0$$
$$= 2D_{nn} + \frac{1}{\nu_n}\frac{d\nu_n}{dx} - k\frac{dR}{dx}\int_{\xi_R}^{\infty}\frac{H_{\nu_n}^{(2)}(\xi)H_{\nu_n}^{(2)}(\xi)\nu_n}{M_nN_n\xi^2}d\xi$$
$$+\frac{d}{dx}\ln\left(\frac{M_n}{N_n}\right). \tag{3.14}$$

Thus,

$$D_{nn} = -\frac{1}{2}\left\{\frac{1}{\nu_n}\frac{d\nu_n}{dx} - \frac{2\nu_nk\,dR/dx}{(4\nu_n^2-1)M_nN_n}\xi_RH_{\nu_n}^{(2)}(\xi_R)\right.$$
$$\times\left[\frac{\partial}{\partial\xi}[H_{\nu_n}^{(2)'}(\xi) - iy_sH_{\nu}^{(2)}(\xi)]\right]_{\xi_R}$$
$$+\frac{\nu_n/\xi_R}{4\nu_n^2-1}[H_{\nu_n}^{(2)}(\xi_R)]^2\frac{k\,dR/dx}{M_nN_n}$$
$$\left.+\frac{d}{dx}\ln\left(\frac{M_n}{N_n}\right)\right\}. \tag{3.15a}$$

Thus, for the choice $M_n = N_n$,

$$D_{nn} = -\frac{1}{2}\left(\frac{1}{\nu_n}\frac{d\nu_n}{dx}+\frac{\nu_nk\,dR/dx}{\nu_n^2-\frac{1}{4}}\right.$$
$$\left.+\frac{\nu_n/\xi_R}{4\nu_n^2-1}[H_{\nu_n}^{(2)}(\xi_R)]^2\frac{k\,dR/dx}{M_nN_n}\right). \tag{3.15b}$$

Substituting the WKB expansion of the Hankel function [(3.13)] in (3.10), we can show that

$$\frac{d\nu_n}{dx} \approx -k\frac{dR}{dx}, \tag{3.16}$$

and it follows from (3.15) that D_{nn} is also proportional to $(dR/dx)R^{-3}$, for $kR \gg 1$. The explicit forms for the transmission and reflection scattering coefficients for $y_s = \text{const}$ [(3.1)] can be obtained by substituting (3.15) into (2.14) and using the relationship between C_{nm} and D_{nm} [(2.13)]. Here, too, we see that $dT_{nn}/dx = 0$ for the choice $M_m = N_m$.

4. SOLUTIONS TO THE COUPLED, FIRST-ORDER, INHOMOGENEOUS, DIFFERENTIAL EQUATIONS FOR THE WAVE AMPLITUDES

In this section, we consider, in detail, solutions to the coupled, first-order, differential equations (2.14) for the forward and backward wave amplitudes $a_n(x)$ and $b_n(x)$, respectively. In view of the terms $\pm J_0\delta(x-x_0)$ appearing in these equations, we can show by direct integration about the immediate neighborhood of the point $x = x_0$ (i.e., $x_0^- \leq x \leq x_0^+$)

that the wave amplitudes must satisfy the periodic boundary conditions

$$[a_n(x)]_{x=x_0+x_L}^{x=x_0} = -J_0 \tag{4.1a}$$

and

$$[b_n(x)]_{x=x_0+x_L}^{x=x_0} = J_0. \tag{4.1b}$$

Various numerical methods for solving coupled, first-order, differential equations are presented in the literature, and the homogeneous forms of Eqs. (2.14) are often encountered in nonuniform waveguide problems.[8–10] In this paper, we shall adopt an iterative approach since it affords further insight into the solution of this problem. Furthermore, as a result of the local Bessel function expansion of the solution, the coupling coefficients between the component waves are shown to be proportional to $(dR/dx)R^{-3}$. Thus, for structures with large radii of curvature, $R(x)$, the iterative procedure would be very efficient.

Neglecting mode conversion, we reduce (2.14) to

$$-\frac{da_n^1}{dx} - \left(\frac{dT_{nn}}{dx} + i\nu_n\right)a_n^1 = J_0\delta(x - x_0) \tag{4.2a}$$

and

$$-\frac{db_n^1}{dx} - \left(\frac{dT_{nn}}{dx} - i\nu_n\right)b_n^1 = -J_0\delta(x - x_0), \tag{4.2b}$$

where the superscript 1 denotes first-order solutions, dT_{nm}/dx is assumed to vanish for $m \neq n$, and dR_{nm}/dx is neglected for all n. The expression for dT_{nn}/dx is

$$\frac{dT_{nn}}{dx} = \frac{1}{2}\frac{d}{dx}\ln\left(\frac{M_n}{N_n}\right). \tag{4.2c}$$

Thus, using (4.1), we get, after some manipulation, for $x_0 \leq x \leq x_0 + x_L$,

$$a_n^1(x) = \frac{iJ_0}{2}\left(\frac{M_n(x_0)N_n(x)}{M_n(x)N_n(x_0)}\right)^{\frac{1}{2}} \times \exp\left[-i\left(\int_{x_0}^{x}\nu_n\frac{du}{R} - \frac{1}{2}\int_0^{x_L}\nu_n\frac{du}{R}\right)\right] \times \left[\sin\left(\frac{1}{2}\int_0^{x_L}\nu_n\frac{du}{R}\right)\right]^{-1} \tag{4.3a}$$

and

$$b_n^1(x) = \frac{iJ_0}{2}\left(\frac{M_n(x_0)N_n(x)}{M_n(x)N_n(x_0)}\right)^{\frac{1}{2}} \times \exp\left[i\left(\int_{x_0}^{x}\nu_n\frac{du}{R} - \frac{1}{2}\int_0^{x_L}\nu_n\frac{du}{R}\right)\right] \times \left[\sin\left(\frac{1}{2}\int_0^{x_L}\nu_n\frac{du}{R}\right)\right]^{-1}. \tag{4.3b}$$

In the above expressions, we recall that both ν_n and R are functions of the variable of integration, u. Note that, for the case $R = \text{const}$ and $z_s \to 0$, the expression for the electric field mode amplitude [(2.9a)] is

$$e_n(x) = iJ_0\cos\nu_n(\phi - \phi_0 - \pi)/\sin\nu_n\pi. \tag{4.4}$$

For $R = \text{const}$ and $z_s \to 0$, the appropriate expression for the product of the normalization coefficients M_nN_n [(2.9)] is

$$M_nN_n = \frac{\xi_R}{2}H_{\nu_n}^{(2)'}(\xi_R)\frac{\partial}{\partial\nu}[H_\nu^{(2)}(\xi_R)]_{\nu_n} = -\frac{2i}{\pi}\frac{\partial}{\partial\nu}[H_\nu^{(2)}(\xi_R)]_{\nu_n}[H_{\nu_n}^{(1)}(\xi_R)]^{-1}, \tag{4.5a}$$

in which we have used the value of the Wronskian,

$$W[H_\nu^{(1)}(\xi), H_\nu^{(2)}(\xi)] = -4i/\pi\xi. \tag{4.5b}$$

Thus, substituting (4.4) into (2.9a) and making use of (4.5a) and the definition of J_0 [(2.12g)], we see that the solution for $E_z(\xi, x)$ [(2.9)] reduces to the expression (2.7a), as must be the case. We note that the expression for the forward mode amplitude $a_n(x)$ [(4.3a)] constitutes not only the direct wave propagating in the positive x direction but also the creeping waves which propagate around the convex cylinder p times, $p = 1, 2, 3, \cdots$. To show this explicitly, it is necessary to expand the sine function in the denominator of (4.3) in an infinite geometric series. For the case $kR \gg 1$ and for lossy surfaces [$\text{Re}\,(Z_s) > 0$], however, we may retain only the direct waves. Thus, Eqs. (4.3) reduce to

$$a_n^1(x) \approx -J_0\left(\frac{M_n(x_0)N_n(x)}{M_n(x)N_n(x_0)}\right)^{\frac{1}{2}}\exp\left(-i\int_{x_0}^{x}\nu_n\frac{du}{R}\right) \tag{4.6a}$$

and

$$b_n^1(x) = -J_0\left(\frac{M_n(x_0)N_n(x)}{M_n(x)N_n(x_0)}\right)^{\frac{1}{2}} \times \exp\left(i\int_{x_0}^{x}\nu_n\frac{du}{R} - \int_0^{x_L}\nu_n\frac{du}{R}\right). \tag{4.6b}$$

We have noted earlier in this paper that, for the choice $M_n = N_n$, the expression for dT_{nn}/dx vanishes and the square root coefficient in (4.6) reduces to unity. For $M_n \neq N_n$, this square root term adjusts the amplitude of the nth-node amplitude such that the energy contained in nth mode is constant. To show this, we note that

$$\int_{\xi_R}^{\infty} e_n(x)\frac{H_{\nu_n}^{(2)}(\xi)}{N_n}h_n(x)\frac{H_{\nu_n}^{(2)}(\xi)}{N_n} = e_n(x)h_n(x)\frac{M_n}{N_n}. \tag{4.7}$$

It is for this reason that we have identified the term dT_{nn}/dx with the WKB-type amplitude factor.

Similarly, the argument of the exponential functions in (4.3) and (4.6) are identified with the WKB-type phase integral or the phase memory concept.[11,3]

The first-order, WKB-type solutions [(4.3) and (4.6)] for the forward and backward wave amplitudes may be substituted in the terms on the right-hand side of (2.14) to obtain second-order solutions for the coupled wave amplitudes. This procedure, while accounting for mode scattering, neglects reconversion. Therefore, we write (2.14a) as

$$-\frac{da_n^2}{dx}-\left(\frac{dT_{nn}}{dx}+\frac{iv_n}{R}\right)a_n^2=J_0\delta(x-x_0)+f_1(x), \tag{4.8a}$$

where the superscript 2 denotes second-order solutions and

$$f_1(x)=\sum_{\substack{n=1\\ n\neq m}}^{\infty}\frac{dT_{nn}}{dx}\,a_n^1(x)+\sum_{n=1}^{\infty}\frac{dR_{nn}}{dx}\,b_n^1(x). \tag{4.8b}$$

Thus, on retaining the direct waves only, the solution to (4.8b) is[12]

$$a_n^2(x)=a_n^1(x)-\int_{x_0}^{x}\exp\left(\int_u^x\gamma_n(v)\,dv\right)f_1(u)\,du, \tag{4.9a}$$

where

$$\gamma_n(x)=\frac{dT_{nn}}{dx}+\frac{iv_n}{R} \tag{4.9b}$$

is the modified propagation coefficient and $a_n^1(x)$ is given by (4.6). If we cannot neglect the creeping waves, $p=1, 2, 3, \cdots$, we substitute for $a_n^1(x)$, in (4.9), the expression (4.3a) times a constant coefficient. This constant is determined by applying the periodic boundary condition (4.1a). The procedure above may be applied to determine the reflected wave amplitudes $b_n(x)$ and repeated to obtain higher-order iterations.

For cylindrical structures of infinite cross section, such as wedges with rounded corners (Fig. 2), the same iterative procedure outlined above may be used. In these cases, however, the boundary conditions are

$$[a_n(x)]_{x=x_0^-}^{x=x_0^+}=-J_0,\quad a_n(-\infty)=0 \tag{4.10a}$$

and

$$[b_n(x)]_{x=x_0^-}^{x=x_0^+}=J_0,\qquad b_n(\infty)=0. \tag{4.10b}$$

Since in this case we have no creeping waves to account for, the first-order WKB solutions are given by (4.6).

An important factor that must still be considered in this paper is the appropriate forms of the expansions for the electromagnetic fields in regions where $R\to\infty$.

In these regions, the Watson transforms (2.9) and (2.10) do not exist. It has been shown that the discrete Watson transform is directly related to the continuous modal expansion[4]

$$E_z(\xi,x)=\tfrac{1}{4}\int_{-\infty}^{\infty}e(v,x)\psi_v(\xi)M_v\,dv, \tag{4.11a}$$

where the path of integration L is along the real axis (see Fig. 3) and the basis function $\psi_v(\xi)$ is

$$\psi_v(\xi)=H_v^{(1)}(\xi)+R_vH_v^{(2)}(\xi). \tag{4.11b}$$

It follows from the boundary condition (2.7c) that

$$R_v=-\frac{H_v^{(1)\prime}(\xi_R)-iy_sH_v^{(1)}(\xi_R)}{H_v^{(2)\prime}(\xi_R)-iy_sH_v^{(2)}(\xi_R)}. \tag{4.11c}$$

The transform function $e(v,\phi)$ is defined, as in (2.9b), as

$$e(v,x)=\int_{\xi_R}^{\infty}E_z(\xi,x)\frac{H_v^{(2)}(\xi)}{M_v}\frac{v}{\xi}\,d\xi, \tag{4.11d}$$

where the normalization coefficient M_v (independent of ξ) may be chosen arbitrarily. For $R\to\infty$, it is not possible in the above continuous spectral expansion to choose M_v equal to the square root of (2.9c) as for the case when R is finite. In this case ($R\to\infty$), it is convenient to choose the normalization coefficient to be

$$M_v=\frac{v}{\xi_R}\frac{H_v^{(2)\prime}(\xi_R)-iy_sH_v^{(2)}(\xi_R)}{-iy_s}. \tag{4.12a}$$

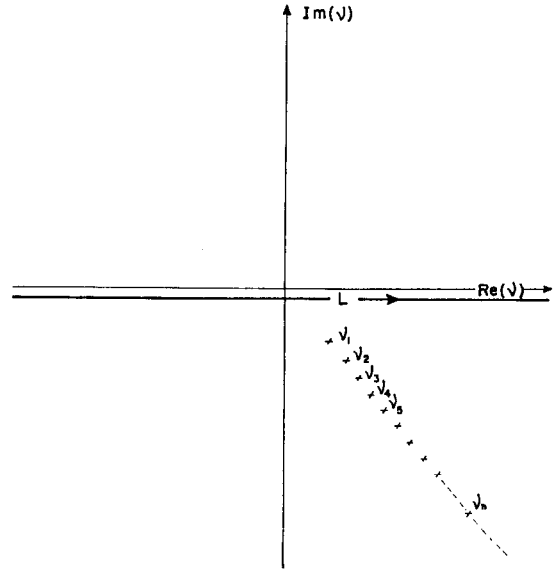

FIG. 3. Integration path in the complex ν plane.

Thus, using the WKB-type Debye–Watson expansion for the Hankel functions (3.13), we can show that

$$\lim_{\xi_R\to\infty}[\psi_\nu(\xi)M_\nu] \to \left(\frac{2S}{\pi\xi_R C}\right)\left(\frac{C}{y_s}+1\right)[e^{ikCy}+R(C)e^{-ikCy}] \quad (4.12b)$$

and

$$\lim_{\xi_R\to\infty}\frac{H_{\nu_n}^{(2)}(\xi)}{M_\nu}\frac{\nu}{\xi_R}\to\frac{e^{-ikCy}}{C/y_s+1}, \quad (4.12c)$$

where

$$C=(1-S^2)^{\frac{1}{2}},\quad \operatorname{Im}C<0,\quad S^2=(\nu^2-\tfrac{1}{4})/\xi^2,$$

and $R(C)$ is the reflection coefficient for horizontally polarized waves over a plane surface:

$$R(C)=\frac{(C/y_s-1)}{(C/y_s+1)}. \quad (4.12d)$$

Thus, the corresponding form for the transform pair (4.11a) and (4.11d) for $R\to\infty$ can be shown to be[4]

$$E_z(y,x)=\frac{1}{2\pi}\int_L e(S,x)\left[\left(\frac{C}{y_s}+1\right)e^{ikCy}+\left(\frac{C}{y_s}-1\right)e^{-ikCy}\right]\frac{S\,dS}{C} \quad (4.13a)$$

and

$$e(S,x)=\int_0^\infty E_z(y,x)\frac{e^{-ikCy}}{C/y_s+1}\,d(ky), \quad (4.13b)$$

where the path of integration L is along the real S axis (Fig. 4). Deforming the path of integration L to the path D_1+D_2 around the branch cut and accounting for the surface wave pole $C=C_0=-y_s$, we can write (4.13) as

$$E_z(y,x)=\frac{1}{2\pi}\int_0^\infty E(C,x)\left[\left(\frac{C}{y_s}+1\right)e^{ikCy}+\left(\frac{C}{y_s}-1\right)e^{-ikCy}\right]dC + E(C_0,x)e^{-ikC_0y}\frac{2}{i} \quad (4.14a)$$

and

$$E(C,x)\equiv e(C,x)-e(-C,x)=\int_0^\infty E_z(\xi,x)\times\frac{[(C/y_s+1)e^{ikCy}+(C/y_s-1)e^{-ikCy}]}{(C/y_s+1)(C/y_s-1)}\,d(ky), \quad (4.14b)$$

for the continuous spectrum along the positive real C axis and for the surface wave mode

$$E(C_0,x)\equiv\lim_{C\to C_0}[(C+y_s)e(C,x)]=y_s\int_0^\infty E_z(\xi,x)e^{-ikC_0y}d(ky)=\lim_{\nu\to\nu_0}e(\nu_0,x)\frac{iH_{\nu_0}^{(2)\prime}(\xi_R)+y_sH_{\nu_0}^{(2)}(\xi_R)}{H_{\nu_0}^{(2)}(\xi_R)}, \quad (4.14c)$$

where the subscript 0 is used to identify the quantities associated with the surface wave. Thus, we have derived the direct relationship between the mode amplitudes for the Watson transform $e_n(x)$ and the mode amplitudes (continuous and discrete) for the Fourier-type transform $E(C,x)$ and $E(C_0,x)$. As an illustrative example, consider the launching of a surface wave in the region $x<0$ ($R\to\infty$; see Fig. 4):

$$E(y,x)=E_0\exp[-ik(S_0x+C_0y)], \quad (4.15a)$$

where E_0 is a constant. The WKB-type expression for the electric field [(2.9)] in the region of the bend where R is finite can be shown to be

$$E(\xi,x)=\frac{a_0(x)H_{\nu_0}^{(2)}(\xi)}{N_0(x)}=E_0\left(\frac{iS_0}{2y_s}\right)^{\frac{1}{2}}H_{\nu_0}^{(2)}(\xi)\times\exp\left(-i\int_0^x\frac{\nu_0}{R}\,du\right)[M_0(x)N_0(x)]^{-\frac{1}{2}}, \quad (4.15b)$$

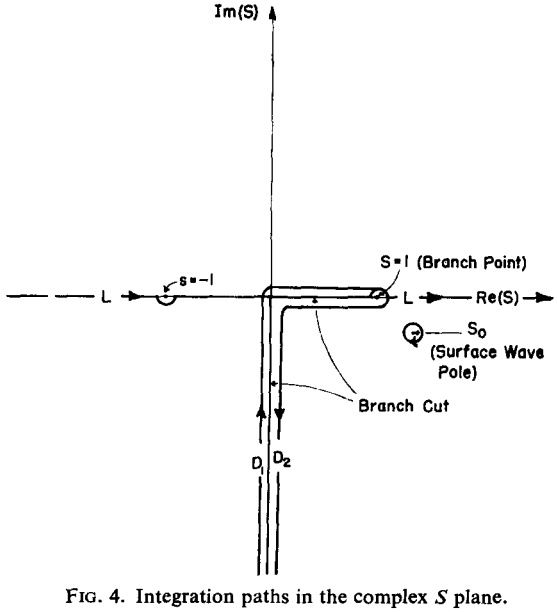

FIG. 4. Integration paths in the complex S plane.

where the product of the normalization coefficients, (2.9c), is

$$M_0(x)N_0(x) = -\frac{\xi_R}{2} H_{\nu_0}^{(2)}(\xi_R) \frac{\partial}{\partial \nu} [H_\nu^{(2)\prime}(\xi_R) - iy_s H_\nu^{(2)}(\xi_R)]_{\nu_0}, \quad (4.15c)$$

and the Debye–Watson expansion of the Hankel functions is used to show that

$$\lim_{x\to 0} [M_0(x)N_0(x)]^{\frac{1}{2}} \to \left(\frac{iS_0}{2y_s}\right)^{\frac{1}{2}} \lim_{x\to 0} [H_{\nu_0}^{(2)}(\xi_R)]. \quad (4.15d)$$

To derive the expressions for the higher-order modes, we substitute the expression in (4.15b) for $a_0(x)$ into the right-hand side of (2.14). Thus, we obtain

$$-\frac{da_n}{dx} - \left(\frac{dT_{nn}}{dx} + i\frac{\nu_n}{R}\right) a_n = \frac{dT_{n0}}{dx} a_0 \quad (4.16a)$$

and

$$-\frac{db_n}{dx} - \left(\frac{dT_{nn}}{dx} - i\frac{\nu_n}{R}\right) b_n = \frac{dR_{n0}}{dx} a_0. \quad (4.16b)$$

The solutions for the higher-order modes are, therefore,

$$a_n(x) = -\int_0^x \frac{dT_{n0}}{du} a_0(u) \times \exp\left[-\int_u^x \left(\frac{dT_{nn}}{dv} + i\frac{\nu_n}{R}\right) dv\right] du \quad (4.17a)$$

and

$$b_n(x) = \int_x^\infty \frac{dR_{n0}}{du} a_0(u) \times \exp\left[-\int_u^x \left(\frac{dT_{nn}}{dv} - i\frac{\nu_n}{R}\right) dv\right] du. \quad (4.17b)$$

A uniform, plane wave, excitation can be considered in a straightforward manner by noting that an electric line source of magnitude

$$I \sim e^{+ikr_0}(r_0)^{\frac{1}{2}}, \quad (4.18)$$

located at $r = r_0$ and $x = x_0$ (Fig. 1), will generate a uniform plane wave at the surface of the cylinder if we let $r_0 \to \infty$.

5. CONCLUDING REMARKS

The electromagnetic fields around a convex cylindrical boundary have been expressed in terms of a complete set of local cylindrical modes. Through the use of a generalized Bessel transform we have shown that the Watson transform for regions with finite radii of curvature merge with the plane wave expansion appropriate over plane boundaries ($R \to \infty$). It has also been shown[4] that, for $R \to 0$, the generalized Bessel transform reduces to the Kontorowich–Lebedev transform.[1] Thus, in our analysis, no restrictions are made on the local radius of curvature of the boundary.

The expansion of the fields in terms of local cylindrical modes reduces Maxwell's equations into a set of first-order, coupled, differential equations. Using the explicit forms of the scattering coefficients, we have shown that the coupling is proportional to $(dR/dx)R^{-3}$ for $R \to \infty$.

An alternative expansion of the solution in terms of local plane waves results in stronger coupling between the component waves. In the local plane wave expansion, we must, of course, characterize the boundary by the reflection coefficients for the local tangent planes.

Although in this paper, we have restricted our attention to 2-dimensional problems, this analysis may be applied to 3-dimensional problems, provided that the following conditions are satisfied[13]: (a) The major contributions to the received fields are from the region along the straight line path between the source and the receiver; (b) the variations of the boundary curvature and surface impedance, transverse to the path of propagation, are small compared to the variations along the path.

ACKNOWLEDGMENTS

The author wishes to thank F. Ullman for his comments and J. R. Wait for stimulating discussions. The manuscript was prepared by Mrs. M. Alles. This work was supported in part by a grant from the University of Nebraska Engineering Research Center.

[1] M. J. Kontorowich and N. N. Lebedev, J. Phys. **1**, 229 (1939).
[2] E. Bahar, Radio Sci. **5**, 1069 (1970).
[3] J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, Oxford, 1962).
[4] E. Bahar, J. Math. Phys. **12**, 179 (1971).
[5] J. R. Wait, J. Math. Phys. **11**, 2851 (1970).
[6] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1952).
[7] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[8] E. Bahar, IEEE Trans. Microwave Theory Techniques **17**, 210 (1969).
[9] S. A. Schelkunoff, Bell System Tech. J. **34**, 995 (1955).
[10] E. Bahar and G. Crain, Proc. IEE **115**, 1395 (1968).
[11] K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge U.P., London, 1961).
[12] E. Bahar, Radio Sci. **2**, 287 (1967).
[13] J. R. Wait, *Electromagnetic Surface Waves*, Advances in Radio Research, Vol. 1 (Academic, New York, 1964), pp. 157–217.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Recursive Method for the Computation of the SO_n, $SO_{n,1}$, and ISO_n Representation Matrix Elements

KURT BERNARDO WOLF
Institute of Theoretical Physics, Goteborg, Sweden

(Received 15 May 1970)

We find a procedure whereby the matrix elements of the finite $SO_{n,1}$ transformations (principal series) can be expressed as a single integral, over a compact domain, of two matrix elements of the SO_n subgroup and a multiplier. In this way we automatically obtain their classification by the canonical chain $SO_{n,1} \supset SO_n \supset \cdots \supset SO_2$. Analytic continuation yields the SO_{n+1} matrix elements in a recursive form. We obtain the asymptotic behavior of the boost matrix elements. The Inönü–Wigner contraction yields the ISO_n representation matrix elements classified by the chain $ISO_n \supset SO_n \supset \cdots \supset SO_2$.

1. INTRODUCTION

The unitary irreducible representation (UIR) matrix elements of the unimodular orthogonal (SO_n), pseudo-orthogonal ($SO_{n,1}$) and inhomogeneous orthogonal (ISO_n) groups have been a fertile field of research due to their repeated appearance in mathematical physics: For SO_2, they are the partial waves of a periodic function; for SO_3, they are the $D^J_{mm'}(\alpha, \beta, \gamma)$ functions. These and the Wigner $d^J_{mm'}(\beta)$ functions[1] have been so extensively used in angular momentum theory that no further remark is needed.

Bargmann's $d^l_{mm'}(\zeta)$ functions for[2] $SO_{2,1}$ have been used in Toller's cross-channel partial wave expansion.[3,4] The SO_4 d matrices[5] were used by Freedman and Wang in order to find the quantum numbers of the daughter Regge poles which belong to a given Toller pole. This, plus the important high-energy behavior of the corresponding scattering amplitude, were found by Sciarrino and Toller[4] using the $SO_{3,1}$ boost matrix elements $d^{M\lambda}_{mjj'}(\zeta)$.[6,7]

Going further, the $SO_{4,1}$ UIR matrix elements have also been calculated.[8–10] In particular, Ström[9] performed the contraction[11] $SO_{4,1} \to ISO_{3,1}$, whereby the D-matrix elements classified by the canonical chain become the matrix elements of Poincaré transformations[10,12] in the chain of subgroups which includes the homogeneous Lorentz group. The matrix elements of[13] SO_5 representations have found applications in nuclear physics,[14] and the theory of master analytic representations[15] has given a method of reaching higher groups.

The importance of the matrix elements of the general SO_n, $SO_{n,1}$, and ISO_n UIR's lies presently in mathematical physics: As group representations, they constitute an orthogonal and complete set of functions[16] on the group manifold, and any well-behaved, square-integrable function on the group can be expanded in terms of them.[16–18]

Thus far, however, they have remained as "certain" functions, some of whose relevant properties were known, but for which one could not write explicit expressions. The reason for this is not difficult to see: The straightforward procedure of obtaining them as eigenfunctions of the set of Casimir operators of the group and its subgroups involves setting up a set of simultaneous differential equations which, together with difference and recursion relations,[1,9,19] gives rise to rather involved expressions which are still under investigation[20] for SO_n and $SO_{n-1,1}$, $n > 5$.

Bargmann's[2] and Toller's[4] work, however, did not involve the solution of differential equations, but rather an integration over the compact subgroup. This was reduced further to a single integral, which has been successfully performed. In this article we set up a procedure which generalizes the above two cases. We shall work, however, only with the component of the group connected to the identity. We thus disregard the parity indices in the UIR labels.

In Sec. 2 we remind the reader how a complete and orthogonal set of functions on a homogeneous space X can be used to set up a multiplier representation of a group G whose action on X is known. The space X is here the SO_n group manifold. The properties and labels of a complete and orthogonal set of functions, the UIR matrix elements for SO_n classified by the canonical[21] chain, are reviewed in Sec. 3. The group G which acts on this space may be, however, larger then SO_n.

In Sec. 4, using a generalization of what is known in the literature as the Gell-Mann operator,[22] we can apply $G = SO_{n,1}$ in such a way that, while the transformations in the SO_n subgroup give rise to "rigid" mappings of the X manifold, the boosts in $SO_{n,1}$ generated by the Gell-Mann operator "deform" X.

In Sec. 5, the complete and orthogonal set of functions over SO_n introduced in Sec. 3 is used to set up a multiplier representation. The matrix elements

of the Gell-Mann operator, proportional to the generalized Wigner $d^J_{LML'}(\zeta)$ functions for $SO_{n,1}$, are thus expressed as an integral over the SO_n subgroup (which is reduced to a single integral over one angle) of two UIR matrix elements of SO_n (simplified to the Wigner d functions for SO_n) and a multiplier.

The asymptotic behavior of the $SO_{n,1}$ d functions as $\zeta \to \infty$ and the contraction[11] $SO_{n,1} \to ISO_n$ can be seen already from the integral form. In fact, from the contraction of $SO_{n,1}$ we obtain the UIR matrix elements of ISO_n classified by the chain $ISO_n \supset SO_n \supset \cdots \supset SO_2$.

The geometrical meaning of the deformation effected on SO_n by the generators built through the Gell-Mann operator is shown, in Appendix A, to be but the natural action of the group $SO_{n,1}$ (in its Iwasawa decomposition $G = KAN$) on itself, modulo AN. A useful integral is calculated in Appendix B.

We want to emphasize that in our procedure

(a) the UIR matrix elements are classified by the canonical chain,

(b) several key properties are apparent from the integral form,

(c) the integration is performed over a compact domain and can be expressed in terms of a sum of products of trigonometric and hypergeometric functions.

We can point also to the possibilities of extending this method, taking a complete and orthogonal set of functions over other groups or homogeneous spaces—noncompact ones, for instance—and considering multiplier representations of a larger group of deformations of it, thus obtaining expressions for the representation matrix elements of noncompact groups classified by chains which can thus include noncompact subgroups.[4,23]

2. MULTIPLIER REPRESENTATIONS

In order to fix our notation, we shall make some well-known definitions.

Let X be a homogeneous space under the group of transformations G, and put $x_1, x_2, \cdots \in X$. A set of functions $\{\phi_n(x)\}$, $n \in N$, discrete, is *orthogonal* on X if

$$\int_X d\mu(x)\overline{\phi_n(x)}\phi_{n'}(x) = \delta_N(n, n'), \qquad (2.1)$$

where $d\mu(x)$ is an appropriate measure on X and $\delta_N(n, n') = 0$ for $n \neq n'$ and will be detailed below.

Furthermore, the set $\{\phi_n(x)\}$ is *complete* on X if

$$\sum_{n\in N} \omega(n)\overline{\phi_n(x_1)}\phi_n(x_2) = \delta_X(x_1, x_2), \qquad (2.2)$$

where $\omega(n)$ is the Plancherel weight on N; $\delta_X(x_1, x_2) = 0$ for $x_1 \neq x_2$ and is normalized in such a way that the integral (2.1) (which is a sum, if X is discrete) fulfills

$$\int_X d\mu(x_1)f(x_1)\delta_X(x_1, x_2) = f(x_2), \qquad (2.3)$$

for any continuous test function f over X. The normalization of (2.1) and (2.2) can be arranged to be such that[17]

$$\sum_{n\in N} \omega(n)\tilde{f}_n\delta_N(n, n') = \tilde{f}_{n'}, \qquad (2.4)$$

and hence $\delta_N(n, n') = [\omega(n)]^{-1}\delta_{n,n'}$.

Any well-behaved function f over X can be expanded in the complete and orthogonal set $\{\phi_n(x)\}$ as

$$f(x) = \sum_{n\in N} \omega(n)\tilde{f}_n\phi_n(x), \qquad (2.5)$$

where $f_n = (\phi_n, f)_X$ is the scalar product between two functions on X, defined as

$$(f, f')_X = \int_X d\mu(x)\overline{f(x)}f'(x) = \sum_{n\in N} \omega(n)\bar{\tilde{f}}_n\tilde{f}'_n = (\tilde{f}, \tilde{f}')_N. \qquad (2.6)$$

The action of G on X, $x \xrightarrow{g} x'(x, g)$, is assumed to be defined such that

$$x'(x''(x, g_1), g_2) = x'(x, g_1g_2) \quad \text{and} \quad x'(x, e) = x$$

for the unit e of the group. When $X = G$, this is satisfied if either $x'(x, g) = xg$ or $x'(x, g) = g^{-1}x$, but may be of a more general nature when $X \neq G$.

The action of G on $f(x)$ is defined through

$$f(x) \xrightarrow{g} U^{(\lambda)}(g)f(x) = M^{(\lambda)}(x, g)f(x'(x, g)), \qquad (2.7)$$

where the multiplier[2] $M^{(\lambda)}(x, g)$ satisfies

$$M^{(\lambda)}(x, g_1)M^{(\lambda)}(x'(x, g_1), g_2) = M^{(\lambda)}(x, g_1g_2)$$

and $M^{(\lambda)}(x, e) = 1$ and does not vanish over $X \times G$.

A multiplier can be written in the form[2,16]

$$M^{(\lambda)}(x, g) = [\rho(x)/\rho(x'(x, g))]^{\lambda}, \qquad (2.8)$$

where $\rho(x)$ is some function over X.

The requirement of *unitarity* of the representation

$$(U^{(\lambda)}(g)f, U^{(\lambda)}(g)f) = (f, f')$$

implies, through (2.6) and (2.7),

$$\frac{d\mu(x'(x, g))}{d\mu(x)} = |M^{(\lambda)}(x, g)|^2, \qquad (2.9)$$

if we restrict the form of the multiplier and the possible values of λ in (2.8). In particular, if $X = G$ and $d\mu(x)$ is the Haar measure, the ratio (2.9) is unity and the multiplier may only be a phase.

We can construct a matrix representation of G as

$$D^{\lambda}_{nn'}(g) = [\omega(n)\omega(n')]^{\frac{3}{2}}(\phi_n, U^{(\lambda)}(g)\phi_{n'}), \quad (2.10)$$

where the rows and columns are labeled by the (discrete) index $n \in N$. We can check through (2.2) that (2.10) follows the group multiplication law and that

$$D^{\lambda}_{nn'}(e) = \delta_{n,n'}$$

while, if (2.9) is satisfied, the representation (2.10) is unitary, i.e.,

$$D^{\lambda}_{nn'}(g^{-1}) = \overline{D^{\lambda}_{n'n}(g)}.$$

At this stage, however, we cannot make any statement as to the irreducibility of (2.10) nor as to whether we can find all unitary representations in this way.

Next, we want to express the transformation (2.7) as generated by a Lie algebra of operators.[2] Assume $g(\zeta)$ belongs to a 1-dimensional subgroup of G parametrized by a variable ζ, whose generator is $N^{(\lambda)}$, i.e.,

$$U^{(\lambda)}(g(\zeta))f(x) = \exp(\zeta N^{(\lambda)})f(x). \quad (2.11)$$

The differential form of $N^{(\lambda)}$ is thus

$$N^{(\lambda)}f(x) = \frac{d}{d\zeta}[M^{(\lambda)}(x, g(\zeta))f(x'[x, g(\zeta)])]_{\zeta=0}. \quad (2.12)$$

When the multiplier $M^{(\lambda)}(x, g)$ is taken in its form (2.8), it is straightforward to see that $N^{(\lambda)}$ can be written as

$$N^{(\lambda)}f(x) = (N^{(0)} - \lambda Q)f(x), \quad (2.13)$$

where $N^{(0)}$ is the generator of the vector representation

$$\exp(\zeta N^{(0)})f(x) = f(x'[x, g(\zeta)]) \quad (2.14)$$

and

$$Q = [N^{(0)}\rho(x)]/\rho(x). \quad (2.15)$$

3. THE ORTHOGONAL GROUP

The n-dimensional (unimodular) orthogonal group SO_n has been extensively studied.[21] We need, however, a brief survey of its properties in order to define the problem.

We introduce the "Euler-angle" parameters[17,18,24] in SO_n (enclosing collective variables in braces) by

$$R_n(\{\theta\}^{(n)}) = R_{n-1}(\{\theta\}^{(n-1)})H_n(\{\theta^{(n)}\}),$$
$$H_n(\{\theta^{(n)}\}) = \nu_{n-1,n}(\theta^{(n)}_{n-1,n}) \times \cdots \times \nu_{23}(\theta^{(n)}_{23})\nu_{12}(\theta^{(n)}_{12}), \quad (3.1)$$

where $R_k \in SO_k$ and $r_{pq}(\theta)$ is a rotation by θ in the (p, q) plane; the ranges are $0 \leq \theta_{12} < 2\pi$ and $0 \leq \theta_{k-1,k} \leq \pi$, $k = 3, 4, \cdots, n$. In this way, we express the SO_n manifold as the product of the SO_{n-1} manifold with [the $(n-1)$-dimensional surface of] the n-dimensional sphere S_n. Notice that $R_3(\alpha, \beta, \gamma) = r_{12}(\alpha)r_{23}(\beta)r_{12}(\gamma)$ differs from the more general usage which writes $r_{13}(\beta)$ as the middle factor. This will cause no inconvenience, however.

The Haar measure can now be split according to (3.1) as $dR_n = dR_{n-1}\, dH_n$, where

$$dH_n = \sin^{n-2}\theta_{n-1,n}\, d\theta_{n-1,n}\, dH_{n-1}, \quad dH_2 = d\theta_{12}. \quad (3.2)$$

From (3.2) and the ranges specified above, it can be seen that the volume of SO_n is $\mathrm{vol}\, SO_n = \mathrm{vol}\, SO_{n-1}S_n$, where $S_n = 2\pi^{\frac{1}{2}n}/\Gamma(\frac{1}{2}n)$, and $\mathrm{vol}\, SO_2 = 2\pi$.

The basis vectors for the unitary irreducible representations of SO_n, classified by the canonical chain of subgroups $SO_n \supset SO_{n-1} \supset \cdots \supset SO_2$, are labeled with the Gel'fand–Tsetlin[21] kets

$$\left|\begin{array}{ccccc} J_{n,1} & J_{n,2} & \cdots & J_{n,[n/2]} \\ J_{n-1,1} & J_{n-1,2} & \cdots & J_{n-1,[(n-1)/2]} \\ \cdots & \cdots & \cdots & \cdots \\ J_{4,1} & J_{4,2} & & \\ J_{3,1} & & & \\ J_{2,1} & & & \end{array}\right\rangle, \quad (3.3)$$

where $[k/2]$ is the largest integer smaller or equal to $k/2$. This ket transforms as the $J_k \equiv \{J_{k,1}, J_{k,2}, \cdots, J_{k,[k/2]}\}$ UIR of SO_k, $k = 2, \cdots, n$, while the representation row is labeled $\overline{J_{k-1}} \equiv \{J_{k-1}, J_{k-2}, \cdots, J_2\}$. For the single-valued UIR's of SO_n, all J_{ab} are integers constrained by the "zigzag" inequalities

$$\begin{array}{ccccccccc} J_{k,1} & & & & & & & & \\ \vee & & & & & & & & \\ J_{k-1,1} & \geq & J_{k,2} & & & & & & \\ & & \vee & & & & & & \\ & & J_{k-1,2} & \geq & J_{k,3} & & & & \\ & & & & \vee & & & & \\ & & & & J_{k-1,3} & \geq & \cdot & & \\ & & & & & & \cdot & & \\ & & & & & & & \cdot & \\ & & & & & & & & \cdot \end{array}, \quad k = 3, \cdots, n, \quad (3.4a)$$

which end, to the right of (3.3), as

$$\begin{array}{ccccccc} \cdot & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \geq & J_{k,[k/2]-1} & & \\ & & \cdot & & & & \\ & & & \cdot & \vee & & \\ & & & & \cdot & & \\ & & & & J_{k-1,[(k-1)/2]-1} & \geq & J_{k,[k/2]} \\ & & & & & & \vee \\ & & & & & & |J_{k-1,[(k-1)/2]}|, \end{array} \quad k \text{ odd}, \quad (3.4b)$$

or

$$\begin{array}{ccc} \cdot & & \\ & \cdot & \\ & & \cdot \\ & & J_{k,[k/2]-1} \\ & \cdot & \\ & \cdot & \text{VI} \\ & & \cdot \\ & & J_{k-1,[(k-1)/2]} \geq |J_{k,[k/2]}|, \quad k \text{ even.} \end{array} \tag{3.4c}$$

In order to economize subindices, we will agree on the following notation: Let J (resp. L and M) stand for J_n (resp. J_{n-1} and J_{n-2}), the UIR label of $G = SO_n$ (resp. $H = SO_{n-1}$ and $K = SO_{n-2}$). The row labels are $\bar{L} = \bar{J}_{n-1}$ (resp. $\bar{M} = \bar{J}_{n-2}$ and $\bar{N} = \bar{J}_{n-3}$), and hence $\bar{L} = \{L, \bar{M}\}$ and $\bar{M} = \{M, \bar{N}\}$; $\dim_n J$ (resp. $\dim_{n-1} L$ and $\dim_{n-2} M$) denote the dimension of the UIR. The scalar representation of SO_k, $k = 2, \cdots, n$, is $J_k = 0 = \{0, \cdots, 0\}$.

The representation D-matrices for SO_n are defined as

$$D^J_{\bar{L},\bar{L}'}(R_n(\{\theta\}^{(n)})) = \langle J\bar{L}|\, R_n(\{\theta\}^{(n)})\, |J\bar{L}'\rangle, \tag{3.5}$$

where we have written the ket and bra (3.3) horizontally. The generalized Wigner d matrices (to be calculated in Sec. 5) are defined through

$$d^J_{LML'}(\theta) = \langle JLM\bar{N}|\, v_{n-1,n}(\theta)\, |JL'M\bar{N}\rangle \tag{3.6a}$$

and are seen to be diagonal in M, the representation label of K, and independent of its row-label $\bar{N}$, since $r_{n-1,n}(\theta)$ commutes with all transformations in K. Similarly,

$$D^J_{L\bar{M},L'\bar{M}'}(R_{n-1}) = \delta_{L,L'} D^L_{\bar{M},\bar{M}'}(R_{n-1}) \tag{3.7}$$

is independent of J, the UIR label of G, and diagonal in that of H.

In particular, for SO_2 the $d^J_{LML'}(\theta)$ are $d^J(\theta) = e^{iJ\theta}$ i.e., the indices L, M, and L' are absent; for SO_3, the d matrices are $d^J_{LL'}(\theta)$, the usual Wigner d matrices[1] for rotations around the x axis. For SO_n, $n \geq 4$, we have the general expression (3.6).

Equations (3.1), (3.5), (3.6a), and (3.7) allow us to write (omitting arguments in an obvious way)

$$D^J_{\bar{L},\bar{L}'}(R_n) = \sum_{\bar{M}''} D^L_{\bar{M},\bar{M}''}(R_{n-1}) E^J_{L\bar{M}'',L'\bar{M}'}(H_n), \tag{3.8}$$

where

$$E^J_{L\bar{M}'',L'\bar{M}'}(H_n) = d^J_{LM''L'}(\theta^{(n)}_{n-1,n}) E^{L'}_{M''\bar{N}'',M'\bar{N}'}(H_{n-1}) \tag{3.9}$$

and $E^L(H_2(\theta)) = d^L(\theta)$. Thus we see that the D-matrix elements (3.8) can be expressed in terms of the Wigner d-matrix elements (3.6). Only the latter need therefore be calculated explicitly.

Most of the interesting properties of the D and d matrices can be found before their explicit calculation. Chief among them are the orthogonality and completeness relations (2.1) and (2.2), which read[16,17]

$$\int_G dR_n \overline{D^J_{L_1L_2}(R_n)} D^{J'}_{L_1'L_2'}(R_n) = \delta_{L_1,L_1'}\delta_{L_2,L_2'}\delta_{J,J'} \frac{\text{vol } G}{\dim_{n'} J}, \tag{3.10}$$

$$\sum_J \frac{\dim_n J}{\text{vol } G} \sum_{L,L'} \overline{D^J_{L,L'}(R_n)} D^J_{L,L'}(R'_n) = \delta_G(R_n, R_{n'}), \tag{3.11}$$

where $\delta_{L,L'}$, etc., stand for a product of Kronecker δ's in the individual indices. The Plancherel weight is seen to be $w(J) = \dim_n J/\text{vol } G$, while the role of the index n in (2.1) is taken by the triad of collective indices $(J, \bar{L}, \bar{L}')$.

From (3.8) and the splitting of the Haar measure, we can find the "orthogonality" relations for the E matrices as

$$\int_{S_n} dH_n \sum_{\bar{M}_1} \overline{E^J_{L_1\bar{M}_1,L_2\bar{M}_2}(H_n)} E_{L_1\bar{M}_1,L_2'\bar{M}_2'}(H_n) = \delta_{L_2,L_2'}\delta_{\bar{M}_2,\bar{M}_2'}\delta_{J,J'} \frac{\text{vol } G}{\text{vol } H} \frac{\dim_{n-1}(L_1)}{\dim_n(J)}, \tag{3.12}$$

while (3.2), (3.9), and (3.12) yield, for the d matrices,

$$\int_0^\pi \sin^{n-2}\theta\, d\theta \sum_M \frac{\dim_{n-2} M}{\text{vol } K} \overline{d^J_{L_1ML_2}(\theta)}\, d^{J'}_{L_1ML_2}(\theta) = \delta_{J,J'} \frac{\dim_{n-1} \bar{L}_1 \dim_{n-2} \bar{L}_2}{\dim_n J} \frac{\text{vol } G}{(\text{vol } H)^2}. \tag{3.13}$$

Thus, while for SO_2 we have in $\{d^J(\theta)\}$ a complete and orthogonal set of functions, for SO_3 the Wigner $\{d^J_{LL'}(\theta)\}$ constitute an orthogonal set in the index J. The set is complete for $L = L' = 0$.[17,18] For SO_n, $n \geq 4$, the general result is (3.13), and this includes the sum over M-labels. Indeed, it is not difficult to show that $\{E_{0\bar{0}}, J_{L\bar{M}}(H_n)\}$ is an orthogonal and complete set of functions on $SO_{n-1}\backslash SO_n$ and the same result holds for $\{d^J_{000}(\theta)\}$ on $SO_{n-1}\backslash SO_n/SO_{n-1}$.[17]

We shall use (3.10), (3.11), and (3.13) in order to build the D matrices (2.10) after we have defined, in the next section, the group of transformations we wish to represent.

The parametrization of $R^P_n \in SO_{n-1,1}$ follows the definitions (3.1), (3.8), and (3.9) with

$$R^P_n = R_{n-1}H^P_n = R_{n-1}b_{n-1,n}(\zeta)H_{n-1}, \tag{3.14}$$

where $b_{n-1,n}(\zeta)$ is a boost in the $(n-1)$th direction through a hyperbolic angle ζ, $0 \leq \zeta < \infty$. The metric tensor has nonzero components $g_{11} = \cdots = g_{n-1,n-1} = -g_{nn} = 1$, and the $SO_{n-1,1}$ manifold is expressed as the product of the SO_{n-1} manifold and [the $(n-1)$-dimensional surface of] the n-dimensional hyperboloid. The Haar measure is $dR_n^P = dR_{n-1}\, dH_n^P$, where

$$dH_n^P = \sinh^{n-2}\zeta\, d\zeta\, dH_{n-1}. \tag{3.15}$$

The Gel'fand–Tsetlin kets for $SO_{n-1,1}$, classified by the canonical chain[25] $SO_{n-1,1} \supset SO_{n-1} \supset \cdots \supset SO_2$, can also be written in the form (3.3), where, for one-valued representations, all indices (except $J_{n,1}$) are integer and follow the "zigzag" inequalities (3.4). The domain of the index $J_{n,1} \equiv \lambda$ is the complex plane. We are at present interested in the principal series[25,26] of UIR's which corresponds to $\lambda = -\frac{1}{2}(n-1) + i\tau$, τ real, and the finite-dimensional (nonunitary) representations which lie at $\lambda = 0, 1, 2, \cdots$ and for which the rest of the inequalities (3.4) hold.

The symbol ${}^P d^J_{LML'}(\zeta)$ will be used for the boost matrix elements of the pseudo-orthogonal group defined, in analogy to (3.6a), as

$${}^P d^J_{LML'}(\zeta) = \langle JLM\bar{N}|\, b_{n-1,n}(\zeta)\, |JL'M\bar{N}\rangle. \tag{3.6b}$$

The orthogonality and completeness relations (3.10) and (3.11) must be carefully justified, as both $\dim_n J$ and vol G are infinite, and an integral with the Plancherel measure $dw(J)$ takes the place of the sum in (3.11). However, we shall not come to need them.

The ISO_{n-1} group is the semidirect product of T_{n-1}, the translation group in $n-1$ dimensions, and SO_{n-1}. Its elements are commonly written as (x, R_{n-1}), $x \in T_{n-1}$ and $R_{n-1} \in SO_{n-1}$, with the usual semidirect product law. The group manifold of ISO_{n-1} is thus the product of the $(n-1)$-dimensional Euclidean space and the SO_{n-1} manifold.

We shall parametrize the former in spherical coordinates, expressing $R_n^I \in ISO_{n-1}$ as

$$R_n^I = R_{n-1}H_n^I = R_{n-1}t_{n-1}(\xi)H_{n-1}, \tag{3.16}$$

where $t_{n-1}(\xi)$ is the translation along the $(n-1)$th direction, $0 \leq \xi < \infty$. In terms of the more usual notation,

$$\begin{aligned} R_n^I &= (0, R_{n-1})(t_{n-1}(\xi), I)(0, H_{n-1}) \\ &= (R_{n-1}t_{n-1}(\xi), R_{n-1}H_{n-1}). \end{aligned}$$

The Haar measure is $dR_n^I = dR_{n-1}\, dH_n^I$, where

$$dH_n^I = \xi^{n-2}\, d\xi\, dH_{n-1}. \tag{3.17}$$

Again, we can use the Gel'fand–Tsetlin kets[24] (3.3) where, for the one-valued representations, all indices (but $J_{n,1}$) are integer and follow (3.4). The index $J_{n-1} \equiv r$ is real.

The symbol ${}^I d^J_{LML'}(\xi)$ will be used for the radial translation matrix elements written, in analogy with (3.6a) and (3.6b), as

$${}^I d^J_{LML}(\xi) = \langle JLM\bar{N}|\, t_{n-1}(\xi)\, |JL'M\bar{N}\rangle. \tag{3.6c}$$

The remarks following Eq. (3.6b) apply to ISO_{n-1}.

4. DEFORMATION OF THE GROUP MANIFOLD

Let $M_{\mu\nu}$ be the generator of a rotation $r_{\mu\nu}(\theta)$ in the (μ, ν) plane, $\mu, \nu = 1, 2, \cdots, n$, of the n-dimensional Euclidean space ($g_{\mu\nu} = \delta_{\mu\nu}$), i.e.,

$$\exp(\theta M_{\mu\nu}) = r_{\mu\nu}(\theta).$$

In terms of the Cartesian coordinates x_μ, they may be represented as

$$M_{\mu\nu} = x_\mu \frac{\partial}{\partial x_\nu} - x_\nu \frac{\partial}{\partial x_\mu}, \tag{4.1}$$

and can be checked to obey the commutation relations of the generators of an so_n algebra:

$$\begin{aligned} [M_{\mu\nu}, M_{\rho\sigma}] = {} & g_{\mu\sigma}M_{\nu\rho} + g_{\nu\rho}M_{\mu\sigma} \\ & + g_{\sigma\nu}M_{\rho\mu} + g_{\rho\mu}M_{\sigma\nu}, \end{aligned} \tag{4.2a}$$

while

$$[M_{\mu\nu}, x_\rho] = g_{\nu\rho}x_\mu - g_{\mu\rho}x_\nu. \tag{4.2b}$$

Finally, we build the second-order Casimir operator of so_n as

$$\Phi^{(n)} = \tfrac{1}{2}\sum_{\mu,\nu} M_{\mu\nu}M_{\mu\nu}. \tag{4.3}$$

Now we construct[22]

$$M_{\mu,n+1} \equiv \tfrac{1}{2}[x_\mu, \Phi^{(n)}] = \tfrac{1}{2}\sum_\nu (M_{\mu\nu}x_\nu + x_\nu M_{\mu\nu}) \tag{4.4}$$

and check that the operators (4.4) together with (4.1) satisfy (4.2a) when we enlarge the range of the indices μ, ν, etc., to $1, 2, \cdots, n, n+1$, with $g_{\mu,n+1} = -x^2\delta_{\mu,n+1}$, where $x^2 = \sum x_\mu x_\mu$. Thus we have built, for $x^2 = 1$, an $so_{n,1}$ algebra (4.2a) out of the iso_n enveloping algebra.

Furthermore, we define the operators

$$M^{(\sigma)}_{\mu,n+1} \equiv M_{\mu,n+1} + \sigma x_\mu, \tag{4.5}$$

and check that (4.5) too, together with (4.1), generates an $so_{n,1}$ algebra whose second-order Casimir operator is

$$\begin{aligned} \Phi^{(n,1)}(\sigma) &= \Phi^{(n)} - \sum_\mu M^{(\sigma)}_{\mu,n+1}M^{(\sigma)}_{\mu,n+1} \\ &= \Phi^{(n,1)}(0) - \sigma^2 x^2. \end{aligned} \tag{4.6}$$

We assume that we already know the representation D-matrices of the SO_n group, and now we want to construct a representation (2.10) for the $SO_{n,1}$ group. As we need only the ${}^{P}d$ matrices for $SO_{n,1}$, we consider the boost generator

$$M^{(\sigma)}_{n,n+1} = x_n \sum_\mu x_\mu \frac{\partial}{\partial x_\mu} - x^2 \frac{\partial}{\partial x_\mu} + [\tfrac{1}{2}(n-1) + \sigma]x_\mu, \tag{4.7}$$

which we have written explicitly in terms of the Cartesian coordinates through (4.1) and (4.4).

We introduce the spherical coordinate system[24] in the n-dimensional space which is best suited for the description of the SO_n group manifold, as

$$x(\{\theta^{(n)}\}) = R_n^{-1}(\{\theta\}^{(n)})x(\{0\}), \tag{4.8a}$$

where $x(\{0\}) = (0, \cdots, 0, x)$, i.e., $x_n = x \cos\theta_{n-1}$, and

$$x_{n-p} = x \sin\theta_{n-1} \cdots \sin\theta_{n-p} \cos\theta_{n-p-1}, \quad p = 1, \cdots, n-1, \tag{4.8b}$$

where we have put $\theta_q \equiv \theta^{(n)}_{q,q+1}$ for economy.

When acting on functions $f(\{\theta\})$ of the angular coordinates, the operator (4.7) can be written as

$$M^{(\sigma)}_{n,n+1} = \sin\theta \frac{\partial}{\partial\theta} + [\tfrac{1}{2}(n-1) + \sigma]\cos\theta, \tag{4.9}$$

where we have put $\theta \equiv \theta_{n-1}$ for short. (In this connection, see Appendix A.)

In order to identify (4.9) as the generator $N^{(\lambda)}$ of a multiplier representation (2.11) in its form (2.13), we set

$$-\lambda = \tfrac{1}{2}(n-1) + \sigma, \tag{4.10}$$

$$N^{(0)} = \sin\theta \frac{\partial}{\partial\theta} = \frac{\partial}{\partial\omega}, \tag{4.11}$$

where $\omega = \ln\tan(\frac{1}{2}\theta)$ and $Q = \cos\theta$, whereby

$$\rho(\theta) = \sin\theta \tag{4.12}$$

provides an appropriate construction of the multiplier (2.7).

The transformation in the parameter ω brought about by the operator $\exp(\zeta N^{(0)})$ is a translation by ζ, i.e., $\omega \to \omega' = \omega + \zeta$. Hence, in the parameter[27] θ

$$\tan\tfrac{1}{2}\theta \to \tan\tfrac{1}{2}\theta' = \tan\tfrac{1}{2}\theta\, e^{\zeta}. \tag{4.13}$$

Therefore, we can state that, while the operators (4.1) generate *rigid* rotations of the space (4.8b) and therefore on the group manifold of SO_n through (4.8a), the operator (4.9) is the generator of a *deformation* of the group manifold which affects only the parameter $\theta \equiv \theta^{(n)}_{n-1,n}$ through (4.13). (See Appendix A.)

We shall work with functions on the SO_n manifold, rather than on the n-sphere (the homogeneous space $SO_{n-1}\backslash SO_n$), since not all representations of the former can be realized on the latter.[18]

5. THE d-MATRIX ELEMENTS

As was stated in Sec. 3, we can use the set of $D^J_{L_1L_2}(R_n)$ functions which is orthogonal and complete, in order to build a representation of the group of transformations on SO_n through the general procedure (2.10). The specific transformation we are interested in is the deformation (4.13) with the multiplier (2.8) built out of (4.12). Thus, we shall build the D matrices of $SO_{n,1}$ with rows and columns specified by the UIR labels of its canonical chain.

We choose the action on the group to be $g''(g', g) = g'g$. [See text around Eq. (2.9).] Then, for $R_n \in SO_n$,

$$(D^{J_1}_{L_1,L_1'}, U^{(\lambda)}(R_n)D^{J_2}_{L_2',L_2'}) = \delta_{L_1,L_2}\delta_{J_1,J_2}\frac{\mathrm{vol}\,G}{\dim_n J} D^{J_1}_{L_1',L_2'}(R_n), \tag{5.1}$$

because of (2.7), (2.9), and (3.10).

Hence, we can write, for any (fixed, allowed) $\bar{L}$,

$$D^J_{L_1,L_2}(R_n) = \frac{\dim_n J}{\mathrm{vol}\,G}(D^J_{\bar{L},L_1}, U^{(\lambda)}(R_n)D^J_{\bar{L},L_2}), \tag{5.2}$$

incorporating the requirements of (2.10) and the independence of λ and $\bar{L}$, as in (3.7).[28]

Using the operator (4.9) and Eqs. (4.10)–(4.13), we have

$$\begin{aligned}(D^{J_1}_{L_1,L_1'}, \exp(\zeta N^{(\lambda)})D^{J_2}_{L_2',L_2'}) &= \int_R dR_n \overline{D^{J_1}_{L_1'L_1}(R_{n-1}\nu_{n-1,n}(\theta)H_{n-1})}\left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} \\ &\quad \times D^{J_2}_{L_2'L_2'}(R_{n-1}\nu_{n-1,n}(\theta')H_{n-1}) \\ &= \delta_{L_1,L_2}\delta_{L_1',L_2'}\frac{(\mathrm{vol}\,H)^2}{\dim_{n-1}L_1 \cdot \dim_{n-1}L_1'}\sum_M \frac{\dim_{n-2}M}{\mathrm{vol}\,K} \\ &\quad \times \int_0^{\pi} \sin^{n-2}\theta\, d\theta \overline{d^{J_1}_{L_1ML_1'}(\theta)}\left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} d^{J_2}_{L_1ML_1'}(\theta'),\end{aligned} \tag{5.3}$$

where we have used (3.8) and (3.9), as well as (3.10) and (3.12), for SO_{n-1}.

In line with (5.2) and (3.6b), we set

$${}^{P}d^{(\lambda,L)}_{JL'J'}(\zeta) = \frac{(\dim_n J \cdot \dim_n J')^{\frac{1}{2}}}{\mathrm{vol}\,G} \times [D^J_{\bar{L}L'}, \exp(\zeta N^{(\lambda)})D^{J'}_{\bar{L}L'}], \tag{5.4}$$

and we can check that

(a) the representation property is fulfilled, i.e.,

$$\sum_{J''} {}^{P}d^{\{\lambda L\}}_{JL'J''}(\zeta_1)\, {}^{P}d^{\{\lambda L\}}_{J''L'J'}(\zeta_2) = {}^{P}d^{\{\lambda L\}}_{JL'J'}(\zeta_1 + \zeta_2),$$

(b) due to (3.13),

$${}^{P}d^{\{\lambda L\}}_{JL'J'}(0) = \delta_{J,J'}.$$

Hence (5.4) indeed provides a representation of $SO_{n,1}$.

Our method of construction gives automatically the Gel'fand–Tsetlin classification of the representation D matrices. Indeed, from (3.4) (recall $J_{n,q} \equiv J_q$, $J_{n-1,r} \equiv L_r$),

$$J_q \geq L_q \geq J_{q+1}, \quad q = 1, \cdots, [n/2] - 1,$$

and

$$0 \leq L_{[(n-1)/2]} \geq |J_{[n/2]}|, \quad n \text{ even}, \tag{5.5}$$

or

$$0 \leq J_{[n/2]} \geq |L_{[(n-1)/2]}|, \quad n \text{ odd},$$

and similar relations between L and J', L' and J, and L' and J'.

Thus, the $SO_{n,1}$ UIR labels in (5.4),

$$I = \{I_1, \cdots, I_{[(n+1)/2]}\} \equiv \{\lambda, L\} = \{\lambda, L_1, \cdots, L_{[(n-1)/2]}\}, \tag{5.6}$$

also fulfill (5.5) with respect to its row indices (J, J') through the replacements $n \to n+1$, $J \to I$, and $L \to J$, except for the index $I_1 = \lambda$ which is, so far, allowed to be complex and unrestricted. This we shall now investigate.

In order to fulfill the unitarity condition (2.9), we notice that (4.13) yields

$$\frac{\sin^{n-2}\theta'}{\sin^{n-2}\theta}\frac{d\theta'}{d\theta} = \left(\frac{\sin\theta'}{\sin\theta}\right)^{n-1}, \tag{5.7}$$

and thus, if we demand that $-\lambda - \bar{\lambda} = n - 1$, we shall satisfy (4.10) when

$$-\lambda = \tfrac{1}{2}(n-1) + i\tau, \quad \tau \text{ real}. \tag{5.8}$$

This provides the principal series of UIR's of $SO_{n,1}$.[20] As the rest of the labels (i.e., $I_2, \cdots, I_{[(n+1)/2]}$) satisfy (3.4)–(5.5), they are, indeed, UIR labels of the $SO_{n,1}$ representation matrices.

We are also interested in the finite-dimensional nonunitary irreducible representations of $SO_{n,1}$ since, when we perform the Weyl continuation [i.e., when we consider the parameter ζ in (4.13) as $\zeta = i\theta_n$, $0 \leq \theta_n \leq \pi$], we obtain the UIR matrices of SO_{n+1}.

It is known that the so_{n+1} second-order Casimir operator (4.3) has eigenvalues

$$\phi^{(n+1)} = -l(l+n-1) + \text{integer}, \quad l = 0, 1, 2, \cdots. \tag{5.9}$$

Under the substitution $-l = \frac{1}{2}(n-1) + \sigma$, Eq. (5.9) takes the form

$$\phi^{(n+1)} = [\tfrac{1}{2}(n-1)]^2 - \sigma^2 + \text{integer},$$
$$\sigma = -\tfrac{1}{2}(n-1), -\tfrac{1}{2}(n-1) - 1, \cdots, \tag{5.10}$$

which has the same dependence on σ as (4.6).

The values of σ in (5.10) give the UIR's of SO_{n+1}, and we can now identify l in (5.9) with λ in (5.6) and see that (5.4), with $\lambda = 0, 1, 2, \cdots$, will provide the UIR representations of SO_{n+1}.

The explicit form of the d matrices (5.4) is, from (5.3),

$${}^{P}d^{\{\lambda,L\}}_{JL'J'}(\zeta) = \frac{(\dim_n J \cdot \dim_n J')^{\frac{1}{2}}}{\dim_{n-1} L \dim_{n-1} L'} \frac{(\text{vol } H)^2}{\text{vol } G \text{ vol } K} \sum_M \dim_{n-2} M \times \int_0^\pi \sin^{n-2}\theta\, d\theta \overline{d^J_{LML'}(\theta)} \left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} d^{J'}_{LML'}(\theta'). \tag{5.11}$$

The expressions for $\dim_n J$, $\dim_{n-1} L$, and $\dim_{n-2} M$ can be found from the branching relations (3.4), but are, in general, rather cumbersome to express in closed form [for SO_2, $\dim_2 J = 1$; for SO_3, $\dim_3 J = 2J+1$; for SO_4, $\dim_4 (J,0) = J^2$]. Therefore, we shall leave them thus indicated. The second factor in (5.11) is

$$\frac{(\text{vol } H)^2}{\text{vol } G \cdot \text{vol } K} = \frac{S_{n-1}}{S_n} = \frac{\Gamma(\frac{1}{2}n)}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}(n-1))}. \tag{5.12}$$

Equation (5.11) provides thus, when $\zeta = i\theta_n$ and $\lambda = 0, 1, 2, \cdots$, an inductive procedure whereby the d matrices of SO_{n+1} can be found in terms of those of SO_n.

The first step of the procedure is SO_2, where we have

$$d^J(\theta) = e^{iJ\theta} \quad \text{using} \int_0^{2\pi} \text{since} \quad 0 \leq \theta < 2\pi; \tag{5.13a}$$

but it can also be taken to be SO_4 (since the d matrices for $SO_{2,1}$, SO_3, and $SO_{3,1}$ are well known), since, due to the local isomorphism $SO_4 \cong SO_3 \times SO_3$, we have the simple form[5]

$$d^{I_1 I_2}_{JLJ'}(\theta) = \sum_m C(\tfrac{1}{2}(I_1+I_2), \tfrac{1}{2}(I_1-I_2), J; \tfrac{1}{2}(L+m), \tfrac{1}{2}(L-m), L) \times C(\tfrac{1}{2}(I_1+I_2), \tfrac{1}{2}(I_1-I_2), J'; \tfrac{1}{2}(L+m), \tfrac{1}{2}(L-m), L) e^{im\theta}, \tag{5.13b}$$

where $C(\cdots)$ are the SO_3 Clebsch–Gordan coefficients.

The general form of the $d^I_{JLJ'}(\theta)$ functions for SO_n and $SO_{n,1}$ will not be attempted here beyond the recursion formula (5.11), which may be of more practical use.

The simplest cases, SO_2, $SO_{2,1}$,[2] SO_3,[1] $SO_{3,1}$,[4,6] and[5] SO_4 have been calculated as finite sums of trigonometric (and hypergeometric) functions. The expression (5.13b) for SO_4 is noteworthy for its simplicity. The appearance of 3-J symbols in the coefficient in (5.13b) and Holman's result[13] for SO_5, which involves 9-J symbols, suggests that a relatively compact expression for the general d matrices may yet be found.

The recursion formula (5.11) can be used to determine the asymptotic behavior of ${}^P d^I_{JLJ'}(\zeta)$ as $\zeta \to \infty$. Indeed, from (4.13), we have

$$\sin\theta/\sin\theta' = \cosh\zeta - \sinh\zeta\cos\theta,$$

and hence

$${}^P d^{\{\lambda,L\}}_{JL'J'}(\zeta) \xrightarrow[\zeta\to\infty]{} e^{\lambda\zeta}\Delta^{\lambda L}_{JL'J'}, \tag{5.14a}$$

where

$$\Delta^{\lambda L}_{JL'J'} = \frac{(\dim_n J \dim_n J')^{\frac12}}{\dim_{n-1} L \dim_{n-1} L'} \frac{\Gamma(\frac12 n)}{\pi^{\frac12}\Gamma(\frac12(n-1))} \times \sum_M \dim_{n-2} M d^{J'}_{LML'}(\pi) \times \int_0^\pi \sin^{n-2}\theta\, d\theta \overline{d^J_{LML'}(\theta)} \sin^{2\lambda}\tfrac12\theta. \tag{5.14b}$$

In Appendix B we perform an integral which may be helpful in the evaluation of (5.11) and (5.14).

The ${}^I d$ matrices for the ISO_n groups are found as the matrix elements (3.6c) of the transformation generated by $x_n = r\cos\theta$, $\theta \equiv \theta^{(n)}_{n-1,n}$. In its form (2.7),

$$U^{(\gamma)}(\xi) f(\{\theta\}) = \exp(i\xi\gamma\cos\theta) f(\{\theta\}) \tag{5.15}$$

is unitary [i.e., satisfies (2.9)] for r real. Notice, however, that it produces no deformation of the group and thus cannot be put in the form (2.8).

The second-order Casimir operator $R^2 = \sum x_\mu x_\mu$ has the eigenvalue r^2, whence we can write, as for (5.11),

$${}^I d^{\{\gamma,L\}}_{JLJ'}(\xi) = \frac{(\dim_n J \dim_n J')^{\frac12}}{\dim_{n-1} L \dim_{n-1} L'} \frac{(\mathrm{vol}\, H)^2}{\mathrm{vol}\, G \cdot \mathrm{vol}\, K} \times \sum_M \dim_{n-2} M \int_0^\pi \sin^{n-2}\theta\, d\theta \overline{d^J_{LML'}(\theta)} \times \exp(i\gamma\xi\cos\theta) d^{J'}_{LML'}(\theta). \tag{5.16}$$

Furthermore, we can check that (5.15) and (5.16) are indeed contractions[11] of the corresponding expressions (4.13) and (5.11), i.e., that

$${}^P d^{\{\lambda,L\}}_{JL'J'}(\zeta) \xrightarrow[\substack{\lambda\to i\infty \\ (i\lambda\zeta=\gamma\xi)}]{} {}^I d^{\{\gamma,L\}}_{JL'J'}(\xi),$$

since

$$(\sin\theta/\sin\theta')^\lambda \xrightarrow[\substack{\lambda\to i\infty \\ (i\lambda\zeta=\gamma\xi)}]{} \exp(i\xi\gamma\cos\theta).$$

ACKNOWLEDGMENTS

The author wishes to thank Professor N. Svartholm for his hospitality at the Institute of Theoretical Physics and the Carl-Bertel Nathhorsts Foundation for financial support. It is a pleasure to acknowledge illuminating discussions with Dr. U. Ottoson, Dr. R. Anderson, and especially Dr. A. Kihlberg, to whom the material presented in Appendix A is due.

APPENDIX A

The Iwasawa decomposition[29] of[8c] $G = SO_{n,1}$ can be written uniquely as $g = k(\{\theta\}) \cdot a(\eta) \cdot n(\bar\xi)$, that is,

$$g = \begin{pmatrix} & & 0 \\ & k(\{\theta\}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 & 0 \\ 0 & \cosh\eta & \sinh\eta \\ 0 & \sinh\eta & \cosh\eta \end{pmatrix} \times \begin{pmatrix} \mathbb{1} & -\underline{\xi} & \bar\xi \\ \underline{\xi} & 1-\Xi & \Xi \\ \underline{\xi} & -\Xi & 1+\Xi \end{pmatrix},$$

where $k(\{\theta\}) \in K = SO_n$, the maximal compact subgroup of G parametrized as in (3.1), whereby we have $k_{nn} = \cos\theta$ (using $\theta \equiv \theta^{(n)}_{n-1,n}$), $a(\eta) \in A$, the Abelian subgroup of G which corresponds to the boost in the nth direction in (3.14), $n(\bar\xi) \in N$, the nilpotent subgroup of G, where $\bar\xi$ is the column vector $(\xi_1, \xi_2, \cdots, \xi_{n-1})$, $\underline{\xi}$ its transpose, and

$$\Xi = \tfrac12(\xi_1^2 + \cdots + \xi_{n-1}^2).$$

Consider now the transformation induced by

$$g \xrightarrow{\zeta} g' = a^{-1}(\zeta) g = k(\{\theta'\}) a(\eta') n(\bar\xi'). \tag{A1}$$

Direct calculation yields

$$\cos\theta \xrightarrow{\zeta} \cos\theta' = (\cosh\zeta\cos\theta - \sinh\zeta) \times (\cosh\zeta - \sinh\zeta\cos\theta)^{-1}, \tag{A2a}$$

$$\exp\eta \xrightarrow{\zeta} \exp\eta' = \exp\eta(\cosh\zeta - \sinh\zeta\cos\theta). \tag{A2b}$$

Notice that (A2a) is the same transformation as (4.13). The infinitesimal generator [as in (2.1) and (2.2)] of the transformation (A2) on the space of functions on G/N is thus

$$N = \sin\theta\frac{\partial}{\partial\theta} - \cos\theta\frac{\partial}{\partial\eta}. \tag{A3}$$

The parameter η of the Abelian subgroup $A \subset G$ does not appear as such in (A3) and thus[30] $\partial/\partial\eta$ commutes with (A3) as well as with all the generators of K and of N [whose action can be written in terms of those of K and A through (3.1)]. Hence, we can choose that subspace of functions on $G/AN \cong K$ which corresponds to an eigenvalue λ under $\partial/\partial\eta$ and now the operator (A3) takes exactly the form (4.9)–(4.10),which was obtained from the Gell-Mann formula (4.4)–(4.5) on the space K. The deformation which the latter produces on K is thus seen to be the same as the natural action of $G = KAN$ on itself (modulo AN). [Notice, however, that this is *not* true, modulo H_n, had we taken the decomposition (3.1).]

Through a suitable choice of λ, the operator (A3) can be made anti-Hermitian,[31] and we know a complete and orthogonal set of functions on K: the D matrices for SO_n. Although it is suggestive to consider a similar set on $G/AN \cong K$ as a subset of those on G, the theory of complete sets of functions on homogeneous spaces with noncompact stabilizers is lacking. Some of the difficulties have been pointed out in Ref. 18.

APPENDIX B

Before solving the integrals in (5.11), (5.14b), and (5.16), we have to decide in which form we expect the integrand to appear and try to put the solution in the same form. The cases which are known suggest that $d^I_{JLJ'}(\theta)$ will appear as a sum of powers of $\sin\theta$ and $e^{i\theta}$ for the compact cases and $\sinh\zeta$ and e^{ζ} for the noncompact ones.

We will therefore perform the integral

$$I^{n,\lambda}_{p,q,p',q'}(\zeta) \equiv \int_0^{\pi} \sin^{n-2}\theta\, d\theta (e^{ip\theta}\sin^q\theta)\left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda}(e^{ip'\theta'}\sin^{q'}\theta'), \tag{B1}$$

where p, q, p', and q' are integers and where θ and θ' are related by (4.13). There ζ is real and λ, in general, complex. If we want to be able to make the analytic continuation from $SO_{n,1}$ to SO_{n+1} easily, we need a form where we can replace ζ by $i\theta_n$, $0 \leq \theta_n \leq \pi$, and then let λ be a nonnegative integer.

We express (B1), expanding the exponentials by the binomial theorem, as

$$\begin{aligned} &I^{n,\lambda}_{p,q,p',q'}(\zeta) \\ &= 2^{n+q+q'-2}\sum_{\gamma=0}^{2|p|}\sum_{\gamma'=0}^{2|p'|}\binom{2|p|}{\gamma}\binom{2|p'|}{\gamma'} \\ &\quad\times\left(\frac{|p|}{p}i\right)^{\gamma}\left(\frac{|p'|}{p'}i\right)^{\gamma'} J(\lambda+n-2+q+\gamma, \\ &\qquad -\lambda+q'+\gamma', \lambda+n-2+2|p|-\gamma, \\ &\qquad\qquad -\lambda+2|p'|-\gamma';\zeta), \end{aligned} \tag{B2}$$

where

$$J(a,b,c,d;\zeta) = \int_0^{\pi} d\theta \sin^a\tfrac{1}{2}\theta \sin^b\tfrac{1}{2}\theta' \cos^c\tfrac{1}{2}\theta\cos^d\tfrac{1}{2}\theta'. \tag{B3}$$

In order to solve (B3), substitute[4]

$$x \equiv \sin\theta/\sin\theta', \quad dx = \sinh\zeta\sin\theta\, d\theta,$$

and the limits of the integral $[0, \pi]$ become $[e^{-\zeta}, e^{\zeta}]$, and

$$\begin{aligned} &\sin^m\tfrac{1}{2}\theta\sin^{m'}\tfrac{1}{2}\theta' \\ &\qquad = [e^{\zeta m'}(x-e^{-\zeta})^{m+m'}x^{-m'}(2\sinh\zeta)^{-m-m'}]^{\frac{1}{2}}, \\ &\cos^n\tfrac{1}{2}\theta\cos^{n'}\tfrac{1}{2}\theta' \\ &\qquad = [e^{-\zeta n'}(e^{\zeta}-x)^{n+n'}x^{-n'}(2\sinh\zeta)^{-n-n'}]^{\frac{1}{2}}. \end{aligned}$$

Thus, when $a+b$ and $c+d$ are odd and positive, we can expand

$$\begin{aligned} &J(a,b,c,d;\zeta) \\ &\quad = (2\sinh\zeta)^{-\frac{1}{2}(a+b+c+d)}e^{\frac{1}{2}(b-d)\zeta} \\ &\qquad\times\int_{e^{-\zeta}}^{e^{\zeta}} dx\, x^{-\frac{1}{2}(b+d)}(x-e^{-\zeta})^{\frac{1}{2}(a+b-1)}(e^{\zeta}-x)^{\frac{1}{2}(c+d-1)}, \end{aligned}$$

using the binomial theorem, into a finite number of summands. This is the case in passing from SO_3 to[4] $SO_{3,1}$, but it does not seem to be a general property. Thus, we have to effect the further transformation

$$y = (x-e^{-\zeta})(2\sinh\zeta)^{-1},$$

in order to bring it to a form where it can be found[32] to be

$$\begin{aligned} &J(a,b,c,d;\zeta) \\ &\quad = (2\sinh\zeta)^{-\frac{1}{2}(a+b+c+d)}e^{b\zeta} \\ &\qquad\times\frac{\Gamma(\frac{1}{2}(a+b+1))\Gamma(\frac{1}{2}(c+d+1))}{\Gamma(\frac{1}{2}(a+b+c+d)+1)} \\ &\qquad\times F(\tfrac{1}{2}(b+d);\tfrac{1}{2}(a+b+1); \\ &\qquad\qquad \tfrac{1}{2}(a+b+c+d)+1;1-e^{2\zeta}). \end{aligned} \tag{B4}$$

When we use (B1)–(B4) in order to find the d-matrix elements for SO_{n+1}, we obtain them in terms of trigonometric (and hypergeometric) functions, i.e., in the same form as we assumed them to be when we choose to construct the form (B1).

[1] E. P. Wigner, *Gruppentheorie* (Vieweg, Braunschweig, 1931), pp. 180, 232.
[2] V. Bargmann, Ann. Math. **48**, 568 (1947).
[3] M. Toller, Nuovo Cimento **37**, 631 (1965); **53**, 671 (1968).
[4] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[5] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961); D. Z. Friedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).
[6] A. Z. Dolginov, Zh. Eksp. Teor. Fiz. **30**, 746 (1956) [Sov. Phys. JETP **3**, 589 (1956)]; S. Ström, Arkiv Fysik **29**, 467 (1965); **33**, 465 (1966).
[7] In the notation used in this article, they appear as $d^{\lambda M}_{jmj'}(\zeta)$.
[8] (a) J. Dixmier, Bull, Soc. Math. France **89**, 9 (1960); (b) A. Kihlberg and S. Ström, Arkiv Fysik **31**, 491 (1966); (c) R. Takahashi, Bull. Soc. Math. France **91**, 289 (1963).
[9] S. Ström, Arkiv Fysik **30**, 455 (1965).
[10] W. J. Holman III, J. Math. Phys. **10**, 1888 (1969).
[11] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
[12] J. S. Lomont and Harry E. Moses, J. Math. Phys. **5**, 294 (1964); **5**, 1438 (1964).
[13] W. J. Holman III, J. Math. Phys. **10**, 1710 (1969).
[14] N. Kammer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968): S. A. Williams and D. L. Pursey, *ibid.*, 1230 (1968); R. T. Sharp and S. C. Pieker, *ibid.*, 663 (1968).
[15] J. G. Kuriyan, M. Mukunda, and E. C. G. Sudarshan, Commun. Math. Phys. **9**, 204 (1968).
[16] N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, Mathematical Monographs Vol. 22, translated by V. N. Singh (American Mathematical Society, Providence, R.I., 1968).
[17] K. B. Wolf, Ph.D. thesis, Tel-Aviv University, 1969; Tel-Aviv University, Preprint 112-69, 1969.
[18] R. L. Anderson and K. B. Wolf, J. Math. Phys. **11**, 3176 (1970).
[19] J. Fisher, J. Niederle, and R. Rączka, J. Math. Phys. **7**, 816 (1966); R. Rączka, N. Limić, and J. Niederle, *ibid.*, 1861 (1966); N. Limić, J. Niederle, and R. Rączka, *ibid.*, 2026 (1966); **8**, 1079 (1967); R. Delbourgo, J. Koller, and R. M. Williams, *ibid.* **10**, 957 (1969).
[20] U. Ottoson, private communication.
[21] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 840 (1950); S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967); M. K. F. Wong, *ibid.*, 1899 (1967).
[22] A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965); A. Böhm, Phys. Rev. **145**, 1212 (1966); J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966).
[23] N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 50 (1968).
[24] See Ref. 16, where a slightly different convention is used.
[25] A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).
[26] U. Ottoson, Commun. Math. Phys. **8**, 228 (1968).
[27] This is essentially the relation which appears in Refs. 2, 4, 8c, and 9.
[28] Notice that, had we taken the second choice in Sec. 2 [see text before Eq. (2.7)], i.e., $g''(g',g) = g^{-1}g'$, we would have

$$D^J_{L_1L_2}(R_n) = \frac{\dim_n J}{\text{vol } G}\,(D^J_{L_2,L}, U^{(\lambda)}(R_n^{-1})D^J_{L_1,L})$$

and a similar involution for (5.4).
[29] K. Iwasawa, Ann. Math. **50**, 507 (1949); R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966), Chap. 7.
[30] In fact, it can be checked that $\partial/\partial\eta$ is the infinitesimal generator of the transformation $g \to g'' = ga(\zeta)$.
[31] R. Harish-Chandra, Proc. Natl. Acad. Sci. (U.S.) **37**, 697 (1951); **40**, 1076 (1952); for applications, see A. Kihlberg. Arkiv Fysik **30**, 121 (1965); **32**, 241 (1966); A. Kihlberg, V. F. Müller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).
[32] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Series, Products and Integrals* (Academic, New York, 1966), Eq. 3.181.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Iterated Integral-Transform Trial Functions

R. L. SOMORJAI
Division of Chemistry, National Research Council of Canada, Ottawa 7, Canada

(Received 7 July 1970)

The concept of iterated integral-transform trial functions is introduced. Its formal correspondence with the iterative solution of integral equations is established. Extensions and generalizations are indicated, and some of the advantages of the approach are discussed. Ways are suggested to make tractable the multidimensional integrals that arise in the method.

1. INTRODUCTION

Recently I proposed[1] the use of integral-transform (IT) trial functions in quantum-mechanical calculations. The conceptual simplicity of the basic idea enhances the computational successes that we achieved with IT trial functions.[2–6] This simplicity makes possible extensions and generalizations quite natural. The systematic construction of special, correlated many-particle wavefunctions,[7] various generalizations of the conventional scaling procedure and their natural relation to correlation,[8] and the construction of new *molecular* functions from *atomic* bases[9] are the most important examples of such extensions. In this work, a further generalization will be introduced, the concept of *iterated* IT trial functions.

2. INTEGRAL TRANSFORM FUNCTIONS

Integral-transform trial functions may be constructed by the prescription

$$F_1(x) = \int_{D_0} S_0(t)F_0(tx)\,d\mu(t). \tag{2.1}$$

In Eq. (2.1) $F_1(x)$ is an approximation to $F(x)$, the exact solution to the eigenvalue equation $HF(x) = EF(x)$, $F_0(x)$ is the *known* exact solution of a model

[1] E. P. Wigner, *Gruppentheorie* (Vieweg, Braunschweig, 1931), pp. 180, 232.
[2] V. Bargmann, Ann. Math. **48**, 568 (1947).
[3] M. Toller, Nuovo Cimento **37**, 631 (1965); **53**, 671 (1968).
[4] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[5] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961); D. Z. Friedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).
[6] A. Z. Dolginov, Zh. Eksp. Teor. Fiz. **30**, 746 (1956) [Sov. Phys. JETP **3**, 589 (1956)]; S. Ström, Arkiv Fysik **29**, 467 (1965); **33**, 465 (1966).
[7] In the notation used in this article, they appear as $d^{\lambda M}_{jmj'}(\zeta)$.
[8] (a) J. Dixmier, Bull, Soc. Math. France **89**, 9 (1960); (b) A. Kihlberg and S. Ström, Arkiv Fysik **31**, 491 (1966); (c) R. Takahashi, Bull. Soc. Math. France **91**, 289 (1963).
[9] S. Ström, Arkiv Fysik **30**, 455 (1965).
[10] W. J. Holman III, J. Math. Phys. **10**, 1888 (1969).
[11] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
[12] J. S. Lomont and Harry E. Moses, J. Math. Phys. **5**, 294 (1964); **5**, 1438 (1964).
[13] W. J. Holman III, J. Math. Phys. **10**, 1710 (1969).
[14] N. Kammer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968): S. A. Williams and D. L. Pursey, *ibid.*, 1230 (1968); R. T. Sharp and S. C. Pieker, *ibid.*, 663 (1968).
[15] J. G. Kuriyan, M. Mukunda, and E. C. G. Sudarshan, Commun. Math. Phys. **9**, 204 (1968).
[16] N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, Mathematical Monographs Vol. 22, translated by V. N. Singh (American Mathematical Society, Providence, R.I., 1968).
[17] K. B. Wolf, Ph.D. thesis, Tel-Aviv University, 1969; Tel-Aviv University, Preprint 112-69, 1969.
[18] R. L. Anderson and K. B. Wolf, J. Math. Phys. **11**, 3176 (1970).
[19] J. Fisher, J. Niederle, and R. Rączka, J. Math. Phys. **7**, 816 (1966); R. Rączka, N. Limić, and J. Niederle, *ibid.*, 1861 (1966); N. Limić, J. Niederle, and R. Rączka, *ibid.*, 2026 (1966); **8**, 1079 (1967); R. Delbourgo, J. Koller, and R. M. Williams, *ibid.* **10**, 957 (1969).
[20] U. Ottoson, private communication.
[21] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 840 (1950); S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967); M. K. F. Wong, *ibid.*, 1899 (1967).
[22] A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965); A. Böhm, Phys. Rev. **145**, 1212 (1966); J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966).
[23] N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 50 (1968).
[24] See Ref. 16, where a slightly different convention is used.
[25] A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).
[26] U. Ottoson, Commun. Math. Phys. **8**, 228 (1968).
[27] This is essentially the relation which appears in Refs. 2, 4, 8c, and 9.
[28] Notice that, had we taken the second choice in Sec. 2 [see text before Eq. (2.7)], i.e., $g''(g',g) = g^{-1}g'$, we would have

$$D^J_{L_1L_2}(R_n) = \frac{\dim_n J}{\text{vol } G}\,(D^J_{L_2,L}, U^{(\lambda)}(R_n^{-1})D^J_{L_1,L})$$

and a similar involution for (5.4).
[29] K. Iwasawa, Ann. Math. **50**, 507 (1949); R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966), Chap. 7.
[30] In fact, it can be checked that $\partial/\partial\eta$ is the infinitesimal generator of the transformation $g \to g'' = ga(\zeta)$.
[31] R. Harish-Chandra, Proc. Natl. Acad. Sci. (U.S.) **37**, 697 (1951); **40**, 1076 (1952); for applications, see A. Kihlberg. Arkiv Fysik **30**, 121 (1965); **32**, 241 (1966); A. Kihlberg, V. F. Müller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).
[32] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Series, Products and Integrals* (Academic, New York, 1966), Eq. 3.181.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Iterated Integral-Transform Trial Functions

R. L. SOMORJAI
Division of Chemistry, National Research Council of Canada, Ottawa 7, Canada

(Received 7 July 1970)

The concept of iterated integral-transform trial functions is introduced. Its formal correspondence with the iterative solution of integral equations is established. Extensions and generalizations are indicated, and some of the advantages of the approach are discussed. Ways are suggested to make tractable the multidimensional integrals that arise in the method.

1. INTRODUCTION

Recently I proposed[1] the use of integral-transform (IT) trial functions in quantum-mechanical calculations. The conceptual simplicity of the basic idea enhances the computational successes that we achieved with IT trial functions.[2–6] This simplicity makes possible extensions and generalizations quite natural. The systematic construction of special, correlated many-particle wavefunctions,[7] various generalizations of the conventional scaling procedure and their natural relation to correlation,[8] and the construction of new *molecular* functions from *atomic* bases[9] are the most important examples of such extensions. In this work, a further generalization will be introduced, the concept of *iterated* IT trial functions.

2. INTEGRAL TRANSFORM FUNCTIONS

Integral-transform trial functions may be constructed by the prescription

$$F_1(x) = \int_{D_0} S_0(t)F_0(tx)\,d\mu(t). \tag{2.1}$$

In Eq. (2.1) $F_1(x)$ is an approximation to $F(x)$, the exact solution to the eigenvalue equation $HF(x) = EF(x)$, $F_0(x)$ is the *known* exact solution of a model

Hamiltonian $H_0 = H - V$, D_0 is a suitable integration domain of the nonphysical t-space, $\mu(t)$ is its measure, and $S_0(t)$ is the unknown shape function to be determined. In fact, Eq. (2.1) may be considered the (formal) *continuous* analog of a conventional expansion of $F(x)$ in eigenvectors $F_0(x; n)$ of $H_0(x)$:

$$F(x) \approx F_1(x; N) = \sum_{n=1}^{N} C_n F_0(x; n). \tag{2.2}$$

If $\{F_0(x; n)\}$ forms a complete set, $\lim F_1(x; N) = F(x)$ as $N \to \infty$. Similarly, if $F_0(tx)$ were complete in D_0 with respect to the *continuous* index t, $F_1(x)$ would become $F(x)$ for the exact $S(t)$.

In practice, $F_1(x; N)$ is usually replaced by the more flexible $G_1(x; N, \alpha)$:

$$G_1(x; N, \alpha) = \sum_{n=1}^{N} D_n F_0(\alpha_n x; n). \tag{2.3}$$

For fixed N, G_1 is more accurate than $F_1(x; N)$ because of the N additional adjustable parameters α_n. The connection between (2.1) and (2.3) is more than a formal one: By interpreting (2.1) as a Stieltjes integral, an appropriate choice of S_0 and μ will result in G_1.

The exact $S(t)$ satisfies a generally difficult integral equation. Instead of attempting to solve this, one parametrizes some preselected trial $S_0(t; T)$ and optimizes the parameters $\{T\}$ variationally. The choice of an appropriate analytical form for $S_0(t; T)$ is facilitated by the observation[7] that as $V = H - H_0$ approaches zero, $S_0(t) \to \delta(t - 1)$. Thus the chosen $S_0(t; T)$ should become the δ function for certain limiting values of its parameters $\{T\}$. A useful class of such δ-converging sequences is[10]

$$S_0(t; T) = \tfrac{1}{2} T f(Tt) \Big/ \int_0^\infty f(s)\, ds,$$

provided that $f(x)$ is even, $f(0) \neq 0$, and the integral in the denominator exists and is absolutely convergent with respect to its upper limit. Then $\lim S_0(t; T) = \delta(t)$ as $T \to \infty$.

3. ITERATED INTEGRAL TRANSFORMS

A. Basic Definition

The prescription of Eq. (2.1) is equally valid if $F_0(tx)$ is replaced by $F_1(tx)$. In general, the nth iterated IT function is generated from its predecessor, $F_{n-1}(x)$, by the formula

$$\begin{aligned} F_n(x) &= \int_{D_{n-1}} S_{n-1}(t) F_{n-1}(tx)\, dt \\ &= \int_{D_{n-1}} \cdots \int_{D_0} F_0\left(x \prod_{i=0}^{n-1} t_i\right) \prod_{i=0}^{n-1} S_i(t_i)\, dt_i. \end{aligned} \tag{3.1}$$

The n-fold integral in Eq. (3.1) is the result of applying n times the equivalent of Eq. (2.1) to $F_0(x)$. At each iteration stage both the domain and the shape function may be changed. Before we indicate variants of Eq. (3.1) and their practical usefulness, we shall exhibit and discuss an important *formal* relationship between the integral equation formulation of our eigenvalue equation $H|F\rangle = E|F\rangle$ and the *exact* shape function $S(t)$.

B. Integral Equations and $S(t)$

Assume that $H_0(t)|t\rangle = \epsilon(t)|t\rangle$ holds for all t, t *a continuous index*. (Standard Dirac notation is used.) In the x representation (coordinate representation) this reads $H_0(x, t)\langle x \mid t\rangle = \epsilon(t)\langle x \mid t\rangle$, and we identify $F_0(tx)$ with $\langle x \mid t\rangle$ and $H_0(x)$ with $H_0(x, t)$. We also assume that $\{|t\rangle\}$ forms a *complete* set, i.e., $\int |t\rangle\langle t|\, dt = 1$. Then we may expand the eigenfunction $|F\rangle$ of the total Hamiltonian H in $\{|t\rangle\}$:

$$|F\rangle = \int dt\, |t\rangle\langle t \mid F\rangle, \tag{3.2}$$

i.e.,

$$\begin{aligned} \langle x \mid F\rangle \equiv F(x) &= \int dt \langle t \mid F\rangle\langle x \mid t\rangle \\ &= \int dt S(t) F_0(tx). \end{aligned} \tag{3.3}$$

Thus $S(t)$ is an expansion coefficient, but it is a *continuous* function of the state index t. We can recast Eq. (3.2) in the form of an integral equation by the following formal manipulations[11]:

$$\begin{aligned} |F\rangle &= \int dt \frac{|t\rangle\langle t|}{E - \epsilon(t)} [E - \epsilon(t)]\, |F\rangle \\ &= \int dt \frac{|t\rangle\langle t|}{E - \epsilon(t)} |VF\rangle. \end{aligned} \tag{3.4}$$

The last step in Eq. (3.4) follows from $H = H_0(t) + V(t)$ and from the equation $[E - H_0(t)]|F\rangle = V|F\rangle$. Introducing the resolution of the identity

$$\int dy\, |y\rangle\langle y| = 1,$$

$\{|y\rangle\}$ an *arbitrary* complete set, we obtain

$$\begin{aligned} |F\rangle &= \int dt \left(\int dy \frac{\langle t \mid y\rangle}{E - \epsilon(t)} \langle y \mid VF\rangle \right) |t\rangle \\ &= \int dy \left(\int dt \frac{|t\rangle\langle t \mid y\rangle}{E - \epsilon(t)} \langle y \mid VF\rangle \right). \end{aligned} \tag{3.5}$$

In the x representation the two alternative expressions of Eq. (3.5) give

$$F(x) = \int dt S(t) F_0(tx) \tag{3.6}$$

and

$$F(x) = \int dy G(x, y; E)V(y)F(y), \tag{3.7}$$

provided that the following identifications are made:

$$S(t) \equiv \int dy \frac{F_0(ty)V(y, t)F(y)}{E - \epsilon(t)}, \tag{3.8}$$

$$G(x, y; E)V(y) \equiv \int dt \frac{F_0(ty)F_0(tx)V(y, t)}{E - \epsilon(t)}. \tag{3.9}$$

Equation (3.7) is the conventional integral equation reformulation of the Schrödinger equation; $G(x, y; E)$ is the energy-dependent Green's function, but expanded in the continuous eigenset of $H_0(x, t)$. Equations (3.6) and (3.3) are identical, and Eq. (3.8) is the fundamental relationship we sought.

The only unusual feature in the above comparison is the use of continuous-index sets for the expansions. These are by no means uncommon [the well-known plane waves, $\exp(-itx)$, form such a complete continuous set]. Nevertheless, they play a crucial role in the integral transform method and are instrumental in its success.

It is instructive to compare the integral equation approach with the integral-transform trial function method. To solve Eq. (3.7), one needs to know the "perturbation" $V(y)$ as well as the Green's function $G(x, y; E)$. The latter is particularly difficult to calculate analytically, even for the simplest quantum mechanical systems.[12] In contrast, in using Eq. (3.6), the only unknown quantity is $S(t)$, and neither the partitioning of H into H_0 and V nor the eigenvalue spectrum $\{\epsilon(t)\}$ need be known. Of course, an expression completely analogous to Eq. (3.8) exists for the *discrete* expansion coefficients C_n of Eq. (2.2) [or for D_n of Eq. (2.3)]. However, the advantage of using $F_1(x)$ instead of $F_1(x; N)$ or even $G_1(x; N, \alpha)$ as a trial function in variational calculations is that only the single function $F_0(x)$ has to be known instead of the N basis functions $\{F_0(x; n)\}$. We generate our set by simply "scaling" the variables in $F_0(x)$. Because of the smoothing-out effect of integration, one expects F_1 to be less sensitive to errors in a trial $S(t)$ than its counterparts $F_1(x; N)$ and G_1 are to errors in $\{C_n\}$ and $\{D_n, \alpha_n\}$, respectively.

The integral equation (3.7) may be solved by some sort of iteration procedure. In the simplest version, one generates $F_n(x)$ by replacing $F(y)$ in the integrand by $F_{n-1}(y)$, $n = 1, 2, \cdots$. Its equivalent in the integral transform method arises if the trial $S_0(t)$ is generated from Eq. (3.8) by replacing $F(y)$ by $F_0(y)$. Thus, at least formally, the mathematical foundations of the iterated IT method are established. However, we do not need Eq. (3.8) to generate an approximation to $S(t)$, an important practical advantage. Furthermore, the idea behind Eq. (3.1) is more akin to a *variation–iteration* approach, since $F_n(x)$ is optimized variationally [by optimizing $S_{n-1}(t)$]. This happens at each iteration, independent of any previous optimization, and consequently the convergence to $F(x)$ may be expected to be faster than in the case of the simple iteration method.

C. Generalizations

The formal correspondence I have established between the integral equation approach and the iterated integral transforms refers to the *total* solution of the Schrödinger equation, i.e., x in $F(x)$ denotes the collection of all particle coordinates. However, because of the variational character of the iterated IT trial function, this restriction may be relaxed. Thus, one may choose to iterate on the basic building blocks, the orbitals, as the other extreme possibility. In particular, the iteration may be confined to the *radial* part of the orbital.

Each iteration introduces a double integration. This could rapidly lead to an "integration catastrophe." One may avoid this by eliminating some of the integrations via an extension of the shape function. To illustrate, consider $F_2(x)$:

$$F_2(x) = \int_{D_1} \int_{D_0} dt_0\, dt_1 S_0(t_0)S_1(t_1)F_0(t_0t_1x). \tag{3.10}$$

We eliminate the integration with respect to t_1 by replacing $S_1(t_1)$ with $S_1(t_1)\delta(t_1 - f(t_0))$. This gives

$$\bar{F}_2(x) = \int_{D_0} dt_0 S_0(t_0)S_1[f(t_0)]F_0[t_0f(t_0)x]. \tag{3.11}$$

One may consider $\bar{F}_2(x)$ as a special version of the generalization G_2 of Eq. (3.10):

$$G_2(x) = \int_{D_1} \int_{D_0} dt_0\, dt_1 S_{01}(t_0, t_1)F_0(t_0t_1x). \tag{3.12}$$

The "generalized second iterate" G_2 of F_0 reduces to F_2 if one replaces S_{01} by S_0S_1; it becomes $\bar{F}_2$ if the coupling of t_0 and t_1 is via the δ function. Properly speaking, G_2 is not an iterate since it cannot be generated from some precursor. This is because the variables t_0 and t_1 are assumed coupled in $S_{01}(t_0, t_1)$.

The extension of $\bar{F}_2(x)$ to an nth-generalized iterate $\bar{F}_{nk}(x)$ is in the spirit of the integral transform correlated wavefunction approach introduced elsewhere.[7] Thus, $\bar{F}_{nk}(x)$ is obtained from $F_n(x)$ of Eq. (3.1) by

introducing the generalized function $\delta(P_0, P_1, \cdots, P_{k-1})$, $k \leq n-1$, $P_i = P_i(t_0, t_1, \cdots, t_{n-1})$:

$$\bar{F}_{nk}(x) = \underbrace{\int \cdots \int}_{n\text{-fold}} \delta(P_0, \cdots, P_{k-1}) \times F_0\left(x \prod_{i=0}^{n-1} t_i\right) \prod_{i=0}^{n-1} S_i(t_i)\, dt_i. \quad (3.13)$$

The role of the generalized function $\delta(P_0, \cdots, P_{k-1})$ is to reduce the n-fold integral of Eq. (3.1) to the more tractable $(n-k)$-fold integral of Eq. (3.13). The latter is confined to the manifold $P_0 = P_1 = \cdots = P_{k-1} = 0$. $\bar{F}_2(x)$ of Eq. (3.11) is identical to $\bar{F}_{21}(x)$ of Eq. (3.13), with $P_0 = t_1 - f(t_0)$. For detailed mathematical definitions, Ref. 7 should be consulted.

4. DISCUSSION

I have indicated a *formal* relationship between the iterated IT approach and the simplest iterative solution of the integral equation form of the Schrödinger equation. It has also been shown how to generalize iterated IT functions while making them computationally more tractable. These conclusions are further indications of the superiority of IT trial functions in comparison with conventional discrete-set-variational functions. The latter cannot be improved *iteratively* (unless the more difficult integral equation formulation is used), because that would require that a *new* discrete set, of which $F_1(x; N)$ is the first member, is known or can be generated simply. In contrast, the very flexible prescription of Eq. (3.1) demands only the *formal* scaling of the previous iterate and the almost trivial choice of a new shape function $S_k(t)$ (which may even be S_{k-1}).[13]

Some of the important advantages of the iterated integral-transform functions approach are listed below.

(i) The approach provides a *systematic* way of generating new trial functions that are guaranteed to be more accurate than previous iterates. Furthermore, the generation of new, *complete sets* is equally feasible[14]; by iteration these sets can be improved systematically.

(ii) The number of parameters in the S_k may be kept low, since at each iteration these are optimized *independently* of the parameters in previous shape functions.

(iii) The choice of $S_k(t)$ can be based on computational convenience, i.e., either rendering $F_{k+1}(x)$ and/or integrals involving $F_{k+1}(x)$ analytically tractable or making the integrands slowly varying, smooth functions of the integration variables. For example, when $k > 1$, the simplest fractional-integral transform[4] [$S_k = 1$, $D_k = (A, B)$] should probably be sufficient and useful since one expects (A, B) to bracket unity quite closely and, therefore, fine details of shape lose their importance. Furthermore, numerical integrations over the domain D_k ought to be efficient.

(iv) The special case $S_k(t; T_k) = S_0(t; T_0)$, all k, corresponds to solving an integral equation by simple iteration. It has the great computational advantage that the parameters $\{T_0\}$ need be optimized but once. Losing the variational character of the energy is unlikely to be important at this stage. Of course, S_0 should already be sufficiently accurate, i.e., the simple iterative approach should converge, and at a reasonable rate.

One disadvantage of the iterated integral transform method is the multidimensional numerical integrations one is forced to do in general. However, the advantages more than compensate for this, especially when (iii) and (iv) are given special consideration. Furthermore, by using the generalized δ function, we can reduce the dimensionality of the IT trial function, as was indicated in Sec. 3C. Numerical integration techniques that take into account the δ-function-like behavior of $S_k(t)$ are under investigation. In particular, the Hilbert transform method[15] appears to be promising, especially since one has essentially unlimited freedom in selecting $S_k(t)$. For an appropriately chosen $S_k(t)$ one may, using the above method, replace a *real* integration with a sharply peaked integrand by a *complex* integration over a much more smoothly and slowly varying integrand. The "peakedness" of the real integrand is handled *analytically*. Since it provides the major contribution to the numerical value of the integral, the accuracy of the residual complex integral is much less critical. A more primitive version of the above approach uses the special properties of the integrands we may encounter. Thus, all our integrals are of the form

$$I = \underbrace{\int \cdots \int}_{k} \prod_{i=1}^{k} S_i(t_i)\, dt_i F(t_1, t_2, \cdots, t_k),$$

which we propose to break up into two integrals:

$$I = CI_1 + I_2 = C \prod_{i=1}^{k} \left(\int S_i(t_i)\, dt_i\right) + \int \cdots \int \{F(t_1, t_2, \cdots, t_k) - C\} \times \prod_{i=1}^{k} S_i(t_i)\, dt_i.$$

The advantage of evaluating I this way is that I_1 is, at worst, a product of k 1-dimensional integrals. Furthermore, the "peakedness" of the integrand is caused by the S_i, and so the relative contribution of I_2

to I is small. The integrand of I_2 is smoother, and consequently fewer integrand evaluations are necessary for a given accuracy. The constant C can be chosen to minimize the contribution of I_2 to I. The simplest way of doing this is to first evaluate I approximately, $I \approx I_a$. I_1 is then computed accurately, and C is chosen to be $C \approx I_a/I_1$; then $I_2 \approx 0$.

The integral transform trial functions used in previous calculations may be considered as first iterates in the *iterated* IT scheme. Calculations with higher iterates are now in progress.

ACKNOWLEDGMENT

I thank Dr. S. Baranovsky for interesting discussions and Dr. D. Bishop for a reading of the manuscript.

[1] R. L. Somorjai, Chem. Phys. Letters **2**, 399 (1968).
[2] R. L. Somorjai, Chem. Phys. Letters **3**, 395 (1969).
[3] D. M. Bishop and R. L. Somorjai, J. Math. Phys. **11**, 1150 (1970).
[4] R. L. Somorjai and D. M. Bishop, Phys. Rev. A **1**,1013 (1970).
[5] C. P. Yue and R. L. Somorjai, Chem. Phys. Letters **5**, 537 (1970).
[6] R. L. Somorjai and C. P. Yue, J. Chem. Phys. **53**, 1657 (1970).
[7] R. L. Somorjai, Phys. Rev. Letters **23**, 329 (1969).
[8] R. L. Somorjai (to be published).
[9] R. L. Somorjai (to be published).
[10] B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U.P., Cambridge, 1955), p. 70.
[11] The author thanks Dr. S. Baranovsky for suggesting this approach.
[12] H. F. Hameka, J. Chem. Phys. **49**, 2002 (1968).
[13] *Note Added in Proof:* It must be kept in mind that $F_0(tx)$, considered as an expansion set with continuous index t, may be overcomplete. The following discussion assumes that such occurrence has been guarded against.
[14] R. L. Somorjai, invited paper at the Third Canadian Symposium on Theoretical Chemistry, Toronto, 1969.
[15] W. E. Smith and J. N. Lyness, Math. Comp. **23**, 231 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Existence of a Solution of a Stochastic Integral Equation in Turbulence Theory

W. J. PADGETT AND C. P. TSOKOS
Department of Statistics and Statistical Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 7 July 1970)

In the theory of turbulence, the random position of a tagged point in a continuous fluid in turbulent motion, $\mathbf{r}(t;\omega)$, is a vector-valued random function of time $t \geq 0$, $\omega \in \Omega$, where Ω is the supporting set of the underlying probability space (Ω, B, P). If $\mathbf{u}(\mathbf{r}, t;\omega)$ is the Eulerian velocity field, then $\mathbf{r}(t;\omega)$ satisfies the stochastic integral equation $\mathbf{r}(t;\omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega)\, d\Upsilon$, $t \geq 0$. General conditions under which a random solution of this stochastic integral equation exists are given in the form of a theorem, and the theorem is proved using the concepts of admissibility with respect to an operator on a Banach space and fixed-point methods of functional analysis.

1. INTRODUCTION

A theoretical approach to the study of a continuous fluid in turbulent motion is virtually impossible except in a stochastic framework because the velocity fluctuations are random.

Consider a tagged material point in a continuous fluid in turbulent motion. The position of the designated point at time $t > 0$ which started at time zero at some reference point, say, the origin, is a vector-valued random variable $\mathbf{r}(t;\omega)$, $\omega \in \Omega$, where Ω is the supporting set of the underlying probability space (Ω, B, P). At position $\mathbf{r}(t;\omega)$, the velocity of the point is given by the Eulerian velocity field $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$, which is a vector-valued random variable for each $\mathbf{r}$ and $t \geq 0$. The Eulerian approach to describing flow in fluid mechanics is that the velocity of a point in a continuous fluid moving by a fixed position at any time $t \geq 0$ can be given. If the flow is turbulent, however, then the velocity is random, and the problem is to determine $\mathbf{r}(t;\omega)$ if the statistical properties of $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$ are known.

The position of the designated point is given by the stochastic integral equation

$$\mathbf{r}(t;\omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega)\, d\Upsilon, \qquad (1.1)$$

where

(i) $\omega \in \Omega$,

(ii) $\mathbf{r}(t;\omega)$ is the unknown vector-valued random function which gives the coordinates of the position of the tagged point in the fluid for each time $t \geq 0$,

(iii) $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$ is the Eulerian velocity field specified in laboratory coordinates for each $t \geq 0$.

to I is small. The integrand of I_2 is smoother, and consequently fewer integrand evaluations are necessary for a given accuracy. The constant C can be chosen to minimize the contribution of I_2 to I. The simplest way of doing this is to first evaluate I approximately, $I \approx I_a$. I_1 is then computed accurately, and C is chosen to be $C \approx I_a/I_1$; then $I_2 \approx 0$.

The integral transform trial functions used in previous calculations may be considered as first iterates in the *iterated* IT scheme. Calculations with higher iterates are now in progress.

ACKNOWLEDGMENT

I thank Dr. S. Baranovsky for interesting discussions and Dr. D. Bishop for a reading of the manuscript.

[1] R. L. Somorjai, Chem. Phys. Letters **2**, 399 (1968).
[2] R. L. Somorjai, Chem. Phys. Letters **3**, 395 (1969).
[3] D. M. Bishop and R. L. Somorjai, J. Math. Phys. **11**, 1150 (1970).
[4] R. L. Somorjai and D. M. Bishop, Phys. Rev. A **1**,1013 (1970).
[5] C. P. Yue and R. L. Somorjai, Chem. Phys. Letters **5**, 537 (1970).
[6] R. L. Somorjai and C. P. Yue, J. Chem. Phys. **53**, 1657 (1970).
[7] R. L. Somorjai, Phys. Rev. Letters **23**, 329 (1969).
[8] R. L. Somorjai (to be published).
[9] R. L. Somorjai (to be published).
[10] B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U.P., Cambridge, 1955), p. 70.
[11] The author thanks Dr. S. Baranovsky for suggesting this approach.
[12] H. F. Hameka, J. Chem. Phys. **49**, 2002 (1968).
[13] *Note Added in Proof:* It must be kept in mind that $F_0(tx)$, considered as an expansion set with continuous index t, may be overcomplete. The following discussion assumes that such occurrence has been guarded against.
[14] R. L. Somorjai, invited paper at the Third Canadian Symposium on Theoretical Chemistry, Toronto, 1969.
[15] W. E. Smith and J. N. Lyness, Math. Comp. **23**, 231 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Existence of a Solution of a Stochastic Integral Equation in Turbulence Theory

W. J. PADGETT AND C. P. TSOKOS
Department of Statistics and Statistical Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 7 July 1970)

In the theory of turbulence, the random position of a tagged point in a continuous fluid in turbulent motion, $\mathbf{r}(t;\omega)$, is a vector-valued random function of time $t \geq 0$, $\omega \in \Omega$, where Ω is the supporting set of the underlying probability space (Ω, B, P). If $\mathbf{u}(\mathbf{r}, t;\omega)$ is the Eulerian velocity field, then $\mathbf{r}(t;\omega)$ satisfies the stochastic integral equation $\mathbf{r}(t;\omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega)\, d\Upsilon$, $t \geq 0$. General conditions under which a random solution of this stochastic integral equation exists are given in the form of a theorem, and the theorem is proved using the concepts of admissibility with respect to an operator on a Banach space and fixed-point methods of functional analysis.

1. INTRODUCTION

A theoretical approach to the study of a continuous fluid in turbulent motion is virtually impossible except in a stochastic framework because the velocity fluctuations are random.

Consider a tagged material point in a continuous fluid in turbulent motion. The position of the designated point at time $t > 0$ which started at time zero at some reference point, say, the origin, is a vector-valued random variable $\mathbf{r}(t;\omega)$, $\omega \in \Omega$, where Ω is the supporting set of the underlying probability space (Ω, B, P). At position $\mathbf{r}(t;\omega)$, the velocity of the point is given by the Eulerian velocity field $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$, which is a vector-valued random variable for each $\mathbf{r}$ and $t \geq 0$. The Eulerian approach to describing flow in fluid mechanics is that the velocity of a point in a continuous fluid moving by a fixed position at any time $t \geq 0$ can be given. If the flow is turbulent, however, then the velocity is random, and the problem is to determine $\mathbf{r}(t;\omega)$ if the statistical properties of $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$ are known.

The position of the designated point is given by the stochastic integral equation

$$\mathbf{r}(t;\omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega)\, d\Upsilon, \qquad (1.1)$$

where

(i) $\omega \in \Omega$,

(ii) $\mathbf{r}(t;\omega)$ is the unknown vector-valued random function which gives the coordinates of the position of the tagged point in the fluid for each time $t \geq 0$,

(iii) $\mathbf{u}(\mathbf{r}(t;\omega), t;\omega)$ is the Eulerian velocity field specified in laboratory coordinates for each $t \geq 0$.

In Ref. 1, Lumley approached the problem by considering a discretized version of Eq. (1.1). Let

$$\boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t, \Upsilon;\omega) = \mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega), \quad 0 \le \Upsilon \le t,$$
$$= 0, \qquad \Upsilon > t.$$

Then the discretized version of (1.1) that was considered by Lumley[1] is given by

$$\mathbf{r}_n(t_i;\omega) = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t_i, \Upsilon;\omega)\, d\Upsilon$$
$$\doteq \sum_{k=1}^{n} \boldsymbol{\phi}(\mathbf{r}(t_k;\omega), t_i, t_k;\omega)\Delta_k,$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of the interval $[0, t]$, $t_i \in (t_{i-1}, t_{i+1})$, $i = 1, 2, \cdots, n-1$, and $\Delta_k = t_k - t_{k-1}$. An extension of the Rice–Kac theorem was used to find the probability that the discretized version of (1.1) has a solution in a small time interval. However, the convergence of the sequence $\mathbf{r}_n$ to a solution $\mathbf{r}$ as $n \to \infty$, and hence as the partition of the interval $[0, t]$ becomes finer, was not considered.

The aim of this paper is to present some conditions under which an equation of the type (1.1) admits a random solution for each $t \ge 0$. By a *random solution* of (1.1) we shall mean the following: The random vector-valued function $\mathbf{r}(t;\omega)$ is a *random solution* of the stochastic integral equation (1.1) if, for each fixed $t \ge 0$, $\mathbf{r}(t;\omega)$ is a vector random variable and satisfies Eq. (1.1) almost everywhere. We shall use here the concept of admissibility and some techniques similar to those developed by Tsokos.[2]

2. PRELIMINARIES

We shall consider the nonlinear stochastic integral equation (1.1) in the form

$$\mathbf{r}(t;\omega) = \int_0^t \boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t, \Upsilon;\omega)\, d\Upsilon, \qquad (2.1)$$

where $\mathbf{r}(t;\omega)$ is the unknown vector-valued random variable, for each $t \ge 0$, and $\boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t, \Upsilon;\omega)$ is the stochastic kernel which equals $\mathbf{u}(\mathbf{r}(\Upsilon;\omega), \Upsilon;\omega)$, for $0 \le \Upsilon \le t < \infty$, and 0, otherwise.

We assume $\mathbf{r}(t;\omega)$ to be a three-component vector-valued continuous function from $[0, \infty) = R_+$ into the space $L_p(\Omega, B, P)$, $1 \le p < \infty$, that is, for each $t \in R_+$,

$$\|\mathbf{r}(t;\omega)\|_p = \left(\int_\Omega |\mathbf{r}(t;\omega)|^p\, dP(\omega)\right)^{1/p} < \infty.$$

Then, for each $t \in R_+$ and $\omega \in \Omega$, $\mathbf{r}(t;\omega)$ is a point in 3-dimensional Euclidean space $(r_1(t;\omega), r_2(t;\omega), r_3(t;\omega))$, and

$$|\mathbf{r}(t;\omega)|^2 = r_1^2(t;\omega) + r_2^2(t;\omega) + r_3^2(t;\omega).$$

We shall also assume that $\boldsymbol{\phi}$ is a function such that, for each fixed pair (t, Υ) in the set

$$\Delta = \{(t, \Upsilon): 0 \le \Upsilon \le t < \infty\},$$

$\boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t, \Upsilon;\omega)$, $\omega \in \Omega$, is continuous in $\mathbf{r}$, and for each $\mathbf{r}$, $\omega \in \Omega$, $\boldsymbol{\phi}(\mathbf{r}, t, \Upsilon;\omega)$ is a continuous function on Δ. Then $\boldsymbol{\phi}$ is of the form $(\phi_1(\mathbf{r}, t, \Upsilon;\omega), \phi_2(\mathbf{r}, t, \Upsilon;\omega), \phi_3(\mathbf{r}, t, \Upsilon;\omega))$, a point in 3-dimensional Euclidean space for each fixed $\mathbf{r}$, (t, Υ), and $\omega \in \Omega$, and for each fixed $\mathbf{r}$, $(t, \Upsilon) \in \Delta$, we have $\boldsymbol{\phi}(\mathbf{r}, t, \Upsilon;\omega) \in L_p(\Omega, B, P)$, with $|\boldsymbol{\phi}(\mathbf{r}(\Upsilon;\omega), t, \Upsilon;\omega)|$ defined as for $\mathbf{r}(t;\omega)$ above.

Let $C_p(R_+, L_p(\Omega, B, P)) = C_p$ be the space of all three-component vector-valued continuous functions from R_+ into $L_p(\Omega, B, P)$ with the topology of uniform convergence on each compact interval $[0, T]$, $T > 0$. That is, the sequence $\mathbf{r}_n(t;\omega) \in C_p$ converges to $\mathbf{r}(t;\omega) \in C_p$ if and only if

$$\lim_{n\to\infty} \left(\int_\Omega |\mathbf{r}_n(t;\omega) - \mathbf{r}(t;\omega)|^p\, dP(\omega)\right)^{1/p} = 0$$

uniformly on every interval $[0, T]$.

Let D and E be a pair of Banach spaces such that $D, E \subset C_p$, and suppose that T is a linear operator from C_p into itself.

Definition 2.1: The pair of Banach spaces (E, D) is said to be *admissible* with respect to T if and only if $TE \subset D$.

Definition 2.2: The space D is said to be *stronger* than the space C_p if every convergent sequence in D with respect to its norm also converges in C_p, but the converse is generally not true.

We state the following lemma with respect to the Banach spaces D and E and the linear operator T above.[2]

Lemma 2.1: Suppose T is a continuous operator from C_p into itself. If the pair of Banach spaces D and E are stronger than C_p and if (E, D) is admissible with respect to T, then T is a continuous operator from E into D.

It is known from a theorem of Banach[3] that, since T is a continuous operator, it is bounded, and we can find a constant $K > 0$ such that

$$\|(T\mathbf{x})(t;\omega)\|_D \le K\, \|\mathbf{x}(t;\omega)\|_E.$$

We shall now state a well-known theorem which is used extensively in the existence proofs for solutions of deterministic nonlinear integral equations.[4]

Theorem 2.1 (Banach's Fixed-Point Theorem): If F is a contraction mapping from a subset W of a Banach space D into itself, then there exists a unique point $\mathbf{x}$ in W such that $F(\mathbf{x}) = \mathbf{x}$, that is, a unique fixed point of the operator F exists in W.

The norm $\|\mathbf{r}(t; \omega)\|_D$ is defined to be the supremum of $\|\mathbf{r}(t; \omega)\|_p$ for $t \geq 0$.

3. AN EXISTENCE THEOREM

We now consider Eq. (2.1) under the conditions stated in the following theorem.

Theorem 3.1: If Eq. (2.1) satisfies the following conditions, then there exists a unique random solution of (2.1):

(i) D and E are Banach spaces stronger than C_p, and the pair (E, D) is admissible with respect to the operator T given by

$$(T\mathbf{r})(t; \omega) = \int_0^t \mathbf{r}(\Upsilon; \omega)\, d\Upsilon;$$

(ii)

$$\boldsymbol{\phi}(\mathbf{r}(\Upsilon; \omega), t, \Upsilon; \omega) = \mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega), \quad 0 \leq \Upsilon \leq t < \infty,$$
$$= 0, \quad t < \Upsilon < \infty,$$

is a mapping from the set

$$W = \{\mathbf{r}(t; \omega) : \mathbf{r}(t; \omega) \in D, \|\mathbf{r}(t; \omega)\|_D \leq \rho\}$$

into the space E for some $\rho \geq 0$ satisfying

$$\|\mathbf{u}(\mathbf{r}(t; \omega), t; \omega) - \mathbf{u}(\mathbf{s}(t; \omega), t; \omega)\|_E \leq \lambda \|\mathbf{r}(t; \omega) - \mathbf{s}(t; \omega)\|_D$$

for

$$\mathbf{r}(t; \omega), \quad \mathbf{s}(t; \omega) \in W, \quad \text{and} \quad \lambda \geq 0 \text{ a const};$$

(iii) $\|\mathbf{u}(\mathbf{r}(t; \omega), t; \omega)\|_E \leq \rho/K$, where $K > 0$ is the norm of T and $\lambda K < 1$.

Proof: Define the operator U from W into D by

$$(U\mathbf{r})(t; \omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega)\, d\Upsilon.$$

Since $\mathbf{u}$ is continuous in $\mathbf{r}$, U is a continuous mapping from W to D.

We must show that $U(W) \subset W$ (inclusion property) and that U is a contraction mapping on W. We first show that U is a contraction mapping. Let $\mathbf{r}(t; \omega)$ and $\mathbf{s}(t; \omega)$ be in W. Since the difference of two elements of a Banach space is in the Banach space, we have $(U\mathbf{r})(t; \omega) - (U\mathbf{s})(t; \omega) \in D$. Thus,

$$\begin{aligned}
&\|(U\mathbf{r})(t; \omega) - (U\mathbf{s})(t; \omega)\|_D \\
&\quad = \left\| \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega)\, d\Upsilon \right. \\
&\qquad \left. - \int_0^t \mathbf{u}(\mathbf{s}(\Upsilon; \omega), \Upsilon; \omega)\, d\Upsilon \right\|_D \\
&\quad = \left\| \int_0^t [\mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega) - \mathbf{u}(\mathbf{s}(\Upsilon; \omega), \Upsilon; \omega)]\, d\Upsilon \right\|_D \\
&\quad \leq K \|\mathbf{u}(\mathbf{r}(t; \omega), t; \omega) - \mathbf{u}(\mathbf{s}(t; \omega), t; \omega)\|_E \qquad (3.1)
\end{aligned}$$

from the remark following Lemma 2.1. Using condition (ii) of the theorem, we have from inequality (3.1)

$$\|(U\mathbf{r})(t; \omega) - (U\mathbf{s})(t; \omega)\|_D \leq K\lambda \|\mathbf{r}(t; \omega) - \mathbf{s}(t; \omega)\|_D.$$

Since $K\lambda < 1$ by hypothesis, U is a contraction mapping on W.

Now we must show inclusion. Assume that $\mathbf{r}(t; \omega) \in W$. We have

$$\|(U\mathbf{r})(t; \omega)\|_D = \left\| \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega)\, d\Upsilon \right\|_D \leq K \|\mathbf{u}(\mathbf{r}(t; \omega), t; \omega)\|_E \leq \rho$$

by condition (iii), so that $(U\mathbf{r})(t; \omega) \in W$. Hence, since $\mathbf{r}(t; \omega)$ is arbitrary, $U(W) \subset W$. Therefore, by Banach's fixed-point theorem, there exists a unique element $\mathbf{r}(t; \omega)$ in W such that

$$(U\mathbf{r})(t; \omega) = \int_0^t \mathbf{u}(\mathbf{r}(\Upsilon; \omega), \Upsilon; \omega)\, d\Upsilon = \mathbf{r}(t; \omega)$$

for each $t \in R_+$.

Theorem 3.1 applies to any m-component vector-valued functions as well.

Remarks: The invariance of coordinate systems implies that the coordinate system used does not affect the above results.[5] The random vector function $\mathbf{u}(\mathbf{r}, t; \omega)$ gives the random velocity of a tagged particle at the random position $\mathbf{r}(t; \omega)$ at time $t \geq 0$. That is, the components of $\mathbf{u}$ record the random rate of change of the coordinates of the tagged point at time t, $\partial r_i(t; \omega)/\partial t$, $i = 1, 2, 3$, and the length of $\mathbf{u}$, $|\mathbf{u}(\mathbf{r}, t; \omega)|$, records the random speed of the tagged material point at time $t \geq 0$.

[1] J. L. Lumley, J. Math. Phys. 3, 309 (1962).
[2] C. P. Tsokos, Math. Systems Theory 3, 222 (1969).
[3] S. Banach, *Théorie des opérations linéaires* (Chelsea, New York, 1963), 2nd ed.
[4] P. M. Anselone, *Nonlinear Integral Equations* (U. of Wisconsin Press, Madison, 1964).
[5] I. S. Sokolnikoff and R. M. Redheffer, *Mathematics of Physics and Modern Engineering* (McGraw-Hill, New York, 1966), 2nd ed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Quasigroups of Coordinate Transformations in Curved Space-Time*

M. HALPERN AND S. MALIN
Department of Physics and Astronomy, Colgate University, Hamilton, New York

(Received 17 July 1970)

The Poincaré group is well known to imply a deep connection between the structure of space-time and particle properties. Searching for the analogous mathematical structure for curved space-time, we investigate the role of 10-parametric classes of coordinate systems in curved space-times. It is shown that all coordinate systems can be divided into 10-parametric classes according to the values of $g_{\mu\nu}$ and certain nontensorial combinations of its derivatives to an arbitrarily high order at a given point. It is also shown that 10-parametric classes arise naturally in Riemannian space-time when the maximum freedom in assigning numerical values to such nontensorial quantities is used. The set of transformations between members of a given class does not form a group, in general, but a novel mathematical structure (a "quasigroup"). New results concerning the transformations between normal coordinates and the analytic characterization of geodesic Fermi coordinates are derived.

1. INTRODUCTION

The investigation of the representations of the Poincaré group[1] indicates a deep connection between the geometry of space-time and the properties of particles. The Poincaré group itself is the group of motion in flat space-time. It is indeed remarkable that the eigenvalues of its Casimir operators can be interpreted in terms of the masses and spins of particles.

Since, however, space-time is only approximately flat, the following problem naturally arises: Which mathematical structure forms the generalization of the Poincaré group to curved space-time?

The Poincaré group is defined as the set of all transformations between coordinate systems belonging to a certain 10-parametric class of frames,[2] namely, Cartesian frames.[3] The role of 10-parametric sets of coordinate systems in flat space-time is well understood and usually connected with the degree of symmetry of flat space-time. We begin the present work by a general investigation of the role of 10-parametric sets of coordinate systems in Riemannian space-time, in which no symmetry is assumed.

Section 2 is devoted to classifications of all coordinate systems in Riemannian space-time into 10-parametric classes. It is shown that, without knowledge of the geometrical structure and degree of symmetry of space-time, it is always possible to divide all coordinate systems into 10-parametric classes according to the values of the metric tensor and certain combinations of its derivatives of an arbitrarily high order at a given point. The idea behind these divisions is as follows (Secs. 2 and 3): Combinations of derivatives of the metric tensor can be broken up into two groups, (1) those that form components of tensor and (2) all the other independent combinations. Tensor components cannot, in general, be chosen arbitrarily; if all components vanish in one frame, they vanish in all frames. The combinations belonging to group (2), however, can be specified at will and their values form a basis for a division of all coordinate systems into classes. All such classes turn out to be 10 parametric.

Having made the distinction between tensorial quantities, which describe the geometrical structure, and the above-mentioned nontensorial expressions, one can formulate the physical significance of 10-parametric sets of frames in Riemannian space-time as follows: *10-parametric classes arise naturally when the maximum freedom in assigning numerical values to nontensorial quantities is used.*

Section 4 deals with the set of all transformations between coordinate systems belonging to the same 10-parametric class. Such a set is, of course, a subset of the Einstein group; however, in general it does not form a group. For example, if A, B, C, and D are four frames belonging to the same class, $t\ (A \to B)$ and $t\ (C \to D)$ transformations from A to B and from C to D, respectively, then, if $B = C$,

$$t\ (A \to B) \cdot t\ (C \to D) = t\ (A \to D).$$

If, however, $B \neq C$, then, in general, $t\ (A \to B) \times t\ (C \to D)$ is not a transformation between frames belonging to the same class but transforms A into a different class.

The mathematical structure of these sets of transformations is defined in Sec. 4 as a "quasigroup." A quasigroup does not satisfy the closure requirement of a group. It does satisfy, however, all the other requirements: An identity element is contained in it; together with every element it contains its inverse;

and the associative law is satisfied. In the limit of flat space–time the closure requirement is met, and the quasigroup becomes identical with the Poincaré group.

As examples of quasigroups of transformations, the sets of transformations between normal coordinates and between geodesic Fermi coordinates are investigated; the general expression for an infinitesimal transformation between two normal coordinates is derived (Sec. 4).

In Appendix A we summarize some known results concerning Riemannian and normal coordinates which are used in this work. The material contained in Appendix B is essentially new: It contains an analytic characterization of geodesic Fermi coordinate systems in terms of nontensorial combinations of derivatives of $g_{\mu\nu}$ to an arbitrarily high order at the origin.

Let us emphasize that all the results of this work are independent of the theory of general relativity. No assumptions were made concerning the equations which determine the curvature of space–time.

All the theorems stated in this work will be subject to the same restrictions used by Veblen and Thomas in their classical paper[4]: Only analytic transformations between coordinate systems will be considered, and the metric tensor components are assumed to be analytic functions of the coordinates. These restrictions will not be restated in each theorem. For example, the expression "all coordinate systems," whenever used, refers only to coordinate systems satisfying these restrictions.

The Einstein summation convention is used. Repeated Greek indices imply summation over 1, 2, 3, 4; repeated Latin indices imply summation over 1, 2, 3.

2. CLASSIFICATION OF COORDINATE SYSTEMS TO 10-PARAMETRIC CLASSES

In flat space–time a 10-parametric class of coordinate systems can be defined by a definite choice of the metric tensor at all points, e.g.,

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.1}$$

defines the 10-parametric set of Cartesian coordinates.

In Riemannian space–time, the metric tensor cannot, in general, be given *a priori* at all points. $g_{\mu\nu}$ is determined by (a) the geometrical structure and (b) the choice of coordinate system. Investigation of derivatives of $g_{\mu\nu}$ to an arbitrarily high order at one point enables us to single out the dependence on the choice of frame. A complete determination of those combinations of derivatives which depend solely on choice of frame defines classifications of all coordinate systems to 10-parametric sets.

Mathematically, the simplest such classification, to be introduced in Theorem 1, is based on a set of nontensorial expressions introduced by Veblen and Thomas.[4] These expressions are defined as follows:

$$\Gamma^{\alpha}_{\beta\gamma\delta} = \tfrac{1}{3}P\left(\frac{\partial}{\partial x^{\delta}}\Gamma^{\alpha}_{\beta\gamma} - 2\Gamma^{\alpha}_{\xi\gamma}\Gamma^{\xi}_{\beta\delta}\right),$$

$$\Gamma^{\alpha}_{\beta\gamma\delta\epsilon} = \tfrac{1}{4}P\left(\frac{\partial}{\partial x^{\epsilon}}\Gamma^{\alpha}_{\beta\gamma\delta} - 3\Gamma^{\alpha}_{\xi\gamma\delta}\Gamma^{\xi}_{\beta\epsilon}\right)$$

and, in general,

$$\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu\nu} = \frac{1}{N}P\left(\frac{\partial}{\partial x^{\nu}}\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu} - (N-1)\Gamma^{\alpha}_{\xi\gamma\delta\cdots\mu}\Gamma^{\xi}_{\beta\nu}\right), \tag{2.2}$$

where $\Gamma^{\alpha}_{\beta\gamma}$ are the Christoffel symbols of the second kind, P means that all terms obtained by cyclic permutations of the subscripts should be added together, and N is the number of subscripts.

These expressions occur naturally if one considers the standard form of the equation of geodesics:

$$\frac{d^2xx^{\mu}}{ds^2} + \Gamma^{\mu}_{\alpha\beta}\frac{dx^{\alpha}}{ds}\frac{dx^{\beta}}{ds} = 0. \tag{2.3}$$

Differentiating Eqs. (2.3) successively with respect to s,we get, for all $n \geq 3$,

$$\frac{d^nx^{\mu}}{ds^n} + \Gamma^{\mu}_{\alpha_1\cdots\alpha_n}\frac{dx^{\alpha_1}}{ds}\cdots\frac{dx^{\alpha_n}}{ds} = 0. \tag{2.4}$$

The following classification utilizes the expressions (2.2):

Theorem 1: All coordinate systems can be divided into 10-parametric classes as follows: Each class is characterized by a set of constants $a_{\mu\nu}$, $\mu, \nu = 1, \cdots, 4$, and $b^{\mu}_{\alpha_1\cdots\alpha_n}$, $\mu, \alpha_1, \cdots, \alpha_n = 1, \cdots, 4$, $n \geq 2$, such that, for coordinate systems belonging to the class

$$(g_{\mu\nu})_0 = a_{\mu\nu}, \tag{2.5}$$

$$(\Gamma^{\mu}_{\alpha_1\cdots\alpha_n})_0 = b^{\mu}_{\alpha_1\cdots\alpha_n} \tag{2.6}$$

(a quantity with subscript 0 denotes the value of the quantity at the origin).

Remarks: (1) From Eqs. (2.5) and (2.6) it follows that the a's and b's are completely symmetric in the subscripts and that

$$\det a_{\mu\nu} \neq 0, \tag{2.7}$$

$$\operatorname{sig} a_{\mu\nu} \neq -2. \tag{2.8}$$

(2) This theorem and analogous theorems and statements are not global. They refer to finite but sufficiently small regions of space–time.

Proof: Let us establish a one-to-one correspondence between the set of all coordinate systems satisfying Eqs. (2.5) and (2.6) and the set of all Riemannian coordinate systems satisfying Eq. (2.5).

Let x^μ be a given frame satisfying Eqs. (2.5) and (2.6). Consider a geodesic through its origin with directions given by

$$\xi^\mu = \left(\frac{dx^\mu}{ds}\right)_0. \tag{2.9}$$

Expanding its Eqs. (2.3) in Taylor series, we get from Eqs.(2.4)

$$\begin{aligned} x^\mu &= \xi^\mu s + \sum_{n=2}^{\infty} \frac{1}{n!}\left(\frac{d^n x^\mu}{ds^n}\right)_0 s^n \\ &= \xi^\mu s - \sum_{n=2}^{\infty} \frac{1}{n!}(\Gamma^\mu_{\alpha_1\cdots\alpha_n})_0 \xi^{\alpha_1}\cdots\xi^{\alpha_n} s^n. \end{aligned} \tag{2.10}$$

Consider now the following transformation between the given coordinates x^μ and another set of coordinates y^μ:

$$\begin{aligned} x^\mu &= y^\mu + \sum_{n=2}^{\infty} \frac{1}{n!}\left(\frac{\partial^n x^\mu}{\partial y^{\alpha_1}\cdots\partial y^{\alpha_n}}\right)_0 y^{\alpha_1}\cdots y^{\alpha_n} \\ &= y^\mu - \sum_{n=2}^{\infty} \frac{1}{n!}(\Gamma^\mu_{\alpha_1\cdots\alpha_n})_0 y^{\alpha_1}\cdots y^{\alpha_n}. \end{aligned} \tag{2.11}$$

The Jacobian of the transformation is nonvanishing at the origin and, therefore, the transformation can be inverted. The inverse transformation is, in fact, given explicitly by

$$y^\mu = x^\mu + \sum_{n=2}^{\infty} (\Lambda^\mu_{\alpha_1\cdots\alpha_n})_0 x^{\alpha_1}\cdots x^{\alpha_n}, \tag{2.12}$$

where the Λ-symbols are defined by

$$\begin{aligned} &\Lambda^\alpha_{\beta\gamma} = \Gamma^\alpha_{\beta\gamma}, \\ &\Lambda^\alpha_{\beta\gamma\delta} = \Gamma^\alpha_{\beta\gamma\delta} + P(\Lambda^\alpha_{\mu\beta}\Gamma^\mu_{\gamma\delta}), \\ &\Lambda^\alpha_{\beta\gamma\delta\epsilon} = \Gamma^\alpha_{\beta\gamma\delta\epsilon} + P(\Lambda^\alpha_{\mu\beta}\Gamma^\mu_{\gamma\delta\epsilon} + \Lambda^\alpha_{\mu\nu}\Gamma^\mu_{\beta\gamma}\Gamma^\nu_{\delta\epsilon} + \Lambda^\alpha_{\mu\beta\gamma}\Gamma^\mu_{\delta\epsilon}), \\ &\qquad\qquad \text{etc.} \end{aligned} \tag{2.13}$$

P means that all the terms obtained by cyclic permutation of the subscripts should be added together.

From Eqs. (2.10) and (2.11) it follows that in the coordinate system y^μ all geodesics through the origin have the form

$$y^\mu = \xi^\mu s. \tag{2.14}$$

By definition the y^μ constitutes, therefore, a system of Riemannian coordinates (see Appendix A). Since the linear form of Eqs.(2.14) is conserved if and only if the coordinates undergo linear transformation, it follows that, given an arbitrary coordinate system x^μ, there exists one and only one Riemannian frame y^μ such that the transformation between x^μ and y^μ reduces to the identity transformation in first order. We have shown, therefore, that corresponding to an arbitrary coordinate system x^μ there exists one and only one Riemannian frame y^μ having the same origin and direction of axes. The components of the metric tensor at the origin are the same in both frames.

The transformation between x^μ and y^μ is given by Eqs.(2.11). From these equations it follows that x^μ satisfies Eqs. (2.5) and (2.6) if and only if it is connected with a Riemannian frame y^μ satisfying Eqs. (2.5) by

$$x^\mu = y^\mu - \sum_{n=2}^{\infty} \frac{1}{n!} b^\mu_{\alpha_1\cdots\alpha_n} y^{\alpha_1}\cdots y^{\alpha_n}. \tag{2.15}$$

Therefore, if numerical values are assigned to all the a's and b's [Eqs. (2.5) and (2.6)], then Eqs. (2.15) establish a one-to-one correspondence between all Riemannian frames that satisfy Eqs. (2.5) and all coordinate systems satisfying Eqs. (2.5) and (2.6). From this one-to-one correspondence and Theorem A3 of Appendix A it follows that the class of coordinate systems satisfying Eqs. (2.5) and (2.6) is 10 parametric. We have thus established a division of all coordinate systems into 10-parametric classes according to the values of $(g_{\mu\nu})_0$ and $(\Gamma^\mu_{\alpha_1\cdots\alpha_n})_0$. QED

The particular division thus established will be called the "normal division," because normal coordinate systems are obtained from Eqs. (2.5) and (2.6) by setting

$$a_{\mu\nu} = \eta_{\mu\nu}, \tag{2.16}$$

$$b^\mu_{\alpha_1\cdots\alpha_n} = 0. \tag{2.17}$$

The set of all normal frames is, therefore, one of the classes defined by this division.

As pointed out in Appendix A, the set of expressions

$$\Gamma^\mu_{\alpha_1\cdots\alpha_n}, \quad \mu, \alpha_1, \cdots, \alpha_n = i, \cdots, 4, \tag{2.18}$$

is not a tensor and does not contain any subset that forms a tensor. Since the geometrical structure of space–time remains unspecified, the classification was achieved by assignment of numerical values to a set of nontensorial expressions.

The normal division is based on the quantities (2.2). By different choices of nontensorial expressions it is possible to introduce different classification. In particular, "the Fermi division" is introduced in Theorem 2. This division might prove to be of greater

physical significance, since it contains as one of its classes the set of geodesic Fermi coordinates (see Appendix B).[5]

Theorem 2: All coordinate systems can be divided into 10-parametric classes as follows: Each class is characterized by a set of numbers $\alpha_{\mu\nu}$, $b^{\mu(k)}_{\alpha\beta}$, and $b^{\mu(k)}_{i_1\cdots i_n}$, $k \geq 0, n \geq 3, \mu, \alpha, \beta = 1, \cdots, 4, i_1, \cdots, i_n = 1, 2, 3$, such that for coordinate systems belonging to the class

$$(g_{\mu\nu})_0 = a_{\mu\nu}, \tag{2.19}$$

$$(\Gamma^{\mu}_{\alpha 4})_0 = b^{\mu(0)}_{\alpha 4}, \tag{2.20}$$

$$(\Gamma^{\mu}_{i_1\cdots i_n})_0 = b^{\mu(0)}_{i_1\cdots i_n}, \tag{2.21}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{\alpha 4}}{(\partial x^4)^k}\right)_0 = b^{\mu(k)}_{\alpha 4}, \tag{2.22}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{i_1\cdots i_n}}{(\partial x^4)^k}\right)_0 = b^{\mu(k)}_{i_1\cdots i_n}. \tag{2.23}$$

Remarks: (1) From Eqs. (2.19)–(2.23) it follows that all the a's and b's are completely symmetric in the subscripts and the a's satisfy Eqs. (2.7) and (2.8). (2) The order of a b symbol is defined as the number of its subscripts plus k. For example, $b^{\mu(0)}_{\alpha\beta}$ are second-order b symbols, $b^{\mu(0)}_{i_1 i_2 i_3}$ and $b^{\mu(1)}_{\alpha\beta}$ are third-order b symbols, etc.

Proof: In analogy with the proof of Theorem 1, the present theorem will be proven by establishing a one-to-one correspondence between all coordinate systems that satisfy Eqs. (2.19)–(2.23) with particular values of the a's and the b's and Riemannian frames satisfying Eqs. (2.19) with the same values of $a_{\mu\nu}$.

In the course of proving Theorem 1 we showed that any frame x^μ has one and only one corresponding Riemannian frame y^μ, having the same origin and directions of axes. The transformation between them is the identity transformation up to terms of the first order:

$$x^\mu = y^\mu + \sum_{n=2}^{\infty} \frac{1}{n!}\left(\frac{\partial^n x^\mu}{\partial y^{\alpha_1}\cdots\partial y^{\alpha_n}}\right)_0 y^{\alpha_1}\cdots y^{\alpha_n}. \tag{2.24}$$

By Eqs. (2.11)

$$\left(\frac{\partial^n x^\mu}{\partial y^{\alpha_1}\cdots\partial y^{\alpha_n}}\right)_0 = -(\Gamma^{\mu}_{\alpha_1\cdots\alpha_n})_0; \tag{2.25}$$

Eqs. (2.20) and (2.21) for $n = 2$ will be satisfied, therefore, if and only if the second-order b symbols are

$$b^{\mu(0)}_{\alpha\beta} = -\left(\frac{\partial^2 x^\mu}{\partial y^\alpha \partial y^\beta}\right)_0. \tag{2.26}$$

The third-order b symbols will now be determined by the coefficients of the $n = 3$ term of Eq. (2.24), namely $(\partial^3 x^\mu/\partial y^{\alpha_1}\partial y^{\alpha_2}\partial y^{\alpha_3})_0$: The $b^{\mu(0)}_{i_1 i_2 i_3}$ are obtained simply from Eqs. (2.25); Eq. (2.21) for $n = 3$ is satisfied if and only if

$$b^{\mu(0)}_{i_1 i_2 i_3} = -\left(\frac{\partial^3 x^\mu}{\partial y^{i_1}\partial y^{i_2}\partial y^{i_3}}\right)_0. \tag{2.27}$$

To obtain the $b^{\mu(1)}_{\alpha\beta}$ symbols, consider the transformation of the Christoffel symbols of the second kind:

$$\Gamma^{\mu}_{\sigma\rho}\frac{\partial x^\sigma}{\partial y^\alpha}\frac{\partial x^\rho}{\partial y^\beta} = \Gamma'^{\sigma}_{\alpha\beta}\frac{\partial x^\mu}{\partial y^\sigma} - \frac{\partial^2 x^\mu}{\partial y^\alpha \partial y^\beta} \tag{2.28}$$

(primed quantities referred to the y^μ coordinates, unprimed to the x^μ). Differentiating Eqs. (2.28) with respect to y^4, we obtain

$$\Gamma^{\mu}_{\sigma\rho,\nu}\frac{\partial x^\sigma}{\partial y^\alpha}\frac{\partial x^\rho}{\partial y^\beta}\frac{\partial x^\nu}{\partial y^4} + 2\Gamma^{\mu}_{\sigma\rho}\frac{\partial^2 x^\sigma}{\partial y^\alpha \partial y^4}\frac{\partial x^\rho}{\partial y^\beta} = \Gamma'^{\sigma}_{\alpha\beta,4}\frac{\partial x^\mu}{\partial y^\sigma} + \Gamma'^{\sigma}_{\alpha\beta}\frac{\partial^2 x^\mu}{\partial y^\sigma \partial y^4} - \frac{\partial^3 x^\mu}{\partial y^\alpha \partial y^\beta \partial y^4} \tag{2.29}$$

at the origin, since

$$\left(\frac{\partial x^\mu}{\partial y^\alpha}\right)_0 = \delta^\mu_\alpha, \tag{2.30}$$

and, using Eq. (A2) and Eq. (2.11), we have

$$(\Gamma^{\mu}_{\alpha\beta,4})_0 - 2\left(\frac{\partial^2 x^\mu}{\partial y^\sigma \partial y^\beta}\right)_0\left(\frac{\partial^2 x^\sigma}{\partial y^\alpha \partial y^4}\right)_0 = (\Gamma'^{\mu}_{\alpha\beta,4})_0 - \left(\frac{\partial^3 x^\mu}{\partial y^\alpha \partial y^\beta \partial y^4}\right)_0. \tag{2.31}$$

Thus, Eqs. (2.21) and (2.22) will be satisfied for $k = 1$ if and only if

$$b^{\mu(1)}_{\alpha\beta} = -\left(\frac{\partial^3 x^\mu}{\partial y^\alpha \partial y^\beta \partial y^4}\right)_0 + (\Gamma'^{\mu}_{\alpha\beta,4})_0 + 2\left(\frac{\partial^2 x^\mu}{\partial y^\sigma \partial y^\beta}\right)_0\left(\frac{\partial^2 x^\sigma}{\partial y^\alpha \partial y^4}\right)_0. \tag{2.32}$$

Proceeding by induction, one can express in this way all b symbols of the nth order in terms of the $(\partial^k x^\mu/\partial y^{\alpha_1}\cdots\partial y^{\alpha_k})_0$ for $k \leq n$. For the $b^{\mu(0)}_{i_1\cdots i_n}$ the result is simply (2.25):

$$b^{\mu(0)}_{i_1\cdots i_n} = -\left(\frac{\partial^n x^\mu}{\partial y^{i_1}\cdots\partial y^{i_n}}\right)_0. \tag{2.33}$$

For the $b^{\mu(k)}_{i_1\cdots i_n}$ with $k \geq 1$ the expression is more complicated. It is obtained by differentiating Eq. (2.28) $n - 2$ times and using the result at the origin. In analogy with (2.32) we obtain unique expressions. The one-to-one correspondence between the frames

x^μ satisfying Eqs. (2.19)–(2.23) and the Riemannian coordinates satisfying Eqs. (2.19) with the same numbers $a_{\mu\nu}$ is thus established. QED

One difference between normal and Fermi divisions is apparent by comparing Eqs. (2.25) and (2.32): For the normal division the relation between the b symbols and the coefficients of the transformation is independent of the geometrical structure; for the Fermi division, however, the relation involves terms like $(\Gamma'^{\mu}_{\alpha\beta,4})_0$ which will depend on the geometrical structure. This difference is of no real significance. It follows from the method of proof of Theorems 1 and 2: In both cases a one-to-one correspondence with *Riemannian* frames was established. The above-mentioned difference stems from the fact that all Riemannian frames with a given value of $(g_{\mu\nu})_0$ belong to the same class according to the normal division but to different classes according to the Fermi division.

It is shown in Appendix B that geodesic Fermi coordinates are characterized analytically by the equations

$$(g_{\mu\nu})_0 = \eta_{\mu\nu}, \tag{2.34}$$

$$(\Gamma^{\mu}_{\alpha 4})_0 = 0, \tag{2.35}$$

$$(\Gamma^{\mu}_{i_1\cdots i_n})_0 = 0, \tag{2.36}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{\alpha 4}}{(\partial x^4)^k}\right)_0 = 0, \tag{2.37}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{i_1\cdots i_n}}{(\partial x^4)^k}\right)_0 = 0. \tag{2.38}$$

The set of all geodesic Fermi coordinates is, therefore, one of the classes defined by the Fermi division.

In the two divisions considered it was shown that, without knowledge of the geometrical structure of space–time, it is possible to divide all coordinate systems into 10-parametric classes by assigning numerical values to nontensorial expressions and that different choices lead to different divisions. Still different divisions can be constructed, of course, by other choices of nontensorial expressions.

In both the normal and Fermi divisions the b symbols were uniquely determined by the coefficients $(\partial^n x^\mu/\partial y^{\alpha_1}\cdots\partial y^{\alpha_n})_0$. The general principle behind this unique determination can be explained as follows: Define as an nth-order quantity an expression which involves derivatives of $g_{\mu\nu}$ up to nth order and is linear in the nth-order derivatives. A set of nth-order quantities is defined as independent if no lower-order quantity can be formed from them by a linear combination. Any set of independent quantities can be broken up into two sets: (1) those that differ from components of tensors only by lower-order quantities and (2) all the other independent combinations. Tensor components cannot, in general, be chosen arbitrarily: If all the components vanish in one frame, they vanish in all frames. The combinations belonging to set (2), however, can be specified at will, and their values form a basis for a division of all coordinate systems into classes. The b symbols are such quantities. (It was pointed out in Appendices A and B that the b symbols of the normal and Fermi division do not contain subsets that form tensors.)

The crucial point is this: The number of independent nth-order quantities of set (2) is equal to the number of $(n+2)$-order derivatives $\partial^{n+2}x^\mu/\partial y^{\alpha_1}\cdots\partial y^{\alpha_{n+2}}$ which is $4\cdot(n+4)!/6\cdot(n+1)!$. Now, the total number of nth-order derivatives of $g_{\mu\nu}$ is $10(n+3)!/6\cdot n!$; and indeed, the difference

$$\frac{10}{6}\frac{(n+3)!}{n!} - \frac{4}{6}\frac{(n+4)!}{(n+1)!} = (n-1)(n+2)(n+3) \tag{2.39}$$

is the number of independent quantities in set (1), i.e., the number of components in nth-order tensors.

Examples: (i) $n=1$: $(n-1)(n+2)(n+3)=0$: No tensors can be built from $g_{\mu\nu}$ and its first-order derivatives; (ii) $n=2$: $(n-1)(n+2)(n+3)=20$: the number of independent components of the Riemann tensor; (iii) $n=3$: $(n-1)(n+2)(n+3)=60$: the number of independent covariant derivatives of the Riemann tensor components (the Bianchi identities reduce this number from 80 to 60!), etc.

Let us finally note the following feature of the divisions: In general, if the same transformations formula is applied to all elements of a given class, they will be mapped thereby into elements of several distinct classes. This is a consequence of the fact that the set of all transformations between elements of any class do not form a group. Indeed, in nonhomogeneous spaces application of the same transformation formulas to two coordinate systems with, say, different origins have, in general, different geometrical significance. The mathematical structure of the transformations between elements of the same class will be investigated in Sec. 4.

3. THE PHYSICAL SIGNIFICANCE OF 10-PARAMETRIC CLASSES OF COORDINATE SYSTEMS IN RIEMANNIAN SPACE-TIME

A coordinate system is a correspondence between events in space–time and sets of four numbers (x^1, x^2, x^3, x^4). Once such a correspondence is set up,

the metric tensor at all points of space–time is, in principle, measurable by a system of measurements that utilizes clocks and light signals.[5] Therefore, the metric tensor and all its derivatives to an arbitrarily high-order at one point can be determined by an appropriate set of measurements. Particular choices of division, as discussed in the previous section, correspond, therefore, to particular choices of sets of measurements.

A particular class of frames of reference as discussed in the previous section is thus defined according to preassigned values of the chosen set of expressions. For example, the class of geodesic Fermi frames is defined by preassignment of the metric tensor at a particular point as $\eta_{\mu\nu}$ and all expressions $\Gamma^{\mu}_{i_1\cdots i_n}$, $\partial^k\Gamma^{\mu}_{i_1\cdots i_n}/(\partial x^4)^k$ at this point as zero (Appendix B). We use the term "preassigned" because this assignment is made prior to determination of the geometrical structure.

The physical significance of the results of the previous section is now formulated as follows: A complete knowledge of $g_{\mu\nu}$ in a finite region around a chosen origin requires knowledge of $(g_{\mu\nu})_0$ and of all its derivatives to an arbitrarily high order at the origin. Such knowledge is equivalent to (1) knowing all the components of all the tensors that can be formed from these derivatives and (2) knowing all the rest, i.e., $(g_{\mu\nu})_0$ and a complete set of independent "nontensorial" expressions [such as (2.6) or (2.20)–(2.23)]. We realize now that we are free to preassign results of measurements of all the nontensorial quantities, and such preassignment of numerical values to all of them defines a 10-parametric class. In flat space–time the occurrence of 10-parametric classes is linked with the degree of symmetry of space–time. Here we realize that *10-parametric classes arise naturally when the maximum freedom in preassigning results of measurements of* $(g_{\mu\nu})_0$ *and of these nontensorial quantities is used.*

It follows now that if $g_{\mu\nu}$ and all its derivatives to an arbitrarily high order are measured at the origin of a given coordinate system, then a subset of these measurements, namely, $(g_{\mu\nu})_0$ and the nontensorial quantities, merely determines to which 10-parametric class of a chosen division the frame belongs.

In the flat space–time of special relativity it is possible to divide coordinate systems into 10-parametric classes by the direct physical significance of the coordinates. This is a consequence of the homogeneity of flat space–time. In an inhomogeneous curved space–time, coordinates do not possess simple metric meaning.[6] It is, however, possible to understand the physical meaning of systems of coordinates as a whole, rather than the numerical values of the coordinates, in terms of the above-mentioned sets of measurements. In the limit of flat space–time the two ways of characterizing frames of reference are equivalent. Since, however, the second way makes no reference to the geometry of space–time, it is directly generalizable to curved space–time, the geometrical structure of which remains unspecified.

4. QUASIGROUPS OF TRANSFORMATIONS

Let $c(P, \lambda)$ be a Cartesian coordinate system in flat space–time with origin at P and directions of axes devoted collectively by λ. Denote by $p\,(P, \lambda \to P', \lambda')$ the transformation from $c(P, \lambda)$ to $c(P', \lambda')$. $p\,(P, \lambda \to P', \lambda')$ belongs to the Poincaré group. *A priori*, it depends on the 20 parameters needed to specify P, λ, P', λ'. Since, however, the space–time under consideration is homogeneous, $p\,(P, \lambda \to P', \lambda')$ depends only on the relative positions of the origin and the relative orientation of the axes. The 20 parameters reduce to 10: The Poincaré group is 10 parametric.

In the general case of transformations between coordinate systems belonging to the same 10-parametric class in Riemannian space–time, this reduction does not occur, or occurs only partially: In general, the transformations depend on the initial and final positions and orientations separately, and the set of transformations does not form a group.

We proceed now to calculate explicitly the transformations belonging to one such set: the set of transformations between normal coordinate systems.

The general infinitesimal transformation between two normal coordinate systems will be derived in two stages:

(1) Let y^{μ} and y'^{μ} be two normal coordinate systems with origins at P and P', such that

$$\left(\frac{\partial y'^{\mu}}{\partial y^{\alpha}}\right)_0 = \delta^{\mu}_{\alpha}. \tag{4.1}$$

The transformation between them will be derived as follows: From the equations of transformation of Christoffel symbols

$$\Gamma'^{\mu}_{\sigma\rho}\frac{\partial y'^{\sigma}}{\partial y^{\alpha}}\frac{\partial y'^{\rho}}{\partial y^{\beta}} = \Gamma^{\sigma}_{\alpha\beta}\frac{\partial y'^{\mu}}{\partial y^{\sigma}} - \frac{\partial^2 y'^{\mu}}{\partial y^{\alpha}\partial y^{\beta}} \tag{4.2}$$

and from Eqs. (A2) and (4.1), it follows that

$$\left(\frac{\partial^2 y'^{\mu}}{\partial y^{\alpha}\partial y^{\beta}}\right)_P = -(\Gamma'^{\mu}_{\alpha\beta})_P. \tag{4.3}$$

By differentiating Eqs. (4.2) with respect to y^γ and summing over the cyclic permutations of α, β, γ, it likewise follows from Eqs. (A2) that

$$\left(\frac{\partial^3 y'^\mu}{\partial y^\alpha \partial y^\beta \partial y^\gamma}\right)_P = -(\Gamma'^\mu_{\alpha\beta\gamma})_P. \tag{4.4}$$

[The Γ symbols are defined by Eqs. (2.2).] Proceeding by induction, we obtain

$$\left(\frac{\partial^n y'^\mu}{\partial y^{\alpha_1} \cdots \partial y^{\alpha_n}}\right)_P = -(\Gamma'^\mu_{\alpha_1 \cdots \alpha_n})_P. \tag{4.5}$$

Therefore, the transformation $y'^\mu(y^\alpha)$ is given by

$$y'^\mu(y^\alpha) = y'^\mu(P) + y^\mu - \sum_{n=2}^{\infty} \frac{1}{n!}(\Gamma'^\mu_{\alpha_1 \cdots \alpha_n})_P y^{\alpha_1} \cdots y^{\alpha_n}. \tag{4.6}$$

Let us now specialize to the case of infinitesimal transformation. Let

$$y'^\mu(P) \equiv b^\mu. \tag{4.7}$$

Then, to first order in b^μ, by Eqs. (A2) of Appendix A

$$\begin{aligned}(\Gamma'^\mu_{\alpha_1 \cdots \alpha_n})_P &= (\Gamma'^\mu_{\alpha_1 \cdots \alpha_n})_{P'} + (\Gamma'^\mu_{\alpha_1 \cdots \alpha_n, \nu})_{P'} b^\nu \\ &= (\Gamma'^\mu_{\alpha_1 \cdots \alpha_n, \nu})_{P'} b^\nu,\end{aligned} \tag{4.8}$$

where ν denotes usual differentiation:

$$\Gamma'^\mu_{\alpha_1 \cdots \alpha_n, \nu} \equiv \frac{\partial \Gamma'^\mu_{\alpha_1 \cdots \alpha_n}}{\partial y'^\nu}. \tag{4.9}$$

By continuity we have, again up to first order in b^μ,

$$(\Gamma'^\mu_{\alpha_1 \cdots \alpha_n, \nu})_{P'} b^\nu = (\Gamma^\mu_{\alpha_1 \cdots \alpha_n, \nu})_P b^\nu, \tag{4.10}$$

$$b^\mu = y'^\mu(P) = -y^\mu(P') \equiv -a^\mu. \tag{4.11}$$

Equations (4.6) become, for infinitesimal transformations,

$$y'^\mu(y^\alpha) = y^\mu - a^\mu + D^\mu_\nu(y^\alpha) a^\nu, \tag{4.12}$$

where

$$D^\mu_\nu(y^\alpha) \equiv \sum_{n=2}^{\infty} \frac{1}{n!}(\Gamma^\mu_{\alpha_1 \cdots \alpha_n, \nu})_0 y^{\alpha_1} \cdots y^{\alpha_n}. \tag{4.13}$$

(2) In Appendix A (Theorem A1) we have shown that the set of transformations between normal coordinates with a fixed origin is identical with the homogeneous Lorentz group. An infinitesimal transformation between two normal frames y^μ and y'^μ having the same origin is, therefore, of the form

$$y'^\mu = y^\mu + \tfrac{1}{2} w^{\alpha\beta} (M_{\alpha\beta})^\mu_\rho y^\rho, \tag{4.14}$$

where $M_{\alpha\beta}$ are the 4×4 matrices which correspond to the infinitesimal transformations of the homogeneous Lorentz group and $w^{\alpha\beta}$ are the corresponding parameters.[7]

A general infinitesimal transformation is now obtained by a direct combination of Eqs. (4.12) and (4.14):

$$y'^\mu = y^\mu - a^\mu + \tfrac{1}{2} w^{\alpha\beta} (M_{\alpha\beta})^\mu_\rho y^\rho + D^\mu_\nu(y^\alpha) a^\nu. \tag{4.15}$$

We proceed now to discuss the set of all transformations between normal frames. This set N contains the homogeneous Lorentz group as a subset (Theorem A1, Appendix A), but N itself is different from the Poincaré group. In fact, N does not form a group at all because Eqs. (4.12) depend not only on the infinitesimal parameters but also on the geometrical structure [on the values of $(\Gamma^\mu_{\alpha_1 \ldots \alpha_n, \nu})_0$].

Let us denote by $n(P, \lambda)$ a normal coordinate system with origin at P and directions of axes denoted collectively by λ and by $n(P, \lambda \to P', \lambda') \in N$, the transformation from $n(P, \lambda)$ to $n(P', \lambda')$. Let us define multiplication of transformations in the usual way: If $n_1 \equiv n_1(P_1, \lambda_1 \to P_1', \lambda_1')$ is the transformation $y^\mu = y^\mu(x^\alpha)$ and $n_2 \equiv n_2(P_2, \lambda_2 \to P_2', \lambda_2')$ is the transformation $Z^\xi = Z^\xi(y^\mu)$, then

$$n_1 \cdot n_2 \equiv Z^\xi[y^\mu(x^\alpha)]. \tag{4.16}$$

In contradistinction to the case of flat space, it now follows from (4.12) that for a curved space–time, if $n_1, n_2 \in N$, $n_1 \cdot n_2 \in N$ is not necessarily true. If $P_1' = P_2$, then $n_1 \cdot n_2 \in N$. If, however, $P_1' \neq P_2$, then, in general, $n_1 \cdot n_2 \notin N$. It follows then that the set N is not a group. We call the mathematical structure exemplified by N a "quasigroup," which we define as follows[8]:

Definition: A set $A = \{a_\alpha\}$, where α stands for any number of discrete or continuous parameters, is a quasigroup if:

(1) Corresponding to every element $a_\alpha \in A$, $\exists\, B_\alpha \leq A$ and $\exists\, B'_\alpha \leq A$ such that if $b \in B_\alpha$, the multiplication $a_\alpha \cdot b$ is defined and $a_\alpha \cdot b \in A$ and if $b \in B'_\alpha$, then $b \cdot a_\alpha$ is defined and $b \cdot a \in A$;

(2) The associative law: for any $a, b, c \in A$ if $a \cdot b$ and $b \cdot c$ are defined, then $(a \cdot b) \cdot c$ and $a \cdot (b \cdot c)$ are also defined and

$$(a \cdot b) \cdot c = a \cdot (b \cdot c); \tag{4.17}$$

(3) Existence of unit element: Among the elements of A there is one and only one element e which has the property that $a \cdot e$ and $e \cdot a$ are defined for all $a \in A$ and

$$a \cdot e = e \cdot a = a; \tag{4.18}$$

(4) Existence of an inverse: Corresponding to every element $a \in A$, $\exists\, a' \in B_\alpha$, B'_α such that

$$a \cdot a' = a' \cdot a = e. \tag{4.19}$$

Thus a quasigroup is different from a group in that the product of two elements is not always defined. It resembles a group in the sense that the associative law is satisfied whenever the products are defined, and in having a unit element and an inverse to every element.

The quasigroup of transformation between normal frames contains the homogeneous Lorentz group. This is a consequence of the fact that transformations between normal coordinates with common origin depend only on the relative orientations of axes, (six parameters) not on the initial and final orientations separately. As far as the positions of origins are concerned, however, they do depend on the initial and final positions (eight parameters), not only on the relative positions. The quasigroup N is, therefore, 14 parametric. In the limit of flat space–time, the separate dependence on initial and final positions degenerates into a dependence on the relative positions only, and N becomes identical with the Poincaré group.

As a second example of a quasigroup, consider the set G of transformation between geodesic Fermi coordinate systems $g(P, \lambda)$.

We define a one-to-one correspondence between the elements of the 10-parametric set of normal frames and the elements of the 10-parametric set of geodesic Fermi coordinates as

$$g(P, \lambda) \leftrightarrow n(P, \lambda), \tag{4.20}$$

where P and λ are the same for g and n, i.e., the normal and geodesic Fermi frames that correspond to each other have the same origin and directions of axes. The transformation between the corresponding frames x^μ (geodesic Fermi) and y^μ (normal) was derived in Sec. 2 [Eqs. (2.24) and (2.25)]:

$$x^\mu = y^\mu - \sum_{n=2}^{\infty} (\Gamma^\mu_{\alpha_1 \cdots \alpha_n})_0 y^{\alpha_1} \cdots y^{\alpha_n}. \tag{4.21}$$

The inverse transformation is given by Eq. (2.12).

The transformation between two geodesic Fermi frames $g(P, \lambda)$ and $g(P', \lambda')$ can now be carried out in three steps:

$$g(P, \lambda) \to n(P, \lambda) \to n(P', \lambda') \to g(P', \lambda'),$$

where the first and last steps are given by Eqs. (4.21) and (2.12), respectively, and the second step is given by successive infinitesimal transformations (4.15). It can be shown that the quasigroup G contains the 1-dimensional group of time displacements and the rotation group, i.e., the transformation of time displacements of the origin depends only on the magnitude of displacement and the transformations corresponding to pure rotations depend only on the relative orientations of the axes. The quasigroup G is, therefore, 16 parametric. In the limit of flat space–time it, too, becomes identical with the Poincaré group.

ACKNOWLEDGMENTS

We are grateful to Professor A. O. Barut and Professor R. Richtmayer for stimulating discussions and helpful suggestions.

APPENDIX A: RIEMANNIAN AND NORMAL COORDINATES

A coordinate system is called *Riemannian* if all the geodesics through the origin can be written in the form

$$y^\mu = \xi^\mu s, \tag{A1}$$

where the ξ^μ are constants. It follows from Eqs.(2.10) that a coordinate system is Riemannian if and only if for all $n \geq 2$, $\mu, \alpha_1, \cdots, \alpha_n = 1, \cdots, 4$,

$$(\Gamma^\mu_{\alpha_1 \cdots \alpha_n})_0 = 0. \tag{A2}$$

The Γ symbols were defined in Sec. 2.

It follows from Eqs. (A2) that for any $n \geq 2$ the set of combinations of derivatives of the metric tensor

$$\{\Gamma^\mu_{\alpha_1 \cdots \alpha_n}\}, \quad \mu, \alpha_1, \cdots, \alpha_n = 1, \cdots, 4,$$

does not form a tensor and does not contain any subset that forms a tensor. Indeed, a tensor that vanishes in one frame vanishes in all frames. None of the expressions $\Gamma^\mu_{\alpha_1 \cdots \alpha_n}$ is identically zero (i.e., vanishes irrespective of the geometrical structure and choice of coordinate system), and yet they all vanish in Riemannian frames.

Consider an arbitrary Riemannian frame and denote its metric tensor at the origin by $(g_{\mu\nu})_0$. Consider the quadratic form $(g_{\mu\nu})_0 y^\mu y^\nu$. According to Sylvester's law of inertia there exists a linear transformation with real coefficients

$$y'^\mu = c^\mu_\alpha y^\alpha, \quad \det |c^\mu_\alpha| \neq 0, \tag{A3}$$

such that in the primed system of coordinates the coefficients of the quadratic form are $\pm\delta_{\mu\nu}$; the difference between the number of + and − along the diagonal being equal to the signature, i.e., −2:

$$(g_{\mu\nu})_0 y^\mu y^\nu = \eta_{\mu\nu} y'^\mu y'^\nu, \tag{A4}$$

where $\eta_{\mu\nu}$ is defined in Eq. (2.1).

The transformation (A3) does not effect the form of Eq. (A1). The y'^μ coordinates are, therefore, Riemannian.

Riemannian coordinates that satisfy, in addition to Eqs. (A2), also

$$(g_{\mu\nu})_0 = \eta_{\mu\nu} \tag{A5}$$

are called *normal coordinates.*

Theorem A1: The set of all normal coordinate systems having the same origin is six parametric, and the set of transformations between them is identical with the homogeneous Lorentz group.

Proof: Let y^μ be normal coordinates. The most general transformation from y^μ to another system of Riemannian coordinates y'^μ with the same origin is given by Eqs. (A3). This is a consequence of the fact that the form of Eq. (A1) is conserved if and only if the coordinates undergo linear transformation. The coordinates y'^μ are normal if and only if

$$(g'_{\mu\nu})_0 = \eta_{\mu\nu} = (g_{\alpha\beta})_0 \frac{\partial y^\alpha}{\partial y'^\mu}\frac{\partial y^\beta}{\partial y'^\nu} = \eta_{\alpha\beta}\frac{\partial y^\alpha}{\partial y'^\mu}\frac{\partial y^\beta}{\partial y'^\nu}. \tag{A6}$$

These equations are equivalent to

$$(g_{\mu\nu})_0 = \eta_{\mu\nu} = (g'_{\alpha\beta})\frac{\partial y'^\alpha}{\partial y^\mu}\frac{\partial y'^\beta}{\partial y^\nu} = \eta_{\alpha\beta}c^\alpha_\mu c^\beta_\nu. \tag{A7}$$

The set of 4×4 matrices C with elements C^α_β satisfying Eqs. (A7) is precisely the homogeneous Lorentz group. This group is six parametric, and each matrix C corresponds, according to Eqs. (A3), to one transformation between the given normal coordinates y^μ and another normal frame y'^μ having the same origin. QED

Theorem A1 dealt with the set of normal coordinates having the same origin. Since any point in space can be chosen as origin, the choice of the four coordinates of the origin allows for four additional parameters. We thus have the following theorem:

Theorem A2: The set of all normal coordinates in space–time is 10 parametric.

Geometrically, after a choice of origin (four parameters) has been made, the choice of the metric tensor at the origin specifies the relative directions of the four unit vectors in the directions of the coordinate axes. Equations (A5), for example, mean that this tetrad of unit vectors should be orthogonal (i.e., the directions of the coordinate axes in a normal coordinate system are mutually perpendicular). The choice of $(g_{\mu\nu})_0$ does not specify, however, the absolute orientation of the tetrad in space: A rotation of the tetrad as a whole does not effect $(g_{\mu\nu})_0$. Since the group of rotations in 4 dimensions is six parametric (considering Lorentz transformations as complex rotations), the total number of parameters is 10. Theorem A2 is, therefore, readily generalized as follows:

Theorem A3: Corresponding to any symmetric matrix $a_{\mu\nu}$ with signature -2 and nonvanishing determinant, there exists a 10-parametric set of Riemannian coordinate systems such that

$$(g_{\mu\nu})_0 = a_{\mu\nu}. \tag{A8}$$

Proof: We have previously seen that, corresponding to any set of coefficients of a quadratic form with signature -2 and nonvanishing determinant, there exists a transformation (A3) to a quadratic form with coefficients $\eta_{\mu\nu}$. Since the Jacobian of the transformation is nonvanishing, the inverse transformation exists, and, when applied to normal coordinates, it transforms the metric tensor at the origin from $\eta_{\mu\nu}$ to $a_{\mu\nu}$.

In analogy with (A7), the transformations (A3) that conserve $(g_{\mu\nu})_0$ satisfy

$$a_{\mu\nu} = a_{\alpha\beta}c^\alpha_\mu c^\beta_\nu. \tag{A9}$$

This is a set of 10 equations for 16 unknowns. Therefore, its real solutions are at most six parametric. Choose one particular transformation, $\bar{c}^\mu_\nu$ between normal coordinates and Riemannian coordinates satisfying Eqs. (A8). By successive application of an arbitrary homogeneous Lorentz transformation and the transformation $\bar{c}^\mu_\nu$, a correspondence between all normal frames and all Riemannian frames satisfying Eqs. (A8) is established. Since the former is six parametric (Theorem A1), so is the latter. Allowance of four parameters for choice of origin completes the proof.

In Appendix B we shall make use of 3-dimensional normal coordinates in a spacelike hypersurface. In particular, we need the following:

Theorem A4: The set of all normal frames in a 3-dimensional hypersurface is six parametric. A normal frame is uniquely defined once an origin (three parameters) and three mutually perpendicular directions of axes (three parameters) have been chosen.

This is the analog of Theorem A2 for 3- instead of 4-dimensional space.

APPENDIX B: GEODESIC FERMI COORDINATES

Mathematically, the simplest 10-parametric set of coordinate systems is the set of all 4-dimensional

normal frames of reference. As Synge[9] points out, however, 4-dimensional normal coordinates are unsatisfactory from a physical standpoint. The difficulty can be stated as follows: The time axis of a normal frame of reference is a timelike geodesic which satisfies

$$(g_{\mu\nu})_0 = \eta_{\mu\nu}, \tag{B1}$$

$$[\alpha\beta, \mu]_0 = (\Gamma^{\mu}_{\alpha\beta})_0 = 0, \tag{B2}$$

where $(g_{\mu\nu})_0[\alpha\beta, \mu]$ and $(\Gamma^{\mu}_{\alpha\beta})_0$ are the values of the metric tensor and the Christoffel symbols of the first and second kind at the origin (0, 0, 0, 0). The corresponding equations are not satisfied, however, at the spatial origin at times other than $t = 0$, i.e., in general, for $t \neq 0$,

$$g_{\mu\nu}(0, 0, 0, t) \neq \eta_{\mu\nu}, \tag{B3}$$

$$\Gamma^{\mu}_{\alpha\beta}(0, 0, 0, t) \neq 0. \tag{B4}$$

Thus, in general, a normal coordinate system is not locally Cartesian at the spatial origin (except at time $t = 0$), and, more important, its properties around the spatial origin *change with time*. Physically, it corresponds, therefore, to an observer whose system of measurement changes continuously as time goes by.

This difficulty comes about because time and space coordinates are treated in the same way in the definition of normal coordinates: Equations (A2) and (A5) are completely symmetrical in time and space variables. In reality, however, the nature of our measurements is such that the time axis is distinguished: The observer is constrained to move along it as he takes his measurements.

In contradistinction to normal coordinates, geodesic Fermi coordinates[9] take this special role of the physical observer into consideration. They are defined as follows:

Definition: A geodesic Fermi frame is a coordinate system: (i) It is locally Cartesian along all points of its time axis, i.e.,

$$g_{\mu\nu}(0, 0, 0, x^4) = \eta_{\mu\nu}, \quad -\infty < x^4 < \infty, \tag{B5}$$

$$\Gamma^{\mu}_{\alpha\beta}(0, 0, 0, x^4) = 0, \quad -\infty < x^4 < \infty. \tag{B6}$$

(ii) All its hypersurfaces $x^4 = c$ (for all real numbers c) are geodesic hypersurfaces[10] perpendicular to the time axis, and the coordinates induced on them by setting $x^4 = c$ are 3-dimensional normal coordinates.

Theorem B1: A geodesic Fermi frame is uniquely determined by choice of a point for its origin and of four mutually perpendicular directions at this point (three spacelike and one timelike) for directions of its axes.

Proof: The following properties follow from the definition of quasi-Lorentz frames:

(a) Its time axis is a geodesic because of Eqs. (B6); the time axis

$$x^i = 0, \quad x^4 = s \tag{B7}$$

satisfies the equations of geodesics (2.3).

(b) The 3-dimensional normal coordinates of any hypersurface $x^4 = c$ are such that their x^1, x^2, and x^3 directions at (0, 0, 0, c) are parallel to the x^1, x^2, and x^3 directions at (0, 0, 0, 0) in Levi-Civita's sense of parallelism.

Indeed, by the definition of parallel transfer, the change in the components of any vector R^{μ}, when displaced parallel to itself along an elementary path dx^{β}, is given by

$$dR^{\mu} = -\Gamma^{\mu}_{\alpha\beta} R^{\alpha} dx^{\beta}. \tag{B8}$$

It follows, therefore, from Eq. (B6) that the components of the unit vectors in the x^1, x^2, and x^3 directions do not change by a parallel displacement along the elementary path (0, 0, 0, dx^4).

Given an origin and four mutually perpendicular directions, it follows from (a) that the time axis is uniquely determined as the timelike geodesic in the given timelike direction. By requirement (ii) of the definition all the spacelike hypersurfaces $x^4 = c$ are uniquely determined. From Theorem A4, the three given spacelike directions uniquely determine a normal frame of reference on the geodesic hypersurface $x^4 = 0$ and, by (b) and Theorem A4, once the normal coordinates on $x^4 = 0$ are fixed, the normal coordinates on all surfaces $x^4 = c$ are uniquely determined. QED

Corollary: The set of geodesic Fermi frames in space–time is 10 parametric.

The following theorem amounts to an alternative definition of geodesic Fermi frames. It exhibits the similarities and differences between geodesic Fermi and normal coordinates.

Theorem B2: A coordinate system is a geodesic frame if and only if it is locally Cartesian at all points of the time axis and for all real values of the numbers c^1, c^2, c^3, and c the lines

$$x^i = c^i s, \quad i = 1, 2, 3, \quad x^4 = 0, \tag{B9}$$

where s is the invariant distance, are geodesics.

Proof: We have to show that requirement (ii) of the definition of geodesic Fermi frames is satisfied if and only if all lines of the form (B9) are geodesics.

If (ii) is satisfied, then the hypersurface $x^4 = c$ is geodesic, and the 3-dimensional coordinate systems x^1, x^2, x^3 induced on it by the geodesic Fermi system by setting $x^4 = c$ are normal. By definition of 3-dimensional normal frame all lines of the form

$$x^i = c^i s, \quad i = 1, 2, 3, \tag{B10}$$

are geodesics in the hypersurface. The additional requirement $x^4 = c$ insures that the lines belong to the hypersurface, and from the definition of geodesic hypersurface it follows that (B9) are geodesics in the 4-dimensional space–time.

Conversely, if in a given surface all lines satisfying Eqs.(B10) are geodesics, the coordinate system on that surface is normal. By Eqs.(B9) all the lines generating any hypersurface $x^4 = c$ are geodesics, and all such hypersurfaces are, therefore, geodesics. QED

The following theorem gives an analytic characterization of geodesic Fermi frames in terms of the metric tensor and certain combinations of its derivatives at the origin.

Theorem B3: A system of coordinates is a geodesic Fermi frame if and only if the following conditions are met for all $n \geq 2$, $k \geq 1$, $\mu, \nu, \alpha = 1, \cdots, 4$ and $i_1, \cdots, i_n = 1, 2, 3$:

$$(g_{\mu\nu})_0 = \eta_{\mu\nu}, \tag{B11}$$

$$(\Gamma^{\mu}_{\alpha 4})_0 = 0, \tag{B12}$$

$$(\Gamma^{\mu}_{i_1 \cdots i_n})_0 = 0, \tag{B13}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{\alpha 4}}{(\partial x^4)^k}\right)_0 = 0, \tag{B14}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{i_1 \cdots i_n}}{(\partial x^4)^k}\right)_0 = 0. \tag{B15}$$

Proof: We divide the proof into two parts. In part (1) we show that a system of coordinates is locally Cartesian at all points of the time axis if and only if

$$(g_{\mu\nu})_0 = \eta_{\mu\nu}, \tag{B16}$$

$$(\Gamma^{\mu}_{\alpha\beta})_0 = 0, \tag{B17}$$

$$\left(\frac{\partial^k \Gamma^{\mu}_{\alpha\beta}}{(\partial x^4)^k}\right)_0 = 0, \tag{B18}$$

for all $k \geq 1$, $\mu, \alpha, \beta = 1, \cdots, 4$. In part (2) we show that, for all real values of c^1, c^2, c^3, and c, the lines (B9) are geodesics if and only if Eqs. (B13) and (B15) are satisfied for all $k \geq 1$, $n \geq 2$, $\mu = 1, \cdots, 4$; $i_1, \cdots, i_n = 1, 2, 3$. Because of Theorem B2 this will complete the proof.

(1) By definition a system is locally Cartesian at all points of the time axis if for all x^4

$$g_{\mu\nu}(0, 0, 0, x^4) = \eta_{\mu\nu} \tag{B19}$$

and

$$\Gamma^{\mu}_{\alpha\beta}(0, 0, 0, x^4) = 0. \tag{B20}$$

Let us show that Eqs. (B19) and (B20) are equivalent to Eqs. (B6)–(B8). First, expanding $\Gamma^{\mu}_{\alpha\beta}(0, 0, 0, x^4)$ around the origin,

$$\Gamma^{\mu}_{\alpha\beta}(0, 0, 0, x^4) = (\Gamma^{\mu}_{\alpha\beta})_0 + \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\partial^k \Gamma^{\mu}_{\alpha\beta}}{\partial x^4}\right)_0 (x^4)^k, \tag{B21}$$

we see that Eqs. (B20) are equivalent to Eqs. (B17) and (B18). In the continuation we use the equivalence of Eqs. (B17) and (B18) to

$$[\alpha\beta, \mu]_0 = 0, \tag{B22}$$

$$\left(\frac{\partial^k [\alpha\beta, \mu]}{(\partial x^4)^k}\right)_0 = 0. \tag{B23}$$

This equivalence is proved as follows: At any point P all Christoffel symbols of the second kind vanish if and only if all Christoffel symbols of the first hand vanish (since any of these sets of symbols vanish if and only if all first-order derivatives of the metric tensor vanish at the point). Therefore, Eqs. (B20) hold if and only if

$$[\alpha\beta, \mu](0, 0, 0, x^4) = [\alpha\beta, \mu]_0 + \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\partial^k [\alpha\beta, \mu]}{(\partial x^4)^k}\right)_0 (x^4)^k = 0 \tag{B24}$$

and Eqs. (B7) and (B8) are equivalent to Eqs. (B22) and (B23), respectively.

We are now ready to show that Eqs. (B19) and (B20) follow from (B16)–(B18) and vice versa: Expanding $g_{\mu\nu}(0, 0, 0, x^4)$ around the origin, we get

$$\begin{aligned} g_{\mu\nu}(0, 0, 0, x^4) &= (g_{\mu\nu})_0 + \left(\frac{\partial g_{\mu\nu}}{\partial x^4}\right)_0 x^4 + \sum_{n=2}^{\infty} \frac{1}{n!} \left(\frac{\partial^n g_{\mu\nu}}{(\partial x^4)^n}\right)_0 (x^4)^n \\ &= \eta_{\mu\nu} + ([\mu 4, \nu] + [\nu 4, \mu])_0 x^4 \\ &\quad + \sum_{n=2}^{\infty} \frac{1}{n!} \left\{\frac{\partial^{n-1}}{(\partial x^4)^{n-1}} ([\mu 4, \nu] + [\nu 4, \mu])\right\}_0 (x^4)^n. \end{aligned} \tag{B25}$$

If Eqs. (B16)–(B18) are satisfied, it follows from Eqs. (B21), (B25), and the equivalence of (B22), (B23) to (B17), (B18) that Eqs. (B19) and (B20) are satisfied. Conversely, if Eqs. (B19), (B20) hold, it follows from Eqs.(B21) that Eqs. (B17) and (B18) are true; Eqs. (B16) too follow now because of the above-mentioned equivalence and Eqs.(B25).

(2) Consider the equations of geodesics (2.3). From these equations it follows that all lines (B9) are geodesics if and only if $\Gamma^\mu_{ij}c^ic^j$ vanish along them for $\mu = 1, \cdots, 4$. Expanding $\Gamma^\mu_{ij}c^ic^j$ around the point $(0, 0, 0, c)$ along a given line, we have

$$\Gamma^\mu_{ij}(s)c^ic^j = \Gamma^\mu_{ij}(0, 0, 0, c)c^ic^j + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{d^n\Gamma^\mu_{ij}}{ds^n}(0, 0, 0, c)c^ic^js^n. \quad \text{(B26)}$$

Now,

$$\frac{d\Gamma^\mu_{ij}}{ds} = \Gamma^\mu_{ij,k}c^k, \quad \text{(B27)}$$

$$\frac{d\Gamma^\mu_{ij}}{ds}(0, 0, 0, c)c^ic^j = \Gamma^\mu_{ij,k}(0, 0, 0, c)c^ic^jc^k. \quad \text{(B28)}$$

i, j, and k are dummy indices, and so we can permute them and add up all the permutations to get[11]

$$\frac{d\Gamma^\mu_{ij}}{ds}(0, 0, 0, c)c^ic^j = \Gamma^\mu_{ijk}(0, 0, 0, c)c^ic^jc^k. \quad \text{(B29)}$$

Thus a necessary and sufficient condition for the first term in the sum to vanish is

$$\Gamma^\mu_{ijk}(0, 0, 0, c) = 0. \quad \text{(B30)}$$

Similarly, for any other term

$$\frac{d^n\Gamma^\mu_{ij}}{ds^n} = \Gamma^\mu_{ij,k_1\cdots k_n}c^{k_1}\cdots c^{k_n} \quad \text{(B31)}$$

$$\frac{d^n\Gamma^\mu_{ij}}{ds^n}(0, 0, 0, c)c^ic^j = \Gamma^\mu_{ij,k_1\cdots k_n}(0, 0, 0, c)c^ic^jc^{k_1}\cdots c^{k_n} = \Gamma^\mu_{ijk_1\cdots k_n}(0, 0, 0, c)c^ic^jc^{k_1}\cdots c^{k_n}. \quad \text{(B32)}$$

Thus all terms in the expression vanish if and only if, for all real values of c, $\mu = 1, \cdots, 4$, $n \geq 2$, $i_1, \cdots, i_n = 1, 2, 3$,

$$\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, c) = 0. \quad \text{(B33)}$$

Expanding $\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, c)$ around the origin, we have

$$\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, c) = (\Gamma^\mu_{i_1\cdots i_n})_0 + \sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{\partial^k\Gamma^\mu_{i_1\cdots i_n}}{(\partial x^4)^k}\right)_0 c^k. \quad \text{(B34)}$$

Thus, Eqs. (B33) are equivalent to Eqs. (B13) and (B15). QED

Corollary 1: For any $n \geq 2$, the set of combinations of derivations of the metric tensor,

$$\left\{\Gamma^\mu_{i_1\cdots i_n}, \frac{\partial^k\Gamma^\mu_{i_1\cdots i_{n-k}}}{(\partial x^4)^k}\right\}, \quad \mu = 1, \cdots, 4, \quad i_1, \cdots, i_n = 1, 2, 3, \quad k = 1, \cdots, n-2,$$

does not form a tensor and does not contain any subset that forms a tensor.

This assertion is proved in complete analogy with a similar assertion following Eqs. (A2).

Corollary 2: If x^μ are geodesic Fermi coordinates, then x'^μ defined by

$$x'^i = x^i, \quad i = 1, 2, 3, \quad x'^4 = x^4 + c \quad \text{(B35)}$$

(c any real constant) are also geodesic Fermi coordinates.

In the course of proving Theorem B3, we showed Eqs. (B11)–(B13) to be equivalent to (B19), (B20), and (B23). If these equations are true for x^μ, they are also true for x'^μ defined by (B35).

Corollary 3: In the limit of flat space–time, geodesic Fermi coordinates reduce to Cartesian frames.

Since Cartesian frames are defined by Eqs. (2.1) at all points, Eqs. (B11)–(B15) are satisfied.

An alternative analytic characterization of quasi-Lorentz frames is given in the following.

Theorem B4: Equations (B13) and (B15) are satisfied if and only if at all points

$$\Gamma^\mu_{ij}(x^1, x^2, x^3, x^4)x^ix^j = 0, \quad \mu = 1, \cdots, 4. \quad \text{(B36)}$$

Proof: Expand (B36) around $(0, 0, 0, x^4)$:

$$\Gamma^\mu_{ij}(x^1, x^2, x^3, x^4)x^ix^j = \sum_{k=0}^{\infty} \frac{1}{k!}\Gamma^\mu_{ij,i_1\cdots i_k}(0, 0, 0, x^4)x^ix^jx^{i_1}\cdots x^{i_k} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!}\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, x^4)x^{i_1}\cdots x^{i_n}. \quad \text{(B37)}$$

The last step follows from the definition of the Γ symbols [Eqs. (2.2)], in analogy with the derivation of Eqs. (B31). Thus, Eqs. (B36) are satisfied if and only if for all $n \geq 2$ and all values of x^4,

$$\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, x^4) = 0, \quad \mu = 1, \cdots, 4, \quad i_1, \cdots, i_n = 1, 2, 3. \quad \text{(B38)}$$

Expanding now $\Gamma^\mu_{i_1\cdots i_n}(0, 0, 0, x^4)$ in Taylor's series around the origin, Eqs. (B38) are equivalent to

$$(\Gamma^\mu_{i_1\cdots i_n})_0 = 0, \quad \text{(B39)}$$

$$\left(\frac{\partial^k\Gamma^\mu_{i_1\cdots i_n}}{(\partial x^4)^k}\right)_0 = 0, \quad \text{(B40)}$$

for all $k \geq 1$, $\mu = 1, \cdots, 4$, $i_1 \cdots i_n = 1, 2, 3$. QED

* This work has been supported in part by the Colgate Research Council and the Sloan Foundation.

[1] E. Wigner, Ann. Math. **40**, 149 (1939).

[2] The terms "coordinate system," "frame of reference," and "frame" will be used interchangeably.

[3] The set of Cartesian frames is 10 parametric because a Cartesian coordinate system is uniquely determined once the origin (four parameters) and four mutually orthogonal directions in space–time (six parameters) are given.

[4] O. Veblen and T. Y. Thomas, Trans. Am. Math. Soc. **25**, 551 (1923).

[5] J. L. Synge, *Relativity, The General Theory* (North-Holland, Amsterdam, 1966), pp. 103–09.

[6] A. Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, N.J., 1960), 3rd ed., p. 133.

[7] S. S. Schweber, *An Introduction to Relativistic Quantum Mechanics* (Harper and Row, New York, 1962), p. 40.

[8] To the best of our knowledge such a structure has never been defined before.

[9] Reference 5, pp. 83–85. Synge calls these frames simply "Fermi coordinates." This name usually refers, however, to a much bigger class of coordinate systems, namely, those that satisfy $g_{\mu\nu} = \eta_{\mu\nu}$ and $\Gamma^{\alpha}_{\beta\gamma} = 0$ along a given curve [E. Fermi, Atti Accad. Naz. Lincei Rend. Classe Sci. Fis. Mat. Nat. **31**, 21, 51 (1922); T. Levi-Civita, Math. Ann. **92**, 291 (1926); L. O'Raifeartaigh, Proc. Roy. Irish Acad. **A59**, 15 (1958); L. O'Raifeartaigh and J. L. Synge, Proc. Roy. Soc. (London) **A246**, 299 (1958)]. We prefer, therefore, to use the term "geodesic Fermi coordinates" suggested by A. Schield [*Relativity Theory and Astrophysics, I. Relativity and Cosmology, Lectures of Applied Mathematics* (American Mathematical Society, Providence, R.I., 1967), Vol. 8]. Geodesic Fermi coordinates were recently used by J. L. Anderson, "Maximal Covariance Conditions and Kretschmann's Relativity Group," in *Prospectives in Geometry and Relativity*, B. Hoffman, Ed. (Indiana U.P., Bloomington, 1966).

[10] For definition and discussion of geodesic manifolds see T. Levi-Civita, *The Absolute Differential Calculus* (Blackie, London, 1954), p. 162.

[11] Having proved part (1), we can assume Eqs. (B21) to hold. Equations (2.2) reduce then to the form

$$\Gamma^{\alpha}_{\beta\gamma\delta} = \tfrac{1}{3}P\left(\frac{\partial}{\partial x^{\delta}}\,\Gamma^{\alpha}_{\beta\gamma}\right),$$

$$\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu\nu} = \frac{1}{N}P\left(\frac{\partial}{\partial x^{\mu}}\,\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu}\right).$$

This is the form used in deriving Eqs. (B23) and (B33).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Infinite-Dimensional Representations of the Lorentz Group: Complementary Series of Representations

M. CARMELI
General Physics Research Laboratory, Aerospace Research Laboratories
Wright-Patterson AFB, Ohio 45433

AND

S. MALIN*
Department of Physics and Astronomy, Colgate University, Hamilton, New York 13346

(Received 24 July 1970)

The method used by Carmeli to obtain another form for the principal series of representations of the group $SL(2, C)$ is extended to the complementary series of representations of that group. As a result, a new form for the complementary series of representations of $SL(2, C)$ is obtained which describes the transformation law of an infinite set of numbers under the group translation in a way which is very similar, but as a generalization, to the way spinors appear in the finite-dimensional case.

1. INTRODUCTION

Recently, Carmeli[1] has introduced an infinite set of quantities which are associated with the principal series of representations of the group $SL(2, C)$ in a way which is very similar, but as a generalization, to the way spinors[2] appear in describing the finite-dimensional representations. The transformation law of these quantities,[3] at the same time, defines a new form of the principal series of representations of $SL(2, C)$.

The principal series of representations, however, do not realize all irreducible unitary representations of the group $SL(2, C)$. Rather, every irreducible unitary representation of the group $SL(2, C)$ is unitarily equivalent to a representation of the principal series or the complementary series of representations.[4]

In this paper we extend Carmeli's result to the complementary series of representations, thus establishing new forms for all irreducible unitary representations (to within unitary equivalence) of the group $SL(2, C)$.

In Sec. 2 we summarize the method used for the principal series of representations. In Sec. 3 we generalize the method to the complementary series of representations. The Appendix is devoted to detailed calculation of a normalization factor introduced in the text.

2. SUMMARY OF PREVIOUS WORK

In this section we discuss Carmeli's[1] form for the principal series of representations.

We denote by $L_2^{2s}(SU_2)$ the set of all functions $\phi(u)$,

* This work has been supported in part by the Colgate Research Council and the Sloan Foundation.

[1] E. Wigner, Ann. Math. **40**, 149 (1939).

[2] The terms "coordinate system," "frame of reference," and "frame" will be used interchangeably.

[3] The set of Cartesian frames is 10 parametric because a Cartesian coordinate system is uniquely determined once the origin (four parameters) and four mutually orthogonal directions in space–time (six parameters) are given.

[4] O. Veblen and T. Y. Thomas, Trans. Am. Math. Soc. **25**, 551 (1923).

[5] J. L. Synge, *Relativity, The General Theory* (North-Holland, Amsterdam, 1966), pp. 103–09.

[6] A. Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, N.J., 1960), 3rd ed., p. 133.

[7] S. S. Schweber, *An Introduction to Relativistic Quantum Mechanics* (Harper and Row, New York, 1962), p. 40.

[8] To the best of our knowledge such a structure has never been defined before.

[9] Reference 5, pp. 83–85. Synge calls these frames simply "Fermi coordinates." This name usually refers, however, to a much bigger class of coordinate systems, namely, those that satisfy $g_{\mu\nu} = \eta_{\mu\nu}$ and $\Gamma^{\alpha}_{\beta\gamma} = 0$ along a given curve [E. Fermi, Atti Accad. Naz. Lincei Rend. Classe Sci. Fis. Mat. Nat. **31**, 21, 51 (1922); T. Levi-Civita, Math. Ann. **92**, 291 (1926); L. O'Raifeartaigh, Proc. Roy. Irish Acad. **A59**, 15 (1958); L. O'Raifeartaigh and J. L. Synge, Proc. Roy. Soc. (London) **A246**, 299 (1958)]. We prefer, therefore, to use the term "geodesic Fermi coordinates" suggested by A. Schield [*Relativity Theory and Astrophysics, I. Relativity and Cosmology, Lectures of Applied Mathematics* (American Mathematical Society, Providence, R.I., 1967), Vol. 8]. Geodesic Fermi coordinates were recently used by J. L. Anderson, "Maximal Covariance Conditions and Kretschmann's Relativity Group," in *Prospectives in Geometry and Relativity*, B. Hoffman, Ed. (Indiana U.P., Bloomington, 1966).

[10] For definition and discussion of geodesic manifolds see T. Levi-Civita, *The Absolute Differential Calculus* (Blackie, London, 1954), p. 162.

[11] Having proved part (1), we can assume Eqs. (B21) to hold. Equations (2.2) reduce then to the form

$$\Gamma^{\alpha}_{\beta\gamma\delta} = \tfrac{1}{3}P\left(\frac{\partial}{\partial x^{\delta}}\,\Gamma^{\alpha}_{\beta\gamma}\right),$$

$$\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu\nu} = \frac{1}{N}P\left(\frac{\partial}{\partial x^{\mu}}\,\Gamma^{\alpha}_{\beta\gamma\delta\cdots\mu}\right).$$

This is the form used in deriving Eqs. (B23) and (B33).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Infinite-Dimensional Representations of the Lorentz Group: Complementary Series of Representations

M. CARMELI
General Physics Research Laboratory, Aerospace Research Laboratories
Wright-Patterson AFB, Ohio 45433

AND

S. MALIN*
Department of Physics and Astronomy, Colgate University, Hamilton, New York 13346

(Received 24 July 1970)

The method used by Carmeli to obtain another form for the principal series of representations of the group $SL(2, C)$ is extended to the complementary series of representations of that group. As a result, a new form for the complementary series of representations of $SL(2, C)$ is obtained which describes the transformation law of an infinite set of numbers under the group translation in a way which is very similar, but as a generalization, to the way spinors appear in the finite-dimensional case.

1. INTRODUCTION

Recently, Carmeli[1] has introduced an infinite set of quantities which are associated with the principal series of representations of the group $SL(2, C)$ in a way which is very similar, but as a generalization, to the way spinors[2] appear in describing the finite-dimensional representations. The transformation law of these quantities,[3] at the same time, defines a new form of the principal series of representations of $SL(2, C)$.

The principal series of representations, however, do not realize all irreducible unitary representations of the group $SL(2, C)$. Rather, every irreducible unitary representation of the group $SL(2, C)$ is unitarily equivalent to a representation of the principal series or the complementary series of representations.[4]

In this paper we extend Carmeli's result to the complementary series of representations, thus establishing new forms for all irreducible unitary representations (to within unitary equivalence) of the group $SL(2, C)$.

In Sec. 2 we summarize the method used for the principal series of representations. In Sec. 3 we generalize the method to the complementary series of representations. The Appendix is devoted to detailed calculation of a normalization factor introduced in the text.

2. SUMMARY OF PREVIOUS WORK

In this section we discuss Carmeli's[1] form for the principal series of representations.

We denote by $L_2^{2s}(SU_2)$ the set of all functions $\phi(u)$,

where $u \in SU_2$, which are measurable and satisfy the conditions

$$\phi(\gamma u) = e^{is\psi}\phi(u), \tag{2.1}$$

$$\int |\phi(u)|^2 \, du < \infty, \tag{2.2}$$

where $\gamma \in SU_2$ is given by

$$\gamma = \begin{pmatrix} e^{-\frac{1}{2}i\psi} & 0 \\ 0 & e^{\frac{1}{2}i\psi} \end{pmatrix}. \tag{2.3}$$

$L_2^{2s}(SU_2)$ provides a Hilbert space[4,5] where the scalar product is defined by[6]

$$(\phi_1, \phi_2) = \int \phi_1(u)\bar{\phi}_2(u) \, du. \tag{2.4}$$

The principal series of representations is then given by the formula[4,7]

$$V_g\phi(u) = [\alpha(ug)/\alpha(u\bar{g})]\phi(u\bar{g}), \tag{2.5}$$

where

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

is an element of the group $SL(2, C)$ and $\alpha(g)$ is a function given by

$$\alpha(g) = g_{22}^{2s} \, |g_{22}|^{i\rho - 2s - 2}. \tag{2.6}$$

Here ρ is a real number and $2s$ is an integer.

Consider now all possible systems of numbers ϕ_m^j, where $m = -j, -j+1, \cdots, j$ and $j = |s|, |s|+1, |s|+2, \cdots$, with the condition

$$\sum_{j=|s|}^{\infty} (2j+1) \sum_{m=-j}^{j} |\phi_m^j|^2 < \infty. \tag{2.7}$$

The aggregate of all such systems ϕ_m^j forms a Hilbert space, which we denote by l_2^{2s}, where the scalar product is defined by

$$\sum_{j=|s|}^{\infty} (2j+1) \sum_{m=-j}^{j} \phi_m^j \overline{\psi_m^j}, \tag{2.8}$$

for any two vectors ϕ_m^j and ψ_m^k of l_2^{2s}. With each vector $\phi_m^j \in l_2^{2s}$ we associate the function

$$\phi(u) = \sum_{j=|s|}^{\infty} (2j+1) \sum_{m=-j}^{j} \phi_m^j T_m^j(u), \tag{2.9}$$

where $T_m^j(u)$ is the matrix element $T_{sm}^j(u)$ of irreducible representation of SU_2. Since[4]

$$T_m^j(\gamma u) = e^{is\psi} T_m^j(u),$$

the function given by Eq. (2.9) belongs to the space $L_2^{2s}(SU_2)$. On the other hand, every function in $L_2^{2s}(SU_2)$ can be written in the form (2.9) since $T_m^j(u)$ provide a complete orthogonal set.[5,8] The two spaces $L_2^{2s}(SU_2)$ and l_2^{2s} are, in fact, isometric where the transition from one space to the other can be made by means of the generalized Fourier transform

$$\phi_m^j = \int \phi(u)\overline{T_m^j(u)} \, du. \tag{2.10}$$

Similarly to spinors which appear as coefficients in the polynomials of the space of representation, we see that the numbers ϕ_m^j appear as coefficients in the expansion given by Eq. (2.9) of the functions $\phi(u)$ of the space $L_2^{2s}(SU_2)$. By means of the mapping (2.10), the operator V_g of the representation (2.5) may also be regarded as an operator in the space l_2^{2s}, whose explicit expression we find below. This expression also defines another form of the principal series of representations.

Applying the operator V_g to the function $\phi(u)$ as given by Eq. (2.9), we obtain

$$V_g\phi(u) = \sum_j (2j+1) \sum_m \phi_m^j \frac{\alpha(ug)}{\alpha(u\bar{g})} T_m^j(u\bar{g}) \tag{2.11}$$

or

$$V_g\phi(u) = \sum_j (2j+1) \sum_m \phi_m^j \sum_{j'} (2j'+1) \times \sum_{m'} V_{mm'}^{jj'}(g; s, \rho) T_{m'}^{j'}(u), \tag{2.12}$$

where

$$V_{mm'}^{jj'}(g; s, \rho) = \int \frac{\alpha(ug)}{\alpha(u\bar{g})} T_m^j(u\bar{g}) \overline{T_{m'}^{j'}(u)} \, du. \tag{2.13}$$

Accordingly, we obtain

$$V_g\phi(u) = \sum_j (2j+1) \sum_m \phi_m'^j T_m^j(u), \tag{2.14}$$

where, using Eq. (2.12), we have

$$\phi_{m'}'^{j'} = \sum_{j=|s|}^{\infty} (2j+1) \sum_{m=-j}^{j} V_{mm'}^{jj'}(g; s, \rho) \phi_m^j. \tag{2.15}$$

Thus the operator V_g of the principal series of representations of $SL(2, C)$ in the space l_2^{2s} is the linear transformation determined by Eq. (2.15) describing the law of transformation of the quantities ϕ_m^j, where $j = |s|, |s|+1, |s|+2, \cdots$ and $m = -j, -j+1, \cdots, j$. The matrices $V_{mm'}^{jj'}(g; s, \rho)$ are functions of $g \in SL(2, C)$ and of ρ and s, where ρ is a real number and $2s$ is an integer.

3. GENERALIZATION TO COMPLEMENTARY SERIES

A. Complementary Series of Representations of $SL(2, C)$

We denote by H the set of all bounded measurable functions $\phi(u)$ satisfying the condition

$$\phi(\gamma u) = \phi(u), \tag{3.1}$$

where again u is an element of SU_2 and γ is given by (2.3). We, furthermore, introduce in H the scalar

product

$$\langle \phi_1, \phi_2 \rangle = \pi \iint K(u'u''^{-1}) \phi_1(u') \overline{\phi_2}(u'')\, du'\, du'' \quad (3.2)$$

for ϕ_1 and ϕ_2 belonging to H. Here $K(u'u''^{-1})$ is a function given by

$$K(u) = |u_{21}|^{\sigma-2}, \quad (3.3)$$

where $0 < \sigma < 2$ and the integral on the right-hand side of (3.2) converges absolutely. The space H can be shown[4] to be a Euclidean space whose completion (which is a Hilbert space) we denote by H_σ.

The operators V_g of a representation of the complementary series in the Hilbert space H_σ are then defined by the formula[4,7]

$$V_g \phi(u) = [\alpha(ug)/\alpha(u\tilde{g})] \phi(u\tilde{g}), \quad (3.4)$$

where $\phi \in H$ and $\alpha(g)$ is given by

$$\alpha(g) = |g_{22}|^{-\sigma-2}, \quad (3.5)$$

$0 < \sigma < 2$.

B. Orthogonal Set in the Space H

We now define a set of functions which provides an orthogonal basis in the space H. It is given by

$$t^j_m(u) = N_j T^j_{0m}(u), \quad (3.6)$$

where $T^j_{0m}(u)$ are matrix elements of the irreducible representations of the group SU_2 and N_j is a real normalization factor (see Appendix) whose value is given by

$$N_j = \left(\pi \int K(u) T^j_{00}(u)\, du \right)^{-\frac{1}{2}}. \quad (3.7)$$

Here $m = -j, -j+1, \cdots, j$ and $j = 0, 1, 2, 3, \cdots$.

To show that t^j_m indeed provide an orthogonal basis in H, we calculate the scalar product

$$\begin{aligned} &\langle t^{j_1}_{m_1}, t^{j_2}_{m_2} \rangle \\ &\quad = \pi \iint K(u'u''^{-1}) t^{j_1}_{m_1}(u') \overline{t^{j_2}_{m_2}}(u'')\, du'\, du'' \\ &\quad = \pi N_{j_1} N_{j_2} \iint K(u'u''^{-1}) T^{j_1}_{0m_1}(u')\, du' \bar{T}^{j_2}_{0m_2}(u'')\, du''. \end{aligned} \quad (3.8)$$

By making the transition $u' \to u'u''$ in the above integral,[6] one obtains

$$\begin{aligned} &\langle t^{j_1}_{m_1}, t^{j_2}_{m_2} \rangle \\ &\quad = \pi N_{j_1} N_{j_2} \iint K(u') T^{j_1}_{0m_1}(u'u'')\, du' \bar{T}^{j_2}_{0m_2}(u'')\, du''. \end{aligned}$$

Using the relation

$$T^{j_1}_{0m_1}(u'u'') = \sum_{m=-j_1}^{j_1} T^{j_1}_{0m}(u') T^{j_1}_{mm_1}(u'')$$

in the last integral, we obtain

$$\begin{aligned} \langle t^{j_1}_{m_1}, t^{j_2}_{m_2} \rangle = \pi N_{j_1} N_{j_2} \sum_{m=-j_1}^{j_1} &\int K(u') T^{j_1}_{0m}(u')\, du' \\ &\times \int T^{j_1}_{mm_1}(u'') \bar{T}^{j_2}_{0m_2}(u'')\, du''. \end{aligned} \quad (3.9)$$

Using now the orthogonality relation[8] that the matrices T^j satisfy, we obtain

$$\langle t^{j_1}_{m_1}, t^{j_2}_{m_2} \rangle = \pi N_{j_1} N_{j_2} \left(\int K(u') T^{j_1}_{00}(u')\, du' \right) \frac{\delta^{j_1 j_2} \delta^{m_1 m_2}}{2j_1 + 1}, \quad (3.10)$$

which, by virtue of Eq. (3.7), gives

$$\langle t^{j_1}_{m_1}, t^{j_2}_{m_2} \rangle = \frac{\delta^{j_1 j_2} \delta^{m_1 m_2}}{2j_1 + 1}. \quad (3.11)$$

C. New Form for the Complementary Series of Representations

Consider now all possible systems of numbers ψ^j_m, where $m = -j, -j+1, \cdots, j$ and $j = 0, 1, 2, 3, \cdots$, with the condition

$$\sum_j (2j+1) N_j^{-2} \sum_{m=-j}^{j} |\psi^j_m|^2 < \infty. \quad (3.12)$$

The aggregate of all such systems of numbers forms a Euclidean space which we denote by h, where the scalar product is defined by

$$\sum_j (2j+1) N_j^{-2} \sum_{m=-j}^{j} \phi^j_m \overline{\psi^j_m} \quad (3.13)$$

for any two vectors ϕ^j_m and ψ^j_m of h. With each vector ϕ^j_m of h we associate the function

$$\phi(u) = \sum_j (2j+1) N_j^{-1} \sum_m \phi^j_m t^j_m(u), \quad (3.14)$$

where $t^j_m(u)$ is given by Eq. (3.6). Since

$$t^j_m(\gamma u) = t^j_m(u), \quad (3.15)$$

it follows that the function (3.14) belongs to the space H. On the other hand, every function in H can be written in the form (3.14) since, as we have seen, t^j_m provide a complete[9] orthogonal set in H. In fact, the two spaces H and h are isometric, where the transition from one space to the other is made by means of

$$\phi^j_m = N_j \langle \phi, t^j_m \rangle. \quad (3.16)$$

A simple calculation also shows that[10]

$$\langle \phi, \psi \rangle = \sum_j (2j+1) N_j^{-2} \sum_m \phi^j_m \overline{\psi^j_m}. \quad (3.17)$$

If we denote now by h_σ the completion[11] of the Euclidean space h, then the isometric mapping (3.16) of H on h can be extended in a unique way by continuity to an isometric mapping of H_σ on h_σ. The

operators V_g of a representation of the complementary series in the space H_σ pass over into operators in the space h_σ, which are also denoted by V_g and whose explicit expression we find below. This expression will also define a new form for the complementary series of representations.

Applying the operator V_g to the function $\phi(u)$ written in the form (3.14), we obtain

$$V_g\phi(u) = \sum_j (2j+1)N_j^{-1} \sum_m \phi_m^j \frac{\alpha(ug)}{\alpha(u\bar{g})} t_m^j(u\bar{g}) \quad (3.18)$$

or

$$V_g\phi(u) = \sum_j (2j+1) \sum_m \phi_m^j \sum_{j'} (2j'+1)N_{j'}^{-1} \times \sum_{m'} V_{mm'}^{jj'}(g;\sigma) t_{m'}^{j'}(u), \quad (3.19)$$

where

$$V_{mm'}^{jj'}(g;\sigma) = \pi \frac{N_{j'}}{N_j} \iint K(u'u''^{-1}) \frac{\alpha(u'g)}{\alpha(u'\bar{g})} \times t_m^j(u'\bar{g}) \bar{t}_{m'}^{j'}(u'')\, du'\, du''. \quad (3.20)$$

Accordingly, Eq. (3.19) has the form

$$V_g\phi(u) = \sum_j (2j+1)N_j^{-1} \sum_m \phi_m'^j t_m^j(u), \quad (3.21)$$

where

$$\phi_{m'}'^{j'} = \sum_{j=0}^{\infty} (2j+1) \sum_{m=-j}^{j} V_{mm'}^{jj'}(g;\sigma)\phi_m^j. \quad (3.22)$$

Hence we see that the operator V_g of the complementary series of representations of the group $SL(2, C)$ in the space h_σ is the linear transformation determined by Eq. (3.22), describing the law of transformation of the quantities ϕ_m^j, where $m = -j, -j+1, \cdots, j$ and $j = 0, 1, 2, 3, \cdots$. The matrices $V_{mm'}^{jj'}(g;\sigma)$ are functions of $g \in SL(2, C)$ and of σ, where $0 < \sigma < 2$.[12]

APPENDIX: THE NORMALIZATION FACTOR N_j

The normalization factor N_j was defined in Sec. 3 by

$$N_j^{-2} = \pi \int K(u) T_{00}^j(u)\, du, \quad (A1)$$

where $K(u)$ is a function of $u \in SU_2$ given by

$$K(u) = |u_{21}|^{\sigma-2} \quad (A2)$$

and $0 < \sigma < 2$.

By calculating the scalar product of $T_{00}^j(u)$ with itself, we find

$$\langle T_{00}^j, T_{00}^j \rangle = \frac{\pi}{2j+1} \int K(u) T_{00}^j(u)\, du. \quad (A3)$$

Hence we have

$$N_j^{-2} = (2j+1)\langle T_{00}^j, T_{00}^j \rangle. \quad (A4)$$

Now, the right-hand side of Eq. (A4) is positive. Hence $N_j^{-2} > 0$, and therefore N_j is real.

The evaluation of the integral in Eq. (A1) is straightforward. We have[4]

$$T_{mn}^j(u) = (-1)^{2j-m-n} \left(\frac{(j-m)!\,(j+m)!}{(j-n)!\,(j+n)!}\right)^{\frac{1}{2}} \times \sum_{r=\max(0,-m-n)}^{\min(j-m,j-n)} \binom{j-n}{r} \binom{j+n}{j-m-r} \times u_{11}^r u_{12}^{j-m-r} u_{21}^{j-n-r} u_{22}^{m+n+r}, \quad (A5)$$

where

$$\binom{m}{n} = \frac{m!}{(m-n)!\,n!}. \quad (A6)$$

Accordingly, we have

$$T_{00}^j(u) = (-1)^{2j} \sum_{r=0}^{j} \binom{j}{r}^2 (u_{11}u_{22})^r (u_{12}u_{21})^{j-r}. \quad (A7)$$

Writing now the unitary matrix u in the form

$$u = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad (A8)$$

with

$$|a|^2 + |b|^2 = 1, \quad (A9)$$

we obtain for the integral in (A1)

$$\sum_{r=0}^{j} (-1)^{3j-r} \binom{j}{r}^2 \int |a|^{2r} |b|^{2(j-r-1)+\sigma}\, du. \quad (A10)$$

We now express du in terms of some three real parameters. We take

$$t = |a|^2, \quad \theta = \arg a, \quad \lambda = \arg b. \quad (A11)$$

Then

$$du = \tfrac{1}{4}\pi^{-2}\, dt\, d\theta\, d\lambda, \quad (A12)$$

and the integral in Eq. (A10) becomes

$$\frac{1}{4\pi^2} \int_0^1 dt \int_0^{2\pi} d\theta \int_0^{2\pi} t^r (1-t)^{j-r-1+\frac{1}{2}\sigma}\, d\lambda. \quad (A13)$$

We therefore obtain

$$N_j^{-2} = \pi \sum_{r=0}^{j} (-1)^{3j-r} \binom{j}{r}^2 \int_0^1 t^r (1-t)^{j-r-1+\frac{1}{2}\sigma} dt. \quad (A14)$$

This last integral can easily be evaluated by parts. One obtains

$$\begin{aligned} &\int_0^1 t^r (1-t)^{j-r-1+\frac{1}{2}\sigma}\, dt \\ &\quad = \frac{r!}{(j+\frac{1}{2}\sigma-r)\cdots(j+\frac{1}{2}\sigma-1)} \times \int_0^1 (1-t)^{j+\frac{1}{2}\sigma-1}\, dt \\ &\quad = \frac{r!}{(j+\frac{1}{2}\sigma-r)\cdots(j+\frac{1}{2}\sigma)} \\ &\quad = r! \frac{\Gamma(j+\frac{1}{2}\sigma-r)}{\Gamma(j+\frac{1}{2}\sigma+1)}. \end{aligned} \quad (A15)$$

Using (A15) in (A14), we finally obtain

$$N_j^{-2} = \pi \sum_{r=0}^{j} (-1)^{3j-r} \binom{j}{r}^2 r! \frac{\Gamma(j + \frac{1}{2}\sigma - r)}{\Gamma(j + \frac{1}{2}\sigma + 1)}. \quad \text{(A16)}$$

* Supported in part by the Colgate Research Council and the Sloan Foundation.

[1] M. Carmeli, J. Math. Phys. **11**, 1917 (1970).

[2] The term spinor is used to mean symmetrical spinor.

[3] Just as in the spinor case, these quantities become functions of space–time when applied in physics.

[4] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

[5] M. Carmeli, J. Math. Phys. **10**, 569 (1969).

[6] The integrals in Eqs. (2.2) and (2.4) and throughout this paper are invariant integrals over the group SU_2. We recall that the invariant integral satisfies the conditions

$$\int f(uu_1)\, du = \int f(u_1 u)\, du = \int f(u)\, du,$$

for any $u_1 \in SU_2$, and

$$\int f(u^{-1})\, du = \int f(u)\, du,$$

$$\int f(u)\, du = 1.$$

The concept of invariant integral is discussed in Ref. 4. It is also discussed in detail in J. D. Talman, *Special Functions: A Group Theoretic Approach* (Benjamin, New York, 1968), based on lectures by E. P. Wigner.

[7] We use the notation of Naimark according to which $u\bar{g}$, appearing in the representation formula (2.5) and throughout this paper, denotes an arbitrary matrix from the right coset $\overline{ug}$ (for details see Ref. 4). The explicit calculation of the unitary matrix $u\bar{g}$ is straightforward. If we denote $u\bar{g}$ by u', then u' can be expressed in terms of the two matrices u and $\bar{g}$. Denoting u' by

$$u' = \begin{pmatrix} \alpha' & \beta' \\ -\bar{\beta}' & \bar{\alpha}' \end{pmatrix}, \quad |\alpha'|^2 + |\beta'|^2 = 1,$$

and u and g by

$$u = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}, \quad |\alpha|^2 + |\beta|^2 = 1,$$

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}, \qquad \det g = 1,$$

one then obtains

$$\alpha' = (-\beta\bar{g}_{12} + \alpha\bar{g}_{22})\, |\lambda|^{-1} e^{i\Lambda},$$
$$\beta' = (\beta\bar{g}_{11} - \alpha\bar{g}_{21})\, |\lambda|^{-1} e^{i\Lambda},$$
$$|\lambda|^2 = |\beta\bar{g}_{11} - \alpha\bar{g}_{21}|^2 + |-\beta\bar{g}_{12} + \alpha\bar{g}_{22}|^2.$$

Accordingly, $u\bar{g}$ is determined by means of u and $\bar{g}$ up to an arbitrary phase factor $\exp(i\Lambda)$, where Λ is real.

[8] The functions $T^j_{mn}(u)$ satisfy the orthogonality relation

$$\int T^j_{mn}(u) \overline{T}^{j'}_{m'n'}(u)\, du = (2j+1)^{-1} \delta^{jj'} \delta_{mm'} \delta_{nn'}.$$

[9] The orthogonal set of functions $t^j_m(u)$ is also complete in H. This can be seen by considering the irreducible unitary representation (3.4) of $SL(2, C)$ in the Hilbert space H_σ as an infinite-dimensional unitary representation for the subgroup SU_2 and by decomposing it into its orthogonal sum of the finite-dimensional irreducible representations. If v is an element of SU_2, then Eq. (3.4) gives for the representation of SU_2

$$V_v \phi(u) = \phi(uv),$$

since one can put $u\bar{v} = uv$ in this case (see Ref. 4). Applying the last formula to $t^j_m(u)$, we obtain

$$\begin{aligned} V_v t^j_m(u) &= t^j_m(uv) \\ &= N_j T^j_{0m}(uv) \\ &= N_j \sum_{m'=-j}^{j} T^j_{0m'}(u) T^j_{m'm}(v) \\ &= \sum_{m'=-j}^{j} T^j_{m'm}(v) t^j_{m'}(u). \end{aligned}$$

Hence the operator V_v realizes a representation of SU_2 in the space R_j of the $(2j+1)$ functions t^j_m with $-j \le m \le j$, where the matrix elements of V_v are $T^j_{m'm}(v)$. The representation $v \to V_v$ in the space of functions $t^j_m(u)$, $m = -j, -j+1, \cdots, j$, is irreducible and the t^j_m form a canonical basis in this space. Accordingly, the infinite-dimensional representation of SU_2 in the space H_σ is decomposed into irreducible parts defined in the subspaces R_j of t^j_m, where $m = -j, -j+1, \cdots, j$ and $j = 0, 1, 2, 3, \cdots$. In other words, every function of H_σ, and hence of H, can be represented in the form given by Eq. (3.14).

[10] If ψ is taken to be equal to ϕ, then Eq. (3.17) gives Plancherel's formula

$$\langle \phi, \phi \rangle = \sum_j (2j+1) N_j^{-2} \sum_{m=-j}^{j} |\phi^j_m|^2.$$

[11] Since h is a Euclidean space, then one can complete it, namely include it in a Hilbert space h_σ. See, for example, Ref. 4; for a detailed proof, see L. A. Lyusternik and V. I. Sobolov, *Elements of Functional Analysis* (Moscow, 1951), and M. A. Naimark, *Normed Rings* (Noordhoff, Groningen, The Netherlands, 1959).

[12] It is worthwhile to recall that the complementary series should be *formally* obtained from the principal series when one takes $s = 0$ and $\rho = i\sigma$ in the latter (see Naimark in Refs. 4 and 11). This fact can also be seen for the new forms of representations of the series. By making the transition $u'' \to u''u'$ in Eq. (3.20), one obtains for our matrices for the complementary series

$$V^{jj'}_{mm'}(g; \sigma) = \pi \frac{N_{j'}}{N_j} \iint K(u''^{-1}) \frac{\alpha(u'g)}{\alpha(u'\bar{g})} t^j_m(u'\bar{g}) \bar{t}^{j'}_{m'}(u''u')\, du'\, du''.$$

Using Eqs. (3.6) and

$$T^{j'}_{0m'}(u''u') = \sum_{n=-j'}^{j'} T^{j'}_{0n}(u'') T^{j'}_{nm'}(u'),$$

we obtain

$$\begin{aligned} &V^{jj'}_{mm'}(g; \sigma) \\ &\quad = \pi N_{j'}^2 \sum_n \int K(u''^{-1}) \overline{T}^j_{0n}(u'')\, du'' \int \frac{\alpha(u'g)}{\alpha(u'\bar{g})} T^j_{0m}(u'\bar{g}) \overline{T}^{j'}_{nm'}(u')\, du'. \end{aligned}$$

By Eq. (3.4) the expression $[\alpha(u'g)/\alpha(u'\bar{g})] T^j_{0m}(u'\bar{g})$ is an element of the space H. Hence the second integral vanishes unless $n = 0$, in which case we obtain, by Eqs. (2.13) and (2.6),

$$V^{jj'}_{mm'}(g; \sigma) = \pi N_{j'}^2 \int K(u''^{-1}) \overline{T}^{j'}_{00}(u'')\, du'' [V^{jj'}_{mm'}(g; s, \rho)]_{s=0, \rho=i\sigma},$$

where $V^{jj'}_{mm'}(g; s, \rho)$ are Carmeli's matrices for the principal series. Using the fact that $\overline{T}^j_{mn}(u) = T^j_{nm}(u^{-1})$ and the invariance property of our integrals, we obtain

$$\begin{aligned} V^{jj'}_{mm'}(g; \sigma) &= \pi N_{j'}^2 \int K(u'') T^{j'}_{00}(u'')\, du'' [V^{jj'}_{mm'}(g; s, \rho)]_{s=0, \rho=i\sigma} \\ &= [V^{jj'}_{mm'}(g; s, \rho)]_{s=0, \rho=i\sigma}, \end{aligned}$$

by Eq. (3.7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Lattice Green's Function for the Body-Centered Cubic Lattice

SHIGETOSHI KATSURA* AND TSUYOSHI HORIGUCHI†
Department of Applied Physics and Department of Physics,† Tohoku University, Sendai, Japan*

(Received 23 April 1970)

The lattice Green's function for the body-centered cubic (bcc) lattice

$$I(t) = \frac{1}{\pi^3}\int_0^\pi\int\int \frac{dx\,dy\,dz}{t \pm i\epsilon - \cos x \cos y \cos z}$$

is considered. With the use of the analytic continuation to complex value of t from Maradudin's result for $t > 1$, the value of the real and imaginary parts of the integral $I(t \pm i\epsilon)$ for $0 < t < 1$, $\epsilon \to 0$, is obtained. The expressions valid for $t \to \infty$, $t \gtrsim 1$, $t \lesssim 1$, and $t \sim 0$ are given. They are useful for analyzing the nature of the singularity and for carrying out numerical calculations in all regions of t.

The lattice Green's function[1,2] for the body-centered cubic (bcc) lattice

$$I(t) = \frac{1}{\pi^3}\int_0^\pi\int\int \frac{dx\,dy\,dz}{t \pm i\epsilon - \cos x \cos y \cos z} \tag{1}$$

is considered. It is real for $t > 1$ and complex for $0 < t < 1$, $\epsilon \to 0$. Tables of the integral (1) are given in the literature.[3-8] Most of the methods of calculation are tedious or of slow convergence (except for $t \to \infty$), and the analytic properties of the integral are not yet sufficiently clear. In the present paper, it is shown that the principle of the analytic continuation to the complex value of t from Maradudin's result[9] for $t > 1$ gives the value of the real and the imaginary parts of $I(t \pm i\epsilon)$ for $0 < t < 1$, $\epsilon \to 0$. The integral (1) is transformed into several forms in terms of hypergeometric functions. These expressions are transparent for analyzing the nature of the singularities at $t = 0$ and $t = 1$, and are very simple for carrying out the numerical calculations in all regions of t, i.e., $t \gtrsim 0$, $t \lesssim 1$, $t \gtrsim 1$, and $t \to \infty$.

For $t > 1$, the integral (1) is expressed as a power series in $1/t^2$:

$$I(t) = \frac{1}{t}\sum_{n=0}^{\infty}\left[\frac{1}{\pi^{\frac{1}{2}}}\frac{\Gamma(n+\frac{1}{2})}{n!}\right]^3\left(\frac{1}{t^2}\right)^n = \frac{1}{t}\,{}_3F_2\left(\frac{1}{2},\frac{1}{2},\frac{1}{2};1,1;\frac{1}{t^2}\right). \tag{2}$$

Equation (2) is expressed[9,10] in terms of the complete elliptic integral of the first kind $K(k)$ in the case $t > 1$:

$$I(t) = t^{-1}4\pi^{-2}K^2(k), \tag{3}$$

$$k^2 = \tfrac{1}{2}[1 - (1 - t^{-2})^{\frac{1}{2}}]. \tag{3'}$$

In this paper the case $t < 1$ is considered. The proof of (3) shows that (3) is valid for $|t| > 1$ for complex value of t. Hence, by the principle of analytic continuation, (3) is also valid when t approaches the real axis for $0 < t < 1$ by indenting $t = 1$ from upper or lower part. For practical use, the following transformation is useful.

Expressing the complete elliptic integral in terms of Gauss' hypergeometric function ${}_2F_1$ and using the quadratic transformations[11] for ${}_2F_1$, we have,

$$I(t) = t^{-1}[{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;\tfrac{1}{2} - \tfrac{1}{2}(1 - t^{-2})^{\frac{1}{2}})]^2, \tag{4}$$

$${}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;\tfrac{1}{2} - \tfrac{1}{2}(1 - t^{-2})^{\frac{1}{2}}) = \frac{\pi^{\frac{1}{2}}}{[\Gamma(\frac{3}{4})]^2}\,{}_2F_1(\tfrac{1}{4},\tfrac{1}{4};\tfrac{1}{2};1 - t^{-2}) - \frac{2\pi^{\frac{1}{2}}}{[\Gamma(\frac{1}{4})]^2}(1 - t^{-2})^{\frac{1}{2}}\,{}_2F_1(\tfrac{3}{4},\tfrac{3}{4};\tfrac{3}{2};1 - t^{-2}). \tag{5}$$

Both terms are real for $t > 1$. The first term is real, and the second term is imaginary for $t < 1$. The expression (5) is useful for $t \gtrsim 1$. The expressions useful for $t \lesssim 1$ and $t \gtrsim 0$ will be obtained by transforming (5):

For $0 < t < 1$, $\epsilon \to 0$,

$$[1 - (t \pm i\epsilon)^{-2}]^{\frac{1}{2}} = \pm i(t^{-2} - 1)^{\frac{1}{2}}.$$

Hereafter we consider the case of lower sign.

Transforming ${}_2F_1(\quad;\quad;z)$ into ${}_2F_1(\quad;\quad;z/(1-z))$, we have

lhs of (5)

$$= \frac{\pi^{\frac{1}{2}}}{[\Gamma(\frac{3}{4})]^2}t^{\frac{1}{2}}\,{}_2F_1(\tfrac{1}{4},\tfrac{1}{4};\tfrac{1}{2};1 - t^2) + i\frac{2\pi^{\frac{1}{2}}}{[\Gamma(\frac{1}{4})]^2}t^{\frac{1}{2}}(1 - t^2)^{\frac{1}{2}}\,{}_2F_1(\tfrac{3}{4},\tfrac{3}{4};\tfrac{3}{2};1 - t^2). \tag{6}$$

The expression (6) is useful for $t \lesssim 1$.

Taking the fact $a + b = c$ into consideration, we transform ${}_2F_1(a, b; c; z)$ into

$$\text{Eq. (6)} = \frac{t^{\frac{1}{2}}}{2\pi}\Big[\sum_{n=0}^{\infty}\left(\frac{(\frac{1}{4})_n}{n!}\right)^2 \times t^{2n}[2\psi(n+1) - 2\psi(n+\tfrac{1}{4}) - \log t^2] + i(1-t^2)^{\frac{1}{2}}\sum_{n=0}^{\infty}\left(\frac{(\frac{3}{4})_n}{n!}\right)^2 \times t^{2n}[2\psi(n+1) - 2\psi(n+\tfrac{3}{4}) - \log t^2]\Big]. \quad (7)$$

The expression (7) is useful for $t \geqslant 0$. Substituting (7) into (4), we obtain the leading term for $t \to 0$ from the term of $n = 0$. Inserting

$$\psi(\tfrac{1}{4}) = -\tfrac{1}{2}\pi - \gamma - 3\log 2, \quad (8)$$

$$\psi(\tfrac{3}{4}) = \tfrac{1}{2}\pi - \gamma - 3\log 2 \quad (9)$$

into (7), we have

$$\text{Re } I(t) = (2/\pi)(-\log t + 3\log 2) + O(t^2), \quad (10)$$

$$\text{Im } I(t) = (2/\pi^2)[(-\log t + 3\log 2)^2 - (\tfrac{1}{4}\pi)^2] + O(t^2), \quad (11)$$

where the terms $O(t^2)$ contain those of $t^2(\log t)^2$ and $t^2(\log t)$.

Equation (10) shows that the state density for bcc diverges at the center of the band.[12]

By using (5) and (6), the leading term at $t \sim 1$ is shown to be

$$\text{Re } I(t) \sim t^{-1}\pi[\Gamma(\tfrac{3}{4})]^{-4} + (2/\pi)(1-t^{-2})^{\frac{1}{2}}, \quad (12)$$

$$\text{Re } I(t) = O(1), \quad (13)$$

$$\text{Im } I(t) \sim (2/\pi)(1-t^2)^{\frac{1}{2}}. \quad (14)$$

The combined subroutine including Eqs. (2), (4), (5), (6), and (7) is very convenient and rapid for the calculation of the values of the real and the imaginary parts of the integral (1) for all values of t.

The lattice Green's function to the antiferromagnet for bcc lattice was calculated by Walker *et al.*[6] It is to be noted that the Green's function for the antiferromagnet for bcc is directly related to that for ferromagnet, since

$$\sum \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\epsilon^2 - \epsilon_{\mathbf{k}}^2} = \frac{1}{2\epsilon}\sum \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\epsilon - \epsilon_{\mathbf{k}}}(1 + e^{i\pi(l+m+n)}), \quad (15)$$

where $\mathbf{r} = (l, m, n)$. For all even l, m, and n, the lhs is equal to

$$\frac{1}{\epsilon}\sum \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\epsilon - \epsilon_{\mathbf{k}}},$$

and, for odd $l + m + n$, both hand sides of (15) are equal to zero.

Recently Iwata[13] expressed I_{fcc} $(t > 3)$ and I_{fcc} $(t < -1)$, the lattice Green's function for the face-centered cubic (fcc) lattice, in terms of elliptic integrals. The method of the present paper can be applied to his results and Re I_{fcc} $(-1 < t < 3)$ and Im I_{fcc} $(-1 < t < 3)$ can be obtained. In particular, the singularity at $t \sim -1$ (at the edge of the band) is shown to be of the similar nature as that of bcc at $t \sim 0$ (logarithmic divergence).

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Dr. S. Inawashiro, Dr. T. Morita, and Y. Abe.

[1] Lattice Green's functions are discussed, for example, by E. W. Montroll, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (U. of California Press, Berkeley, Calif., 1955), Vol. III, p. 209, and by A. A. Maradudin, E. W. Montroll, G. H. Weiss, R. Herman, and H. W. Milnes, Acad. Roy. Belg., Classe Sci., Mem: Collection in 4°, **14**, (1960).

[2] S. Katsura, T. Morita, S. Inawashiro, T. Horiguchi, and Y. Abe, J. Math. Phys., to be published.

[3] M. Yussouff and J. Mahanty, Proc. Phys. Soc. **87**, 689 (1966).

[4] P. Vashishta and M. Yussouff, Department of Physics, Indian Institute of Technology, Kanpur, India, Technical Report No. 12/66, 1966.

[5] T. Tonegawa, Ph.D. thesis, Osaka University, 1968.

[6] L. R. Walker, B. B. Celtin, and D. Hone, J. Phys. Chem. Solids **30**, 923 (1969).

[7] T. Morita and T. Horiguchi, J. Math. Phys., to be published.

[8] I. Mannari and C. Kawabata, Department of Physics, Okayama University, Research Notes, No. 15, 1964: the real part for $t > 1$.

[9] Equation (3) is reported to be derived by Maradudin, cited in Ref. 1.

[10] Equation (1) on p. 185 and Eq. (10) on p. 111 in Ref. 11 are used.

[11] Bateman Project, *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill, New York, 1953), Vol. 1.

[12] T. Horiguchi and T. Morita, Phys. Letters **32A**, 191 (1970).

[13] G. Iwata, Nat. Sci. Rept., Ochanomizu Univ., Tokyo, **20**, 13 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

"Short-Ranged" Scalar Gravity in Einstein's Theory

A. DAS
Department of Mathematics, Simon Fraser University, Burnaby 2, British Columbia, Canada

(Received 31 July 1970)

This paper deals with the class of conformally flat universes satisfying Einstein's field equations from a new point of view. It is shown that this class involves a scalar or pseudoscalar field that necessarily satisfies a meson-type field equation. This scalar gravity does not act at a distance, so to say, but supplies cohesive forces inside the source structure. Furthermore, the scalar field may involve spin-zero, massive quanta if quantized. Three exact solutions of this class are furnished here.

1. INTRODUCTION

"The general theory of relativity is bedevilled by the large number of unknown functions—the ten components of g_{ij}. There is little hope of getting physically interesting results without making drastic reduction in their number."[1] The usual way of reducing this number is to study some special universes admitting certain groups of motions. To minimize the number of unknown functions, one has to revert to the conformally flat universes, either subjected to Einstein's equations or otherwise constrained. The non-Einstein scalar theories[2] lack aestheticism as well as experimental validity. On the other hand, Einstein's equations of vacuum imposed on a conformally flat universe turn it just flat[3] or destroy any gravity content. This very reason makes the scalar gravity act only inside the source structures, where it need not be trivial. Moreover, this scalar field satisfies a mesonlike equation which follows from the contraction of Einstein's equations. The pseudoscalar field may be produced from the scalar by multiplying the factor $(-g)^{\frac{1}{2}}$. The quantization of the scalar gravity should be easier because the field equations can be expressed essentially in the Minkowskian metric.

But the sailing is not smooth all the way. The price of the drastic reduction of unknown functions is paid by encountering the highly overdetermined systems of equations. Moreover, in the gravity interpretation of mesons, a crude estimate shows that meson mass turns out to be about 10^4 times the acceptable value.

In this paper we propose a modified field equation in terms of a Riemann tensor, which boils down to Einstein's equation in a conformally flat universe. Three examples of exact solutions of the modified equations are supplied here. The first example is essentially the de Sitter universe. The second one is an exact wave solution of the modified field equations where energy–momentum–stress tensor is borrowed from another scalar field. In the third example a static solution of the modified equations coupled with electromagnetism is derived.

The author believes that gravity is structurally important in the very large or in the very small dimensions. He has especially been working on the application of gravity in the elementary particles.[4] This present attempt is towards isolating that component in the ingredients of Einstein's theory which may be the most relevant in the formation of elementary structures.

2. THE MODIFIED FIELD EQUATIONS

First, the notations and definitions will be clarified. V_4 denotes the 4-dimensional Riemannian manifold of events. An event $x \in V_4$ has the real coordinates x^i (Latin indices take the values 1, 2, 3, 4). The summation convention is followed. V_4 has the index of inertia -2, i.e., the metric form $\Phi = g_{ij}\,dx^i\,dx^j$ is reducible at a regular point to

$$\Phi = \eta_{ij}\,dx^i\,dx^j = -(dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (dx^4)^2.$$

Units are chosen so that $c = G = 1$.

Now the modified field equations are stated, and the immediate consequences thereof are derived in the following theorem.

Theorem 1: Let the Riemann tensor satisfy in a regular domain D of V_4 the following field equation:

$$R_{hijk} = -4\pi(g_{h[k}T_{j]i} + g_{i[j}T_{k]h} + \tfrac{2}{3}Tg_{h[j}g_{k]i}), \quad (2.1)$$

where T_{ij} is a differentiable tensor field with contraction T and square brackets denote antisymmetrization.

Then the following implications are true:

(i) Einstein's equations

$$R_{ij} - \tfrac{1}{2}g_{ij}R = 8\pi T_{ji} \quad (2.2)$$

hold in D.

(ii) The domain D is conformally flat, i.e., $g_{ij} = \phi^2\eta_{ij}$.

(iii) The D is flat if, furthermore, $T_{ij} = 0$.

$$\text{(iv)} \quad \text{(a)}\ T_{ij} = T_{ji}, \quad \text{(b)}\ T^{ij}_{\ \ j} = 0, \quad \text{(c)}\ T_{i[j/k]} - \tfrac{1}{3}g_{i[j}T_{,k]} = 0, \quad (2.3)$$

where the stroke and comma denote covariant and partial derivatives.

(v) $$\Box\phi + \tfrac{4}{3}\pi\tilde{T}_0\phi = 0, \tag{2.4}$$

where $\Box$ is the usual d'Alembertian and

$$\tilde{T}_0 \equiv \eta^{ij}\tilde{T}_{ij}, \quad \tilde{T}_{ij} \equiv T_{ij} - \tfrac{1}{2}g_{ij}T.$$

Proof: (i) First and second contractions of (2.1) give

$$R_{ij} = -8\pi\tilde{T}_{ji}, \tag{2.5}$$

and

$$R = 8\pi T. \tag{2.6}$$

From the above, (2.2) follows immediately.

(ii) The conformal curvature tensor

$$C_{hijk} \equiv R_{hijk} + (\tfrac{1}{2}g_{h[j}R_{k]i} + g_{i[k}R_{j]h}) + \tfrac{1}{6}Rg_{h[k}g_{j]i} = 0, \tag{2.7}$$

in D by (2.1), (2.5), and (2.6). Therefore, the domain is conformally flat.[5]

(iii) This consequence is obvious from (2.1). The physical implication is that this type of gravity dies outside the sources.

(iv) By (2.2) the equations (a) and (b) are obvious. Now, from (2.7), one has

$$C^h_{ijk/h} = 0.$$

The above equation implies (2.3) by (2.1) and (2.7). It should be noted that by (iv) all the identities of Riemann tensor are consistent with (2.1).

(v) From (ii) one is allowed to put

$$g_{ij} = \phi^2\eta_{ij}. \tag{2.8}$$

Plugging (2.8) into (2.6), one obtains (2.4). It should be noted that (2.4) is a meson-type equation where meson mass is related to the source mass.

Theorem 2: The field equations

$$\phi_{,ij} + \tfrac{1}{2}\eta_{ij}\Box\phi - 2\phi^{-1}(\phi_{,i}\phi_{,j} - \tfrac{1}{4}\eta_{ij}\eta^{k\ell}\phi_{,k}\phi_{,\ell}) = -4\pi\tilde{T}_{ij}\phi, \tag{2.9}$$

together with (2.8) hold iff (2.1) does.

Proof: Assume (2.9) and (2.8). By (2.8)

$$C_{hijk} = 0. \tag{2.10}$$

Now by (2.8) and (2.9), Eq. (2.5) is a consequence and, when (2.5) is substituted in (2.10), the field equation (2.1) immediately follows. The converse is obvious.

Theorem 3: The most general coordinate transformations under which the field equations (2.9) together with (2.8) are covariant are arbitrary compositions of the following:

(i) the inhomogeneous Lorentz group

$$x'^i = a^i + l^i_{,j}x^j,$$

(ii) the scale transformations

$$x'^i = \lambda x^i,$$

(iii) the inversions with respect to a hypersphere

$$x'^i = x^i/(\eta_{k\ell}x^kx^\ell).$$

The proof follows from Bianchi's theorem.[6]

Comment: The field equations (2.9) together with the constraints (2.3) are an overdetermined system of equations. Because the number of unknown functions are

$$1(\phi) + 10(T_{ij}) = 11,$$

whereas the number of equations are

$$10(2.9) + 20(2.3) = 30.$$

Note that no coordinate conditions are allowable. In spite of this overdeterminism, some exact solutions are derived in the next section.

3. EXACT SOLUTIONS OF THE MODIFIED FIELD EQUATIONS

In the first example the following choice is made:

$$T_{ij} = (-3K/8\pi)g_{ij}, \tag{3.1}$$

where K is a scalar, not necessarily constant.

Substituting (3.1) into (2.1), one has

$$R_{hijk} = Kg_{h[j}g_{k]i}. \tag{3.2}$$

From the above it is evident that K must be a constant and V_4 is a space of constant curvature. The Riemannian metric form of V_4 is

$$\Phi = \phi^2\eta_{ij}\,dx^i\,dx^j = (1 + \tfrac{1}{4}K\eta_{k1}x^kx^1)^{-2}\eta_{ij}\,dx^i\,dx^j. \tag{3.3}$$

The above form together with (3.1) satisfy all the Eqs. (2.1)–(2.9) and can be recognized as the classical de Sitter universe.

For the second case, a real massless scalar field X is chosen and coupled to the modified gravity equations (2.9). So one has the following system of equations:

$$g^{ij}X_{/ij} = \phi^{-2}(\Box X + 2\phi^{-1}\eta^{k\ell}X_{,k}\phi_{,\ell}) = 0, \tag{3.4}$$

$$\phi_{,ij} + \tfrac{1}{2}\eta_{ij}\Box\phi - 2\phi^{-1}(\phi_{,i}\phi_{,j} - \tfrac{1}{4}\eta_{ij}\eta^{k\ell}\phi_{,k}\phi_{,\ell}) = -4\pi X_{,i}X_{,j}\phi. \tag{3.5}$$

A special class of solutions is investigated by imposing a relationship

$$X = K(\phi). \tag{3.6}$$

With the above, (3.4) becomes

$$\Box\phi + [(\ln K')' + 2\phi^{-1}]\eta^{k\ell}\phi_{,k}\phi_{,\ell} = 0, \tag{3.7}$$

where the prime denotes differentiation with respect to ϕ.

But the contraction of (3.5) yields

$$\Box\phi + \tfrac{4}{3}\pi(K')^2\eta^{k\ell}\phi_{,k}\phi_{,\ell}\phi = 0. \tag{3.8}$$

Comparing (3.7) and (3.8), one has

$$(\ln K')' + 2\phi^{-1} - \tfrac{4}{3}\pi(K')^2\phi = 0, \tag{3.9}$$

where $\eta^{k\ell}\phi_{,k}\phi_{,\ell}$ is not necessarily zero.

The general solution of (3.9) is

$$X = K(\phi) = \pm(\tfrac{3}{4}\pi)^{\frac{1}{2}} \ln|[1 \pm (1 \pm a^2\phi^2)^{\frac{1}{2}}]\phi^{-1}| + b, \tag{3.10}$$

a and b being constants of integration.

Substituting (3.10) into (3.5), one gets

$$\phi_{,ij} + (1 \pm 2a^2\phi^2)(1 \pm a^2\phi^2)^{-1}\phi^{-1}\phi_{,i}\phi_{,j} + \tfrac{1}{2}\eta_{ij}(\Box\phi + \phi^{-1}\eta^{k\ell}\phi_{,k}\phi_{,\ell}) = 0. \tag{3.11}$$

Introducing another function

$$[F(\phi)]^2 = a^{-4}(1 \pm a^2\phi^2), \tag{3.12}$$

where $a \neq 0$, we reduce the equation

$$F_{,ij} + \tfrac{1}{2}\eta_{ij}(\Box F - 3F^{-1}\eta^{k\ell}F_{,k}F_{,\ell}) = 0. \tag{3.13}$$

Solving above for $i \neq j$, one gets

$$F = f_1(x^1) + f_2(x^2) + f_3(x^3) + f_4(x^4), \tag{3.14}$$

where the $f_i(x^i)$ are arbitrary functions of integration.

Substituting (3.14) into (3.13) and solving for $i = j$ case, one finally obtains

$$F = \alpha\eta_{ij}(x^i + \xi^i)(x^j + \xi^j), \tag{3.15}$$

where α and ξ^i are constants of integration.

In case the constant a in (3.10) and (3.11) is zero, one obtains the following solution:

$$X = \pm(\tfrac{3}{4}\pi)^{\frac{1}{2}} \ln|\phi|, \quad F = \tfrac{1}{2}\phi^2 = \alpha_i x^i, \tag{3.16}$$

where the α_i are constants of integration.

Therefore, Eqs. (3.10), (3.12), and (3.15) or else Eqs. (3.16) furnish two exact wave solutions of the system (3.4) and (3.5).

In the third example the combined scalar gravity and electromagnetic field equations are considered in a static setting. Now V_3 denotes a time-constant hypersurface of V_4. A space point $\mathbf{x} \in V_3$ has coordinates x^α (Greek indices take the values 1, 2, 3). Static conditions on electro-gravity are the following:

$$\phi = \phi(\mathbf{x}), \quad F_{\alpha\beta} = 0, \quad F_{4\alpha} = -F_{\alpha 4} = [A(\mathbf{x})]_{,\alpha}. \tag{3.17}$$

The modified field equations (2.9) and electrostatic equations are given by

$$\phi_{,\alpha\beta} - 2\phi^{-1}(\phi_{,\alpha}\phi_{,\beta} - \tfrac{1}{4}\delta_{\alpha\beta}\phi_{,\gamma}\phi_{,\gamma}) = 4\pi(A_{,\alpha}A_{,\beta} - \tfrac{1}{2}\delta_{\alpha\beta}A_{,\gamma}A_{,\gamma})\phi^{-1}, \tag{3.18}$$

$$\phi_{,\alpha}\phi_{,\alpha} = 4\pi A_{,\alpha}A_{,\alpha}, \tag{3.19}$$

$$A_{,\alpha\alpha} = 0. \tag{3.20}$$

Note that both $A(\mathbf{x})$ and $\phi(\mathbf{x})$ are harmonic in E_3. Now a functional relationship will be imposed as

$$\phi = \Phi[A(\mathbf{x})]. \tag{3.21}$$

From (3.19), the linear relationship[7]

$$\phi = a + (4\pi)^{\frac{1}{2}}A, \tag{3.22}$$

where a is a constant, follows immediately.

Plugging (3.22) into (3.18), one obtains

$$\phi_{,\alpha\beta} - 3\phi^{-1}\phi_{,\alpha}\phi_{,\beta} + \delta_{\alpha\beta}\phi^{-1}\phi_{,\gamma}\phi_{,\gamma} = 0. \tag{3.23}$$

First solving for $\alpha \neq \beta$, then for $\alpha = \beta$, one gets the general solution

$$\phi = -m/[(x^1 - \xi^1)^2 + (x^2 - \xi^2)^2 + (x^3 - \xi^3)^2]^{\frac{1}{2}}, \tag{3.24}$$

where m and ξ^α are constants of integration. Equations (3.24) and (3.22) furnish a static solution of the system (3.18), (3.19), and (3.20).

ACKNOWLEDGMENT

The author acknowledges thankfully the financial assistance of the National Research Council of Canada, Grant No. A-3993.

[1] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 317.
[2] D. E. Littlewood, Proc. Cambridge Phil. Soc. **49**, 90 (1953); F. A. E. Pirani, *ibid.* **51**, 535 (1955); P. Rastall, Can. J. Phys. **46**, 2155 (1968).
[3] H. W. Brinkman, Proc. Natl. Acad. **9**, 172 (1923).
[4] A. Das, Proc. Roy. Soc. (London) **A267**, 1 (1962); J. Math. Phys. **4**, 45 (1963); A. Das and C. V. Coffman, *ibid.* **8**, 1720 (1967); N. De, Progr. Theoret. Phys. (Kyoto) **33**, 545 (1965); Nuovo Cimento **39**, 986 (1965).
[5] L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1949), p. 92.
[6] See Ref. 5, p. 94.
[7] H. E. J. Curzon, Proc. Math. Soc. (London) **23**, 477 (1925); S. D. Majumdar, Phys. Rev. **72**, 390 (1947); A. Papapetrou, Proc. Roy. Irish Acad. **A51**, 191 (1947).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Zeros of the Partition Function for the Heisenberg, Ferroelectric, and General Ising Models

MASUO SUZUKI* AND MICHAEL E. FISHER
Baker Laboratory, Cornell University, Ithaca, New York, 14850

(Received 7 July 1970)

The Lee–Yang theorem for the zeros of the partition function of a ferromagnetic Ising model with real pair spin interactions is extended to general Ising models with complex many-spin interactions (satisfying appropriate "ferromagnetic" and spin inversion symmetry conditions). When many-spin interactions are present, all zeros lie on the imaginary H^z-axis for sufficiently low (but fixed) T, but, in general, some leave the imaginary axis as $T \to \infty$. The extended Ising theorem is used to prove the same result for a Heisenberg system of arbitrary spin with the real anisotropic pair interaction Hamiltonian

$$\mathcal{H}_{ij} = -(J_{ij}^x S_i^x S_j^x + J_{ij}^x S_i^x S_j^x + J_{ij}^y S_i^y S_j^y)$$

in an arbitrary transverse field (H^x, H^y) under the "ferromagnetic" condition $J_{ij}^z \geq |J_{ij}^x|$ and $J_{ij}^z \geq |J_{ij}^y|$. The analyticity of the limiting free energy of such a Heisenberg ferromagnet and the absence of a phase transition are thereby established for all (real) nonzero magnetic fields H^z. The Ising theorem is also applied to hydrogen-bonded ferroelectric models to prove, in particular, that the zeros for the KDP model lie on the imaginary electric field axis for all T below the transition temperature T_c.

1. INTRODUCTION AND SUMMARY

In connection with their theory of phase transitions, Lee and Yang[1] discussed the zeros of the partition function $Z_N(z)$ of an Ising model of N spins as a function of the "activity" variable

$$z = \exp(-H^z/k_B T), \tag{1.1}$$

where H^z is the external magnetic field (in energy units). Under the condition that the Ising spins ($\sigma_i = \pm 1$) interact only through pairwise terms, $-J_{ij}\sigma_i\sigma_j$, which are ferromagnetic ($J_{ij} \geq 0$), they proved that all zeros lie on the circle $|z| = 1$ (or, equivalently, on the imaginary H^z-axis, Re $\{H^z\} = 0$). This result has been enormously useful in studying the existence and location of phase transitions in the Ising model and in discussing the analyticity of the limiting thermodynamic free energy and its derivatives. The "circle" theorem has been extended by Asano,[2] Suzuki,[3] and, most completely, by Griffiths[4] to ferromagnetic Ising models of arbitrarily high spin. Existing proofs, however, do not cover many-body spin interactions such as $-J_{ijkl}\sigma_i\sigma_j\sigma_k\sigma_l$.

Recently Heilmann and Lieb[5] have proved a corresponding theorem for the monomer-dimer problem,[6] which has, equally, led to many valuable conclusions. As a by-product of their proof, Heilmann and Lieb were able to prove the circle theorem for a ferromagnetically coupled Heisenberg spin system at high enough temperatures. Earlier Suzuki[7] had given a complementary proof for the Heisenberg model valid for low enough temperatures. In both of these cases the range of temperature (or T^{-1}) for which the proofs are valid shrinks to zero in the thermodynamic limit. However, numerical studies by Kawabata and Suzuki[8] for several small Heisenberg systems indicated that the circle theorem was probably valid for all T. Indeed, very recently Asano[9] has considered the general Heisenberg Hamiltonian

$$\mathcal{H}_N = -\sum_{(ij)} (J_{ij}^x S_i^x S_j^x + J_{ij}^y S_i^y S_j^y + J_{ij}^z S_i^z S_j^z) - 2\sum_{i=1}^{N} (H_i^x S_i^x + H_i^y S_i^y) - 2H^z \sum_{j=1}^{N} \mu_j S_j^z, \tag{1.2}$$

in the uniform cylindrically symmetric case, $\mu_j \equiv 1$, $J_{ij}^x \equiv J_{ij}^y$, and $H_i^x \equiv H_i^y \equiv 0$. In this case $M_N^z = m\sum_j S_j^z$ commutes with $\mathcal{H}_N$, and so the partition function $Z_N(z)$ is, as for an Ising model, still a polynomial in z and z^{-1}. Asano then proves that all the zeros of $Z_N(z)$ lie on the circle $|z| = 1$ provided that

$$J_{ij}^z \geq |J_{ij}^x| = |J_{ij}^y| \geq 0. \tag{1.3}$$

The purpose of the present note is, first, to extend the circle theorem to Ising models with many-spin interactions; for a given set of many-spin interactions of finite order, our results (see Theorem 1 below) establish the circle theorem for all temperatures below a certain fixed temperature (*independent* of the total number of spins N); we also present counterexamples to show that the theorem should *not* apply at sufficiently high temperatures. Secondly, we extend the circle theorem to the nonuniform, fully anisotropic general-spin Heisenberg model ($\mu_i \neq$ const, $J_{ij}^x \neq J_{ij}^y$, $H_i^x \neq H_i^y \neq 0$) under the "ferromagnetic" conditions

$$\mu_i \geq 0, \quad J_{ij}^z \geq |J_{ij}^x|, \quad J_{ij}^z \geq |J_{ij}^y| \tag{1.4}$$

(see Theorem 2 below). We remark that in the fully anisotropic case the partition function is not merely a

polynomial in z and z^{-1}. (Our methods can also provide some rather limited results for Heisenberg models with many-spin interactions.) Lastly, we discuss the zeros of the partition function of a general class of hydrogen-bonded ferroelectric models in the variable $z = \exp(-\mathcal{E}/k_B T)$, where $\mathcal{E}$ is the imposed electric field. In particular, for the KDP model and its extension allowing "doubly ionized" vertices, we prove that the zeros lie on the circle $|z| = 1$ for $T \leq T_c$, where T_c is the transition temperature of the model. (We also make an observation on the class of exactly soluble extended ferroelectric models.)

The techniques of analysis we use owe some important elements to Asano's recent work, but we believe our methods are simpler and more transparent (as well as leading to stronger results). The over-all strategy is as follows. First (from Asano) we introduce a class of multivariable functions $F(z_1, \cdots, z_n)$ which satisfy a certain "Lee–Yang condition" regarding the location of their zeros.[10] Ising spin functions $f(\sigma_1, \cdots, \sigma_n)$ belong to the Lee–Yang class if their "transforms" (i.e., corresponding many-field partition functions) do. A sequence of lemmas establishes that products of spin functions of Lee–Yang type still belong to the class, so that, ultimately, the total Ising partition function is of Lee–Yang type provided that the partial spin interaction Hamiltonians, or corresponding individual Boltzmann factors, satisfy an appropriate condition. The condition (see Theorem 1) essentially states that the Boltzmann factor for the fully aligned states dominates the sum of factors for the other "disordered" states.

To handle the Heisenberg model, the first step is to reduce its partition function, for the case $S = \frac{1}{2}$, to that of a suitable generalized Ising model. This is accomplished by using matrix identities of the type

$$\exp(\mathsf{A} + \mathsf{B}) = \lim_{n\to\infty} [\mathsf{I} + n^{-1}(\mathsf{A} + \mathsf{B}) + n^{-2}\mathsf{C}]^n, = \lim_{n\to\infty} [(\mathsf{I} + n^{-1}\mathsf{A})(\mathsf{I} + n^{-1}\mathsf{B})]^n. \quad (1.5)$$

The utility of such expressions was observed a little while ago by Ginibre,[11] who, however, made principal use of Trotter's formula,[12] which retains A and B as exponents. Asano[9] also used Trotter's formula, but the linear forms (1.5) offer some distinct advantages.

On inserting intermediate states of definite S_1^z, $S_2^z, \cdots, S_n^z$ into the finite n products in (1.5), one obtains a partition function for an Ising lattice of more elaborate structure which contains explicit 4-spin and 2-spin interactions (arising, respectively, from the Heisenberg coupling and the transverse fields). For large enough n the equivalent Ising models satisfy the conditions of Theorem 1 for all T provided that (1.4) holds. On taking the limit $n \to \infty$, one concludes that all the zeros of the partition function $Z_N(H^z)$ of the general Heisenberg model described by (1.2) lie on the imaginary axis, $\mathrm{Re}\{H^z\} = 0$. [See Theorem 2 below.] The case of $S > \frac{1}{2}$ is then discussed by Suzuki's method.[7]

Finally, the ferroelectric models may be dealt with by formulating them in terms of Ising spin variables (in place of arrows drawn on bonds) and applying Theorem 1 directly. Theorem 1 is stated, discussed, and used to prove Theorem 2 and to discuss the ferroelectric models in the next section. The proof of Theorem 1 is postponed to Sec. 3.

2. STATEMENTS AND DISCUSSION

We first state Theorem 1 and discuss its significance for Ising models. We then prove Theorem 2 for Heisenberg models. Ferroelectric models are considered at the end of this section.

Theorem 1: Lee–Yang Theorem for Generalized Ising Models

Let $\sigma_j = \pm 1$, $j = 1, \cdots, N$, be a set of Ising spin variables, and let

$$q_{\nu,r}(\sigma_{j(r,1)}, \sigma_{j(r,2)}, \cdots, \sigma_{j(r,\nu)}), \quad r = 1, 2, \cdots, R \leq \binom{N}{\nu},$$

be a set of (in general, complex-valued) "partial Boltzmann factors" for the subset of ν (≥ 2) spin variables $\{\sigma_{j(r,\alpha)}\}_{\alpha=1,\cdots,\nu}$ satisfying (for all ν and r)

(A) $$q_{\nu,r}(-\sigma_{(1)}, -\sigma_{(2)}, \cdots, -\sigma_{(\nu)}) = q^*_{\nu,r}(\sigma_{(1)}, \sigma_{(2)}, \cdots, \sigma_{(\nu)}) \quad (2.1)$$

and

(B) $$|q_{\nu,r}(1, 1, \cdots, 1)| \geq \tfrac{1}{4} \sum_{\sigma_{(1)}=\pm 1} \cdots \sum_{\sigma_{(\nu)}=\pm 1} |q_{\nu,r}(\sigma_{(1)}, \cdots, \sigma_{(\nu)})|. \quad (2.2)$$

Then any zero $z = \zeta$ of the "total partition function"

$$Z_N(z) = \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} \prod_{j=1} z^{\mu_j \sigma_j} \prod_{\nu,r} q_{\nu,r}(\sigma_{j(r,1)}, \cdots, \sigma_{j(r,\nu)}), \quad (2.3)$$

in which the μ_j are real, nonnegative coefficients, satisfies

$$|\zeta| = 1.$$

Remarks on Theorem 1

(a) The partial Boltzmann factors may be written

$$q_{\nu,r} = \exp[-\beta \mathcal{H}_{\nu,r}(\sigma_{j(r,1)}, \cdots, \sigma_{j(r,\nu)})] \quad (2.4)$$

with

$$\beta = 1/k_B T, \quad (2.5)$$

where $\mathcal{H}_{\nu,r}$ is a partial Hamiltonian for the rth subset of ν spins. A partial Hamiltonian may, however, take the value $+\infty$ or, more generally, may be complex. Condition (A) represents a generalized *spin inversion symmetry* of the (zero field) Hamiltonians. When the $\mathcal{H}_{\nu,r}$ are real, (A) implies that they may be expressed as sums of products of *even* numbers of spin variables; the $q_{\nu,r}$, when real, can be expressed similarly.

(b) With $z = \exp(-\beta H^z)$, as in (1.1), the total Ising Hamiltonian in field H^z may be written

$$\mathcal{H}_N(\sigma_1, \cdots, \sigma_N) = -H^z \sum_{j=1}^{N} \mu_j \sigma_j + \sum_{\nu,r} \mathcal{H}_{\nu,r}. \quad (2.6)$$

Evidently the μ_i (≥ 0) can be interpreted as local magnetic moments; alternatively, one may regard $H_i^z = \mu_i H^z$ as a local (inhomogeneous) magnetic field.

(c) To explore the significance of condition (B), we may write

$$\begin{aligned} E_{\nu,r}^0 &= \operatorname{Re}\{\mathcal{H}_{\nu,r}(1, 1, \cdots, 1)\} \\ &= \operatorname{Re}\{\mathcal{H}_{\nu,r}(-1, -1, \cdots, -1)\} \end{aligned} \quad (2.7)$$

and define $\Delta E_{\nu,r}$ such that for all other, nonfully aligned, spin configurations we have

$$\operatorname{Re}\{\mathcal{H}_{\nu,r}(\sigma_{(1)}, \cdots, \sigma_{(\nu)})\} \geq E_{\nu,r}^0 + \Delta E_{\nu,r}. \quad (2.8)$$

Then it is easy to see that condition (B) is satisfied if, for all ν and r, we have

$$k_B T \leq k_B T_{\nu,r} = \Delta E_{\nu,r}/\ln(2^{\nu-1} - 1). \quad (2.9)$$

When only pair interactions are present ($\nu = 2$, $\mathcal{H}_{2,r} = -J_r \sigma_{r,1} \sigma_{r,2}$), this observation shows that the circle theorem is valid for *all* finite T provided $\operatorname{Re}\{J_r\} > 0$ (so that $\Delta E_{2,r} > 0$). The theorem remains valid for $T = \infty$ or for $\operatorname{Re}\{J_r\} = 0$, by the continuity of the roots of $Z_N(z) = 0$. This result, of course, includes the original Lee–Yang theorem for ferromagnetic Ising models.[1]

(d) When many-spin interactions ($\nu > 2$) are present, it is evident from (B) [or (2.9)] that the roots have been proved to lie on the unit circle only for *sufficiently low* T. However, investigation of the simplest case, namely,

$$\begin{aligned} &\mathcal{H}_4(\sigma_1, \cdots, \sigma_4) \\ &\quad = -J\left(\sum_{(ij)} \sigma_i \sigma_j + \gamma \sigma_1 \sigma_2 \sigma_3 \sigma_4\right) - H^z \sum_{j=1}^{4} \sigma_j, \end{aligned} \quad (2.10)$$

for J real and positive, reveals that the roots can indeed leave the unit circle when $\gamma \neq 0$ and T is sufficiently large. Specifically, in the special case $\gamma = 1$, the roots will lie on the circle if and only if

$$k_B T \leq 8J/\ln 2. \quad (2.11)$$

The sufficiency condition (B) yields $\ln 7$ in place of $\ln 2$. Similarly, when $\gamma = -1$, the necessary and sufficient condition is

$$k_B T \leq 4J/\ln 3, \quad (2.12)$$

whereas condition (B) gives sufficiency, $\ln(2 + \sqrt{7})$ replacing $\ln 3$. More generally, one can prove there are roots *off* the unit circle whenever

$$\begin{aligned} k_B T &\geq 4J(1 + \gamma)/\ln(1 + \tfrac{1}{3}\gamma) \quad \text{for} \quad \gamma > 0 \\ &\geq 2J(1 + |\gamma|)/\ln(1 - \tfrac{1}{3}|\gamma|^{-1}) \quad \text{for} \quad \gamma < 0. \end{aligned} \quad (2.13)$$

Furthermore, if $\gamma < -3$, there are roots off the circle for all T; this borderline is also indicated by condition (B) or (2.9), in which we find $\Delta E = 6J(1 - \frac{1}{3}|\gamma|)$ for $\gamma \simeq -3$.

(e) The theorem may be extended directly to Ising models of higher spin with many-spin interactions by the method of Griffiths.[4]

Heisenberg Model

We turn now to the fully anisotropic Heisenberg model described by the Hamiltonian $\mathcal{H}_N(H^z)$ given in (1.2), which for $S = \frac{1}{2}$ will be explicitly interpreted as a $2^N \times 2^N$ matrix given in the representation of states, $|\sigma_1, \sigma_2, \cdots, \sigma_N\rangle$, in which each S_i^z is diagonal with eigenvalues $\frac{1}{2}\sigma_i = \pm\frac{1}{2}$. The partition function is defined by

$$\begin{aligned} Z_N(H^z) &= \operatorname{tr}\{\exp[-\beta \mathcal{H}_N(H^z)]\} = \sum_{\sigma_1=\pm 1} \cdots \\ &\sum_{\sigma_N=\pm 1} \langle \sigma_1, \cdots, \sigma_N | \exp[-\beta \mathcal{H}_N(H^z)] | \sigma_1, \cdots, \sigma_N \rangle. \end{aligned} \quad (2.14)$$

The exponential of a bounded operator A is usually defined by

$$\exp(\mathsf{A}) = \sum_{t=0}^{\infty} \frac{\mathsf{A}^t}{t!}, \quad (2.15)$$

but it is straightforward to show that this is equivalent to

$$\exp(\mathsf{A}) = \lim_{n\to\infty} (\mathsf{I} + n^{-1}\mathsf{A} + n^{-2}\mathsf{C}_n)^n, \quad (2.16)$$

provided that C_n remains uniformly bounded as $n \to \infty$. It follows from this that, if

$$\mathsf{A}(w) = \sum_{u=1}^{U} \mathsf{A}_u + w \sum_{v=1}^{V} \mathsf{B}_v, \quad U, V < \infty, \quad (2.17)$$

then

$$\mathsf{E}_n(w) = \left(\prod_{u=1}^{U} (\mathsf{I} + n^{-1}\mathsf{A}_u) \prod_{v=1}^{V} \exp(wn^{-1}\mathsf{B}_v)\right)^n \quad (2.18)$$

converges *uniformly* to $\exp \mathsf{A}(w)$ on all compact sets in the complex w plane.

To utilize this formula, we rewrite the total Heisenberg Hamiltonian as

$$\mathcal{H}_N = \sum_{(i,j)} \mathcal{H}_{ij} + \sum_{i=1}^{N} \mathcal{H}_i - 2H^z \sum_{j=1}^{N} \mu_j S_j^z, \tag{2.19}$$

where

$$\mathcal{H}_{ij} = -(J_{ij}^x S_i^x S_j^x + J_{ij}^y S_i^y S_j^y + J_{ij}^z S_i^z S_j^z), \tag{2.20}$$

$$\mathcal{H}_i = -2(H_i^x S_i^x + H_i^y S_i^y). \tag{2.21}$$

From (2.14) and (2.18) we then obtain

$$Z_N(H^z) = \lim_{n\to\infty} \Phi_{N,n}(H^z), \tag{2.22}$$

where

$$\Phi_{N,n}(H^z) = \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} \langle \sigma_1, \cdots, \sigma_N | \prod_{j=1} \exp\left[(2\epsilon H^z \mu_j S_j^z)\right] \prod_{i=1}^{N} \mathsf{Q}_i \prod_{ij} \mathsf{Q}_{ij} | \sigma_1, \cdots, \sigma_N \rangle, \tag{2.23}$$

in which

$$\mathsf{Q}_i(n) = 1 - \epsilon \mathcal{H}_i, \tag{2.24}$$

$$\mathsf{Q}_{ij}(n) = 1 - \epsilon \mathcal{H}_{ij}, \tag{2.25}$$

with

$$\epsilon = \epsilon(n) = \beta/n = 1/nk_{\mathrm{B}}T. \tag{2.26}$$

(The implicit interchange of limit and summation is, of course, justified since the sum contains only finitely many terms; equally, the convergence to the limit remains uniform on compact sets in the complex H^z plane.)

Now, by inserting sets of states $|\{\sigma_{t,k}\}\rangle\langle\{\sigma_{t,k}\}|$, $t = 0, 1, \cdots$, between each factor in the full product in (2.23), we obtain the expansion

$$\begin{aligned}
&\Phi_{N,n}(H^z) \\
&= \sum_{\{\sigma_{t,k}\}=\pm 1} \prod_{k=1}^{N} z^{\mu_k \sigma_{0,k}} \langle\{\sigma_{0,k}\}| \mathsf{Q}_1 |\{\sigma_{1,k}\}\rangle \langle\{\sigma_{1,k}\}| \mathsf{Q}_2 |\{\sigma_{2,k}\}\rangle \cdots \langle\{\sigma_{N,k}\}| \mathsf{Q}_{12} |\{\sigma_{N+1,k}\}\rangle \langle\{\sigma_{N+1,k}\}| \mathsf{Q}_{13} |\{\sigma_{N+2,k}\}\rangle \\
&\quad \times \cdots \times \prod_{k=1}^{N} z^{\mu_k \sigma_{N'',k}} \langle\{\sigma_{N'',k}\}| \mathsf{Q}_1 |\{\sigma_{N''+1,k}\}\rangle \cdots \prod_{k=1}^{N} z^{\mu_k \sigma_{N''(n-1),k}} \langle\{\sigma_{N''(n-1),k}\}| \mathsf{Q}_1 |\{\sigma_{N''(n-1)+1,k}\}\rangle \cdots,
\end{aligned} \tag{2.27}$$

where now

$$z = \exp(\epsilon H^z) \tag{2.28}$$

and

$$N'' = N + N', \tag{2.29}$$

in which N' is the number of (i,j) pairs for which $\mathcal{H}_{ij}$ does not vanish. The matrix elements of Q_i and Q_{ij} simplify as follows:

$$\langle\{\sigma_{t,k}\}| \mathsf{Q}_i |\{\sigma_{u,k}\}\rangle = q_{2,i}(\sigma_{t,i}, \sigma_{u,i}) \prod_{k\neq i} \delta(\sigma_{t,k}, \sigma_{u,k}), \tag{2.30}$$

$$\begin{aligned}
&\langle\{\sigma_{t,k}\}| \mathsf{Q}_{ij} |\{\sigma_{u,k}\}\rangle \\
&\quad = q_{4,ij}(\sigma_{t,i}, \sigma_{t,j}, \sigma_{u,i}, \sigma_{u,j}) \prod_{k\neq i,j} \delta(\sigma_{t,k}, \sigma_{u,k}),
\end{aligned} \tag{2.31}$$

where $\delta(\sigma, \sigma') = \frac{1}{2}(1 + \sigma\sigma')$ is a Kronecker δ function. Because of the δ functions a substantial reduction of (2.27) occurs until, finally, $\Phi_{N,n}$ is expressed as a sum over at most $N(N+1)n$ Ising spins, which can be regarded as forming a lattice with periodic boundary conditions in the "n direction." (The exact structure of this lattice depends on the ordering chosen for the Q_{ij}, which do not, of course, commute; however, the structure does not matter for our purposes.) The magnetic fields H^z act only on n "layers" of N spins each, but can be formally extended to all spins $\sigma_{t,k}$ merely by defining coefficients $\mu_{t,k} \geq 0$ which vanish when appropriate. In addition to the Ising magnetic field, there are evidently pair interactions with partial Boltzmann factors which, by (2.24) and (2.30), can be written in an obvious matrix notation as

$$q_{2,i}(\sigma, \sigma') \equiv \begin{bmatrix} 1 & \epsilon(H_i^x + iH_i^y) \\ \epsilon(H_i^x - iH_i^y) & 1 \end{bmatrix}. \tag{2.32}$$

Lastly, there are also four-spin interactions with partial Boltzmann factors

$$\begin{aligned}
&q_{4,i,j}(\sigma_1, \sigma_2, \sigma_1', \sigma_2') \\
&\equiv \begin{bmatrix} 1+\epsilon J_{ij}^z & 0 & 0 & \epsilon(J_{ij}^x - J_{ij}^y) \\ 0 & 1-\epsilon J_{ij}^z & \epsilon(J_{ij}^x + J_{ij}^y) & 0 \\ 0 & \epsilon(J_{ij}^x + J_{ij}^y) & 1-\epsilon J_{ij}^z & 0 \\ \epsilon(J_{ij}^x - J_{ij}^y) & 0 & 0 & 1+ J_{ij}^z \end{bmatrix}.
\end{aligned} \tag{2.33}$$

In order to apply Theorem 1 to this Ising model, we must check conditions (A) and (B). Provided that H_i^x, H_i^y and J_{ij}^x, J_{ij}^y, and J_{ij}^z are *real*, the spin inversion symmetry condition (A) is easily seen to be satisfied by (2.32) and (2.33). (Note that q_2 will generally be complex.) For $q_{2,i}$, condition (B) yields

$$1 > \tfrac{1}{2} + \tfrac{1}{2}\epsilon[(H_i^x) + (H_i^y)^2]^{\frac{1}{2}}. \tag{2.34}$$

Provided that $T \neq 0$ (and $|H_i^x|$, $|H_i^y| < \infty$), this will always be valid for large enough n [see (2.26)]. With the same proviso (and $|J_{ij}^z| < \infty$), condition (B) with (2.33) yields

$$2J_{ij}^z \geq |J_{ij}^x + J_{ij}^y| + |J_{ij}^x - J_{ij}^y|, \tag{2.35}$$

which is easily seen to be equivalent to

$$J_{ij}^z \geq |J_{ij}^x| \quad \text{and} \quad J_{ij}^z \geq |J_{ij}^y|. \tag{2.36}$$

This condition is the most natural characterization of a *ferromagnetic Heisenberg model with dominant z–z coupling*. It now follows from Theorem 1 that all the zeros of $\Phi_{N,n}(H^z)$ lie on the imaginary axis $\mathrm{Re}\,\{H^z\} = 0$ whenever (2.36) is satisfied and all the μ_j are real and nonnegative.

Finally, in order to take the limit $n \to \infty$, we note that, by (2.27) and (2.28), the function $\Phi_{N,n}(H^z)$ is entire analytic in H^z for all n, and recall that $\Phi_{N,n}(H^z)$ converges uniformly to $Z_N(H^z)$. An appeal to Hurwitz's theorem[13] then ensures that any zeros of $Z_N(H^z)$ lie on the imaginary H^z axis. For $S = \frac{1}{2}$ we have therefore proved Theorem 2.

Theorem 2: Lee–Yang Theorem for Heisenberg Models

All zeros of the partition function $Z_N(H^z)$ [defined in (2.14)], for the real Heisenberg Hamiltonian $\mathcal{H}_N$ defined in (2.19)–(2.21), lie on the imaginary H^z axis ($\mathrm{Re}\,\{H^z\} = 0$) provided that

(C) $T > 0$, $\mu_j \geq 0$, $J_{ij}^z \geq |J_{ij}^x|$, and $J_{ij}^z \geq |J_{ij}^y|$.

To prove the theorem for general S, we may use the method of Suzuki[7] to reduce a general spin Hamiltonian to one for spin $\frac{1}{2}$. For completeness, we outline the procedure here. Suppose, generally, that for spin S we have

$$\mathcal{H}_N = \mathcal{F}(\{S_i^\alpha\}), \quad \alpha = x, y, z, \quad i = 1, 2, \cdots, N. \tag{2.37}$$

Let s^α denote spin operators for $S = \frac{1}{2}$ and then extend $\mathcal{H}_N$ to a spin-$\frac{1}{2}$ Hamiltonian of $2SN$ spins, namely

$$\mathcal{H}_N^{(\frac{1}{2})} = \mathcal{F}\left(\left\{\sum_{t=1}^{2S} s_{i,t}^\alpha\right\}\right). \tag{2.38}$$

Now the operator

$$\mathcal{S}_i^2 = \sum_\alpha \left(\sum_{t=1}^{2S} s_{i,t}^\alpha\right)^2 \tag{2.39}$$

clearly commutes with $\mathcal{H}_N^{(\frac{1}{2})}$. Hence the eigenvectors of $\mathcal{H}_N^{(\frac{1}{2})}$ can be classified by the set $\{S_i'\}$ which specifies the eigenvalues $S_i'(S_i' + 1)$ of the $\mathcal{S}_i^2$. If $\mathcal{P}_N$ is a projection operator onto the subspace specified by $S_i' = S$ for all i, we then have

$$Z_N = \mathrm{tr}^{(S)}\,[\exp(-\beta\mathcal{H}_N)] = \mathrm{tr}^{(\frac{1}{2})}\,[\mathcal{P}_N \exp(-\beta\mathcal{H}_N^{(\frac{1}{2})})], \tag{2.40}$$

where $\mathrm{tr}^{(\frac{1}{2})}$ denotes a trace in the full spin-$\frac{1}{2}$ space. Now a convenient form for the projection operator is

$$\begin{aligned}\mathcal{P}_N &= \lim_{\xi\to\infty} \prod_{i=1}^{N} \exp\{-\tfrac{1}{2}\beta\xi[S(S+1) - \mathcal{S}_i^2]\}, \\ &= \lim_{\xi\to\infty} \exp(-\beta\xi\mathcal{G}_N^{(\frac{1}{2})}),\end{aligned} \tag{2.41}$$

where

$$\mathcal{G}_N^{(\frac{1}{2})} = \sum_{i=1}^{N} \sum_{(t,u)} (\tfrac{1}{4} - \mathbf{s}_{i,t} \cdot \mathbf{s}_{i,u}), \tag{2.42}$$

in which the second sum is over distinct pairs (t, u). Finally, we have

$$Z_N = Z_N^{(\frac{1}{2})} = \lim_{\xi\to\infty} \mathrm{tr}^{(\frac{1}{2})}\{\exp[-\beta(\mathcal{H}_N^{(\frac{1}{2})} + \xi\mathcal{G}_N^{(\frac{1}{2})})]\}, \tag{2.43}$$

which represents a spin-$\frac{1}{2}$ Heisenberg Hamiltonian with additional pair interaction terms given by (2.42). To apply Theorem 1 to this Hamiltonian, we may include the additional terms in the exponential, or B factors in (2.18). The corresponding partial Boltzmann factors corresponding to (2.33) are then easily found to be

$$q_4(\xi; \sigma_1, \sigma_2, \sigma_1', \sigma_2') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2}(1 + e^{-\epsilon\xi}) & \frac{1}{2}(1 - e^{-\epsilon\xi}) & 0 \\ 0 & \frac{1}{2}(1 - e^{-\epsilon\xi}) & \frac{1}{2}(1 + e^{-\epsilon\xi}) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.44}$$

This expression clearly satisfies the spin-inversion symmetry condition (A) of Theorem 1; it also satisfies condition (B) (as an equality) for all ξ, so that the limit $\xi \to \infty$ in (2.43) may be taken when convenient. The pair terms in $\mathcal{H}_N^{(\frac{1}{2})}$ arising from those in $\mathcal{H}_N$ still satisfy (A) and (B) provided that (2.36) holds. This completes the proof.

Extensions of Theorem 2

(a) The theorem is easily extended to pair interactions of the form

$$\mathcal{H}_{ij} = -\sum_\alpha \sum_\beta J_{ij}^{\alpha\beta} S_i^\alpha S_j^\beta, \quad \alpha, \beta = x, y, z. \tag{2.45}$$

In order that the corresponding (linearized) partial Boltzmann factor satisfy condition (A) of Theorem 1,

we must require, dropping the subscripts ij, that

$$J^{xx}, J^{yy}, J^{zz} \text{ and } J^{xy}, J^{yx} \text{ be real, but}$$
$$J^{xz}, J^{yz}, J^{zx}, J^{zy} \text{ be pure imaginary.} \quad (2.46)$$

Condition (B) is then satisfied if

$$\begin{aligned} 2J^{zz} \geq\ & [(J^{xx} - J^{yy})^2 + (J^{xy} + J^{yx})^2]^{\frac{1}{2}} \\ & + [(J^{xx} + J^{yy})^2 + (J^{xy} - J^{yx})^2]^{\frac{1}{2}} \\ & + 2(|J^{xz}|^2 + |J^{yz}|^2)^{\frac{1}{2}} \\ & + 2(|J^{zx}|^2 + |J^{zy}|^2)^{\frac{1}{2}}, \end{aligned} \quad (2.47)$$

which may be compared with (2.35), to which it reduces when $J^{\alpha\beta} \equiv 0$ for $\alpha \neq \beta$. One should note, however, that if J^{xz}, J^{yz}, J^{zx}, or J^{zy} do not vanish but are imaginary, then the Hamiltonian will not be Hermitian.

(b) If ν-spin interactions are introduced into the Hamiltonian, they may be handled as above. However, for $\nu > 2$ they lead to difficulties since large n corresponds essentially to high T, and it will not, in general, be possible to satisfy condition (B) as $n \to \infty$ at fixed T. An exception arises if the many-spin interactions involve only S_i^z operators and commute with the rest of the Hamiltonian. The many-spin terms are then diagonal in the Ising interactions, and bounds valid for sufficiently low T can be obtained as in the pure Ising model. But this special case seems of limited interest.

(c) Finally, we remark that, just as in the case of the Ising model, the restriction of all zeros to the imaginary H^z-axis ensures that the limiting thermodynamic free energy of any Heisenberg model satisfying the "ferromagnetic" condition (C) is analytic for H^z real and nonzero and hence that a phase transition can occur only in zero magnetic field. Techniques like those used in proving the theorems can also be used to prove Griffiths'[14] first inequalities for Heisenberg spin correlation functions, namely,

$$\langle S_i^z S_j^z \cdots S_l^z \rangle \geq 0 \quad \text{for} \quad H^z \geq 0, \quad (2.48)$$

under condition (C). However, Gallavotti[15] has already given a proof under the more general condition

$$\text{(D)} \quad J_{ij}^z \geq |J_{ij}^x| \quad \text{or} \quad J_{ij}^z \geq |J_{ij}^y| \text{ for all } (i,j), \quad (2.49)$$

which is *not* expected to be sufficient to ensure that the zeros lie on the imaginary H^z axis.

Ferroelectric Models

Lattice models of ice and hydrogen-bonded ferroelectrics[16] which have recently been the subject of significant numerical[17] and exact analysis[18] are usually represented by drawing arrows on the bonds of an appropriate four-coordinated lattice. The "ice rules" then state that two arrows must point towards and two point away from each vertex. Different energies are assigned to the different configurations around a vertex, and a dipole moment is associated with each bond. In order to apply Theorem 1 to such models, we will, however, formulate them, and their generalizations, directly in terms of Ising spins.[19] Specifically, we associate a spin variable σ_i with the ith bond according to the following convention:

"vertical" bonds:

$$\sigma = +1 \Leftrightarrow \uparrow \text{ upward arrow}$$
$$= -1 \Leftrightarrow \downarrow \text{ downward arrow;}$$

"horizontal" bonds:

$$\sigma = +1 \Leftrightarrow \rightarrow \text{ rightward arrow}$$
$$= -1 \Leftrightarrow \leftarrow \text{ leftward arrow.}$$

Our discussion will, in fact, apply to an arbitrary tetrahedrally coordinated lattice so that "vertical" and "horizontal" have only the local significance of labeling the two pairs of bonds incident with a particular vertex. Table I lists all sixteen possible configurations about a vertex. The first six configurations, with associated energies ϵ_1, ϵ_2, and ϵ_3, are the allowed ("neutral") ice configurations. We will, however, consider the most general inversion-invariant model with eight distinct bond weights $\epsilon_1, \cdots, \epsilon_8$ as shown in Table I. This model evidently reduces to the standard ferroelectric models by letting $\epsilon_4, \cdots, \epsilon_8$ approach $+\infty$, in which limit the remaining "ionized" configurations cannot occur.

We suppose that the ith bond has an associated dipole moment $\bar{\mu}_i$ and introduce an electric field $\mathcal{E}$ with "direction cosines" l_h or l_v for horizontal or vertical bonds, respectively. We assume the l_i are nonnegative so that the field direction lies in the "first quadrant." The interaction of the system with the field is then described by the Hamiltonian

$$\mathcal{H}_{\mathcal{E}} = -\mathcal{E} \sum_{i=1}^{N} \mu_i \sigma_i \quad \text{with} \quad \mu_i = l_i \bar{\mu}_i \geq 0. \quad (2.50)$$

The model is now clearly equivalent to an Ising model in a field $H \equiv \mathcal{E}$ with at most four-spin interactions: Specifically, a partial Hamiltonian

$$\begin{aligned} \mathcal{H}_r = & -J_0 - (J_1 \sigma_{r1}\sigma_{r2} + J_2 \sigma_{r2}\sigma_{r3} \\ & + J_3 \sigma_{r3}\sigma_{r4} + J_4 \sigma_{r4}\sigma_{r1} + J_5 \sigma_{r1}\sigma_{r3} \\ & + J_6 \sigma_{r2}\sigma_{r4}) - J_7 \sigma_{r1}\sigma_{r2}\sigma_{r3}\sigma_{r4} \end{aligned} \quad (2.51)$$

is associated with each vertex r of the original ferroelectric lattice. Here the spins $\sigma_{r1}, \sigma_{r2}, \sigma_{r3}$, and σ_{r4} correspond to the four bonds incident at r. After some

TABLE I. List of all possible vertex configurations and their weights for the generalized ferroelectric models. Plus spins correspond to upward and rightward pointing arrows; minus spins to downward and leftward pointing arrows. Configurations 1–6 are the allowed "ice" configurations. Configurations 7 and 8 are "doubly ionized," while 9–16 are the eight "singly ionized" configurations.

(1)	(3)	(5)	(7)
(2)	(4)	(6)	(8)
ϵ_1	ϵ_2	ϵ_3	ϵ_4
(9)	(11)	(13)	(15)
(10)	(12)	(14)	(16)
ϵ_5	ϵ_6	ϵ_7	ϵ_8

algebra one finds the explicit relations

$$8J_0 = -\sum_{j=1}^{8}\epsilon_j, \quad 8J_7 = \sum_{j=5}^{8}\epsilon_j - \sum_{j=1}^{4}\epsilon_j \tag{2.52}$$

and, with an obvious notation,

$$\begin{aligned} 8J_1 &= \{2,4,6,8\} - \{1,3,5,7\},\\ 8J_2 &= \{2,3,6,7\} - \{1,4,5,8\},\\ 8J_3 &= \{2,4,5,7\} - \{1,3,6,8\},\\ 8J_4 &= \{2,3,5,8\} - \{1,4,6,7\},\\ 8J_5 &= \{3,4,7,8\} - \{1,2,5,6\},\\ 8J_6 &= \{3,4,5,6\} - \{1,2,7,8\}, \end{aligned} \tag{2.53}$$

where the bonds have been numbered in a clockwise sense around the vertex starting at the left-hand horizontal bond (see Table I).

However, in order to apply Theorem 1, we do not actually need these explicit formulas since condition (B) is simply expressed by

$$e^{-\beta\epsilon_1} \geq \sum_{j=2}^{8} e^{-\beta\epsilon_j}. \tag{2.54}$$

The inversion-symmetry condition (A) is clearly satisfied for real ϵ_j (see Table I). We will analyze the following four special cases.

1. KDP *Model*

In this case we have[16–18]

$$\epsilon_1 = 0, \quad \epsilon_2 = \epsilon_3 = \epsilon,$$

and

$$\epsilon_4 = \cdots = \epsilon_8 = +\infty. \tag{2.55}$$

Condition (B) is then equivalent to

$$T \leq T_0, \tag{2.56}$$

where by (2.54) we have

$$T_0 = \epsilon/k_B \ln 2, \tag{2.57}$$

which happens to be identical with the exact critical or phase transition temperature[16–18] of the model! Thus, for all temperatures below the critical temperature, the zeros of the partition function of the KDP model lie on the imaginary axis in the complex electric field plane. This interesting result might, perhaps, have been anticipated in view of the recent numerical studies by Katsura, Abe, and Ohkouchi[20] on small planar KDP models. The numerical work shows that the zeros leave the imaginary axis above T_c (and, indeed, apparently, fill out a 2-dimensional region).

2. *Model with Double Ions*

Sutherland[21] has recently discussed models with the energies

$$\begin{aligned} &\epsilon_1 = -\tfrac{1}{2}\delta, \quad \epsilon_2 = \tfrac{1}{2}\delta, \quad \epsilon_3 = -\epsilon,\\ &\epsilon_4 = -\gamma, \quad \epsilon_5 = \cdots = \epsilon_8 = +\infty, \end{aligned} \tag{2.58}$$

in which the "doubly ionized" configurations 7 and 8 are allowed (all four arrows inwards or all four outwards). Condition (B) then reduces to $T \leq T_1$, where T_1 is the root of

$$2\sinh(\tfrac{1}{2}\delta/k_B T_1) = \exp(\epsilon/k_B T_1) + \exp(\gamma/k_B T_1). \tag{2.59}$$

On the other hand, by an argument somewhat similar to that used originally by Kramers and Wannier for the Ising model,[22] Sutherland has argued that the transition temperature of the model when the original ferroelectric lattice is the plane square lattice should be just $T_c = T_1$.[21] Again we conclude that the zeros must lie on the imaginary $\mathcal{E}$-axis for all $T \leq T_c$.

3. *Reducible Models*

When $J_7 = 0$, that is, by (2.52), when

$$\epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 = \epsilon_5 + \epsilon_6 + \epsilon_7 + \epsilon_8, \quad (2.60)$$

the general model reduces to an Ising model with only pair interactions. The zeros of the partition function will then lie only on the imaginary $\mathcal{E}$-axis at all temperatures provided that $J_i \geq 0$ for $i = 1, 2, \cdots, 6$. By Eqs. (2.53) these conditions reduce to

$$\begin{aligned} \epsilon_2 &\leq \epsilon_5 + \epsilon_7 \leq \epsilon_3 + \epsilon_4, \\ \epsilon_3 &\leq \epsilon_5 + \epsilon_6 \leq \epsilon_2 + \epsilon_4, \\ \epsilon_4 &\leq \epsilon_6 + \epsilon_7 \leq \epsilon_2 + \epsilon_3, \end{aligned} \quad (2.61)$$

where ϵ_8 is given by (2.60) and, with no loss of generality, we have set $\epsilon_1 \equiv 0$. These inequalities are not empty, for suppose ϵ_2, ϵ_3, and ϵ_4 are the lengths of the sides of a triangle $P_2P_3P_4$ opposite the vertices P_2, P_3, and P_4, respectively. Then, if Q is any interior point of the triangle and ϵ_5, ϵ_6, and ϵ_7 are identified as the lengths of QP_4, QP_2, and QP_3, respectively, it is easy to check that the inequalities (2.61) are satisfied. (In fact, there are many more solutions than these.)

4. *Exactly Soluble Models*

It is worthwhile pointing out that among the reducible models with $J_7 \equiv 0$ there is a rather large class of models which are exactly soluble (for zero field, $\mathcal{E} \equiv 0$) when the original ferroelectric lattice is planar (i.e., has no crossing bonds when drawn on a plane). Specifically, suppose that $J_5 \equiv 0$ or that $J_6 \equiv 0$: In these cases the remaining spin interaction (or Ising lattice) bonds around a vertex form a square with one diagonal. This is a planar graph, and hence the corresponding Ising lattice is planar and its partition function can be calculated by the Pfaffian or dimer method.[23] The condition $J_5 = J_7 = 0$ implies

$$\epsilon_1 + \epsilon_2 = \epsilon_7 + \epsilon_8 \quad \text{and} \quad \epsilon_3 + \epsilon_4 = \epsilon_5 + \epsilon_6, \quad (2.62)$$

while $J_6 = J_7 = 0$ implies

$$\epsilon_1 + \epsilon_2 = \epsilon_5 + \epsilon_6 \quad \text{and} \quad \epsilon_3 + \epsilon_4 = \epsilon_7 + \epsilon_8. \quad (2.63)$$

When the original ferroelectric or bond lattice is a plane square lattice, the corresponding Ising spin lattice will be a triangular lattice with every row of, say, horizontal bonds missing. This lattice is a degenerate form of Utiyama's general checker board lattice,[24] and the exact partition function (and hence critical point) has been given long ago.[24]

This class of soluble models includes those discussed recently by Wu[19] as "extended antiferroelectric models" which were specified generally by

$$\epsilon_1 = \epsilon_2 = \epsilon, \quad \epsilon_3 = 0, \quad \epsilon_4 = b\epsilon,$$

and

$$\epsilon_5 = \epsilon_6 = \epsilon_7 = \epsilon_8 = a\epsilon. \quad (2.64)$$

When $a = 1$ and $b = 2$, we find $J_5 = J_6 = J_7 = 0$; if the ferroelectric (bond) lattice is a plane square lattice, so is the corresponding Ising lattice, and hence $k_B T_c = \frac{1}{2}\epsilon \ln(1 + \sqrt{2})$ as found by Wu.[19]

3. CIRCLE THEOREM FOR GENERAL ISING MODELS

In this section we provide the proof of Theorem 1 which is stated in Sec. 2. As explained, we will introduce a special Lee–Yang class of functions and proceed to develop their essential properties. To this end, let $F(z_1, \cdots, z_n)$ denote a multinomial (of finite degree) in the complex variables z_i and z_i^{-1}, $i = 1, 2, \cdots, n$.

Definition 1: We say F is of Lee–Yang type with respect to the variable z_j and write $F(z_i, \cdots, z_n) \subset \mathbf{L}_j$, provided that any root $z_j = \zeta_j(\{z_i\}_{i \neq j})$ of the equation

$$F(z_1, \cdots, z_j, \cdots, z_n) = 0 \quad (3.1)$$

satisfies

$(\mathbf{L}_j)$ $|\zeta_j| < 1$ whenever $|z_i| \geq 1$ *for all* $i \neq j$ and $|z_k| > 1$ *for some* $k \neq j$.

With this definition we have

Lemma 1: If $F(z_1, \cdots, z_n) \subset \mathbf{L}_j$, then a root $z_j = \zeta_j$ of (3.1) satisfies

$(\mathbf{L}_j')$ $|\zeta_j| \leq 1$ whenever $|z_i| \geq 1$ *for all* $i \neq j$. (3.2)

Proof: The lemma follows directly from Definition 1 and the continuity of the roots of a polynomial as functions of their coefficients, which are here continuous functions (in fact, multinomials) of the z_i for $i \neq j$.

Lemma 2: The relation $F(z_1, \cdots, z_n) \subset \mathbf{L}_j$ implies $F(z_1, \cdots, z_n) \subset \mathbf{L}_h$ for any h and hence for all $h = 1, 2, \cdots, n$.

Proof: We may suppose $h \neq j$. By Definition 1, the conditions

$$|z_i| \geq 1 \quad \text{for all } i$$

and

$$|z_k| > 1 \quad \text{for some} \quad k \neq j, h$$

imply

$$F(z_1, \cdots, z_j, \cdots, z_n) \neq 0. \tag{3.3}$$

Hence any root $z_h = \zeta_h$ of

$$F(z_1, \cdots, z_h, \cdots, z_n) = 0 \tag{3.4}$$

must satisfy

$$(L_h)_{k\neq j} \quad |\zeta_h| < 1 \quad \text{whenever} \quad |z_i| \geq 1 \ \textit{for all} \ i \neq h \quad \text{and} \quad |z_k| > 1 \ \textit{for some} \ k \neq j, h.$$

On the other hand, by Lemma 1, the conditions

$$|z_i'| \geq 1 \quad \text{for all } i, \text{ and } \quad |z_j| > 1,$$

also imply (3.3) above. Hence any root ζ_h of (3.4) must also satisfy

$$(L_h)_{k=j} \quad |\zeta_h| < 1 \quad \text{whenever} \quad |z_i| \geq 1 \quad \text{for all} \quad i \neq h \quad \text{and} \quad |z_j| > 1.$$

The properties $(L_h)_{k\neq j}$ and $(L_h)_{k=j}$ together clearly imply (L_h) as required to prove the lemma.

Lee–Yang Class: As a consequence of this lemma, we may say a multinomial $F(z_1, \cdots, z_n)$ is of *Lee–Yang type*, or belongs to $\mathbf{L}$, if $F(z_1, \cdots, z_n) \subset \mathbf{L}_j$ for any $j = 1, \cdots, n$. This then implies $F \subset \mathbf{L}_j$ for all j.

Complement of Lee–Yang Class: For later use it is convenient to state here a characterization of a function which is *not* of Lee–Yang type, namely:

If $F(z_1, \cdots, z_n) \notin \mathbf{L}$, then there exists a zero $F = 0$ for a set of values $z_i = z_i^{(0)}$ (all i) and a partition $\{i_1, i_2, \cdots, i_\mu \mid i_{\mu+1}, \cdots, i_n\}$ with $1 \leq \mu \leq n$ of the set of indices $\{i\}$ such that

$$|z_{i_\alpha}^{(0)}| > 1 \quad \text{for} \quad \alpha \leq \mu \quad \text{and} \quad |z_{i_\beta}^{(0)}| = 1 \quad \text{for} \quad \beta > \mu. \tag{3.5}$$

The validity of this characterization is evident on a moment's reflection.

Now if $F(z_1, z_2, \cdots, z_n)$ belongs to $\mathbf{L}$, it is clear that the "reduced" functions

$$F(z^{\mu_1}, z^{\mu_2}, z_2, \cdots, z_n), F(z^{\mu_1}, z^{\mu_3}, z^{\mu_2}, z_4, \cdots, z_n), \cdots$$

also belong to $\mathbf{L}$ provided that the μ_i are real and nonnegative since the relations $|z_j| \geq 1$, $|z_k| > 1$, etc., will be preserved. An immediate consequence of Lemma 1 is then

Lemma 3 (Lee–Yang zeros): If $F(z_1, \cdots, z_n) \subset \mathbf{L}$ and $\mu_1, \mu_2, \cdots, \mu_n$ are real and nonnegative, then any root ζ of the equation

$$F(z^{\mu_1}, z^{\mu_2}, \cdots, z^{\mu_n}) = 0 \tag{3.6}$$

satisfies

$$|\zeta| = 1. \tag{3.7}$$

This, of course, is just the type of result we wish to prove for the partition function [with z given by (1.1)]. To discuss partitionlike functions, we introduce the following definition.

Definition 2: Transform of a Spin Function: If $f(\sigma_1, \cdots, \sigma_n)$ is a function (in general, complex valued) of the n spin variables $\sigma_i = \pm 1$, then the (normalized) *trace* operation is defined by

$$\mathrm{Tr}\,\{f\} = 2^{-n} \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_n=\pm 1} f(\sigma_1, \cdots, \sigma_n), \tag{3.8}$$

and hence the *transform* $F = \hat{f}$ by

$$F(z_1, \cdots, z_n) = \mathrm{Tr}\left\{\prod_{i=1} z_i^{\sigma_i} f(\sigma_1, \cdots, \sigma_n)\right\}. \tag{3.9}$$

The transform is a multinomial in z_i and z_i^{-1} and hence

$$\text{when} \quad \hat{f} = F(z_1, \cdots, z_n) \subset \mathbf{L},$$
$$\text{we say} \quad f(\sigma_1, \cdots, \sigma_n) \subset \boldsymbol{l},$$

or, in words, that f is a (spin) function of Lee–Yang type.

We next need

Lemma 4: If $\hat{f} = F(z_1, \cdots, z_n) \subset \mathbf{L}$, then the coefficient A_j of z_j in the expansion of F cannot vanish when $|z_i| \geq 1$ for all $i \neq j$.

Proof: Without loss of generality, suppose $j = 1$ and, by (3.9), write

$$F = A_1 z_1 + B_1 z_1^{-1}. \tag{3.10}$$

Suppose, contrary to the lemma, that $A_1(z_2, \cdots, z_n)$ vanishes when $z_j = \zeta_j$, $j = 2, \cdots, n$, with $|\zeta_j| \geq 1$ for all $j \geq 2$. Then the equation $F(z_1) = 0$ either has an infinite root ($z_1^{-1} = 0$) or, in the event

$$B_1(\zeta_2, \cdots, \zeta_n) = 0,$$

has an arbitrary root. In either case there is a root ζ_1 with $|\zeta_1| > 1$; but this contradicts the datum $F \subset \mathbf{L} \equiv \mathbf{L}_1$. Hence the hypothesis is false, and the lemma is proved. (Asano has proved this lemma in a somewhat different way.)

We dignify the next two steps as propositions; they deal with the composition of spin functions and hence the building up of an Ising lattice by taking products and identifying spin variables.

Proposition 1: If

$$f(\sigma_0; \sigma_1, \cdots, \sigma_m) \quad \text{and} \quad g(\sigma_0'\sigma_1', \cdots, \sigma_n')$$

belong to $\boldsymbol{l}$, then so does

$$h(\sigma; \sigma_1, \cdots, \sigma_m, \sigma_1', \cdots, \sigma_n') = f(\sigma; \sigma_1, \cdots, \sigma_m) g(\sigma; \sigma_1', \cdots, \sigma_n'), \quad (3.11)$$

in which f and g have been *linked* by the *contraction* $\sigma_0 = \sigma_0' = \sigma$.

Proof: By definition the transform of h is

$$H(z; z_1, \cdots, z_m, z_1', \cdots, z_n') = \tfrac{1}{2} z \operatorname{Tr}\left\{\prod_{i\geq 1} z_i^{\sigma_i} \prod_{j\geq 1} z_j'^{\sigma_j'} h(1; \sigma_1, \cdots, \sigma_n')\right\} + \tfrac{1}{2} z^{-1} \operatorname{Tr}\left\{\prod_{i\geq 1} z_i^{\sigma_i} \prod_{j\geq 1} z_j'^{\sigma_j'} h(-1; \sigma_1, \cdots, \sigma_n')\right\}. \quad (3.12)$$

But, on using (3.11), the traces factorize, and we obtain

$$H = 2A_0 A_0' z + 2B_0 B_0' z^{-1}, \quad (3.13)$$

where

$$A_0(z_1, \cdots, z_m) = \tfrac{1}{2} \operatorname{Tr}\left\{\prod_{i\geq 1} z_i^{\sigma_i} f(1; \sigma_1, \cdots, \sigma_m)\right\} \quad (3.14)$$

is just the coefficient of z_0 in $\hat{f} = F(z_0, z_1, \cdots, z_m)$, while

$$B_0(z_1, \cdots, z_m) = \tfrac{1}{2} \operatorname{Tr}\left\{\prod_{i\geq 1} z_i^{\sigma_i} f(-1; \sigma_1, \cdots, \sigma_m)\right\} \quad (3.15)$$

is the coefficient of z_0^{-1} in $\hat{f}$. Similarly $A_0'(z_1', \cdots, z_n')$ and $B_0'(z_1', \cdots, z_n')$ are the coefficients of z_0' and $(z_0')^{-1}$ in $\hat{g}$. By Lemma 2, it is sufficient to prove that $H(z; z_1, \cdots, z_n')$ is of Yang–Lee type with respect to the first variable z. To this end, we suppose that

$$|z_i| \geq 1, \quad i \geq 1, \quad \text{and} \quad |z_j'| \geq 1, \quad j \geq 1,$$
$$|z_k| > 1 \quad \text{for some } k \text{ or } \quad |z_l'| > 1 \quad \text{for some } l. \quad (3.16)$$

By Lemma 4 and the data $f \subset \boldsymbol{l}$ and $g \subset \boldsymbol{l}$, it then follows that neither A_0 nor A_0' vanish. Consequently, by (3.13), any zero $z = \zeta$ of the equation $H(z) = 0$ verifies

$$|\zeta|^2 = |B_0/A_0|\,|B_0'/A_0'|. \quad (3.17)$$

In order to bound the ratios involved, consider the equation

$$F(z_0) = A_0 z_0 + B_0 z_0^{-1} = 0, \quad (3.18)$$

with roots ζ_0 satisfying

$$\zeta_0^2 = -(B_0/A_0). \quad (3.19)$$

Since $f \subset \boldsymbol{l}$, it follows from (3.16) and Lemma 1 that $|\zeta_0| \leq 1$ and hence that $|B_0/A_0| \leq 1$. Similarly, we have $|B_0'/A_0'| \leq 1$. In addition, by the second part of (3.16), at least one of these ratios is strictly less than unity. Combining these results in (3.17) proves that $|\zeta| < 1$ under condition (3.16) and thereby completes the proof.

Lastly, we will require:

Proposition 2: If $f(\sigma_1, \cdots, \sigma_n)$ belongs to $\boldsymbol{l}$, then so does

$$g(\{\sigma\}_{i\neq j,k}, \sigma) = [f(\sigma_1, \cdots, \sigma_m)]_{\sigma_j=\sigma_k=\sigma}, \quad (3.20)$$

in which the spin variables σ_j and σ_k have been *contracted.*

Proof: Without loss of generality, we may take $j = 1$ and $k = 2$. The transform $\hat{g} = G$ can then be written

$$G(z, z_3, \cdots, z_n) = Az + Dz^{-1}, \quad (3.21)$$

where

$$A = \tfrac{1}{2} \operatorname{Tr}\left\{\prod_{i\geq 3} z_i^{\sigma_i} f(1, 1, \sigma_3, \cdots, \sigma_n)\right\} \quad (3.22)$$

and

$$D = \tfrac{1}{2} \operatorname{Tr}\left\{\prod_{i\geq 3} z_i^{\sigma_i} f(-1, -1, \sigma_3, \cdots, \sigma_n)\right\}. \quad (3.23)$$

As before, it is sufficient to prove that G is of Lee–Yang type with respect to z alone. Hence we suppose that

$$|z_i| \geq 1, \quad \text{for all} \quad i \geq 3,$$
$$|z_h| > 1, \quad \text{for some} \quad h \geq 3. \quad (3.24)$$

Now consider

$$\hat{f} = F(z_1, \cdots, z_n) = \tfrac{1}{2} A z_1 z_2 + B z_1 z_2^{-1} + C z_1^{-1} z_2 + \tfrac{1}{2} D z_1^{-1} z_2^{-1}, \quad (3.25)$$

where the coefficients B and C arise from sums over $f(1, -1, \sigma_3, \cdots)$ and $f(-1, 1, \sigma_3, \cdots)$. Now since $F \subset \mathbf{L}$, the coefficient of z_1, namely, $\frac{1}{2} A z_2 + B z_2^{-1}$, cannot, by Lemma 4, vanish under (3.24) with $|z_2| \geq 1$. Hence $|A| > 2\,|B|$ so that A itself cannot vanish. Furthermore, under condition (3.24), any root $z_0 = \zeta_0$ of the reduced equation ($z_1 = z_2$)

$$F(z_0, z_0; z_3, \cdots, z_n) = 0 \quad (3.26)$$

satisfies

$$|\zeta_0| < 1. \quad (3.27)$$

From (3.25) we see that the product of all roots ζ_0 is simply D/A, which is, hence, of modulus less than unity. Finally, under condition (3.24), any root $z = \zeta$ of the equation $G(z) = 0$ will by (3.21) hence satisfy

$$|\zeta|^2 = |D/A| < 1. \quad (3.28)$$

This completes the proof.

These two propositions yield the following important corollary:

Corollary: If $f_r(\{\sigma_{ri}\}) \subset I$ for all r, then any product

$$f_1(\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{14}, \cdots) f_2(\sigma_{21}, \sigma_{22}, \sigma_{23}, \cdots) f_3(\sigma_{31}, \sigma_{32}, \sigma_{33}, \cdots) \cdots,$$

where variables joined by a line are contracted, will belong to I provided that all functions are *linked* by the contractions.

We are now in a position to complete the proof of Theorem 1. First, we note that we may assume that all the partial Boltzmann factors $q_{v,r}(\sigma_1, \cdots, \sigma_v)$ entering the expression (2.3) for the total partition function $Z_N(z)$ are linked. If the circle theorem is proved for one (linked) factor, it clearly holds for the total partition function. It thus follows from the corollary just stated, Definition 2, and Lemma 3 that all zeros of $Z_N(z)$ will lie on $|z| = 1$ provided that the $q_{v,r}(\sigma_{(1)}, \cdots, \sigma_{(v)})$ belong to I for all v and r. Our task is thus to prove this under conditions (A) and (B) of the theorem [see (2.1) and (2.2)].

Our first step (see also Ref. 1) is to prove by using (A) that

(M) any root $z_1 = \zeta_1$ of the equation

$$\hat{q} = Q(z_1, \cdots, z_v) = A_1 z_1 + B_1 z_1^{-1} = 0, \quad (3.29)$$

with

$$|z_i| = 1, \quad \text{for} \quad i \geq 2, \quad \text{and} \quad A_1(z_2, \cdots, z_n) \neq 0, \quad (3.30)$$

satisfies $|\zeta_1| = 1$.

We write q for a typical factor $q_{v,r}$. If A_1 does not vanish in (3.29), we have

$$|\zeta_1|^2 = |B_1|/|A_1|. \quad (3.31)$$

As previously, we may express A_1 as

$$A_1 = \tfrac{1}{2} \operatorname{Tr} \left\{ \prod_{i \geq 2} z_i^{\sigma_i} q(1, \sigma_2, \cdots, \sigma_v) \right\}, \quad (3.32)$$

and, since the σ_i are dummy variables under the trace, we have

$$A_1 = \tfrac{1}{2} \operatorname{Tr} \left\{ \prod_{i \geq 2} z_i^{-\sigma_i} q(1, -\sigma_2, \cdots, -\sigma_v) \right\}. \quad (3.33)$$

Then if q is spin-inversion invariant, as required by (A) [see (2.1)], and if (3.30) is satisfied so that $z_i^{-1} = z_i^*$, we have

$$A_1 = \tfrac{1}{2} \operatorname{Tr} \left\{ \prod_{i \geq 2} (z_i^*)^{\sigma_i} q^*(-1, \sigma_2, \cdots, \sigma_v) \right\} = B_1^*. \quad (3.34)$$

It follows directly from (3.32) that $|\zeta_1| = 1$, as asserted in (M).

Next we consider condition (B) of the theorem as a *strict* inequality. By using the spin inversion symmetry (A) and removing the term $|q(1, 1, \cdots, 1)|$ from the right-hand side, the condition becomes

$$(\mathrm{B})_{\neq} \quad |q(1, 1, \cdots, 1)| > \sum_{\sigma_2 = \pm 1}{}' \cdots \sum_{\sigma_v = \pm 1}{}' |q(1, \sigma_2, \cdots, \sigma_v)|, \quad (3.35)$$

where the primes denote the exclusion of the term $\sigma_2 = \sigma_3 = \cdots = \sigma_n = +1$. Now from (3.32) we find

$$A_1(z_2, \cdots, z_v) = 2^{-v} z_2 z_3 \cdots z_v \Big(q(1, 1, \cdots, 1) + \sum_{\sigma_2 = \pm 1}{}' \cdots \sum_{\sigma_v = \pm 1}{}' \prod_{j \geq 2} z_j^{\sigma_j - 1} q(1, \sigma_2, \cdots, \sigma_v) \Big). \quad (3.36)$$

If we notice that $\sigma_j - 1 \leq 0$, it is clear that

(N) under condition $(\mathrm{B})_{\neq}$ the coefficient A_1 of z_1 in the expansion of $Q(z_1, \cdots, z_v)$ cannot vanish if $|z_i| \geq 1$ for all $i \geq 2$.

Finally, we will prove that this nonvanishing of A_1 implies that q belongs to I as required. To achieve this, we will suppose the contrary and obtain a contradiction to statement (M) above. Now if q does not belong to I, we conclude, from the characterization (3.5) given before, that there exists a solution of

$$Q(z_1, z_2, \cdots, z_v) = 0, \quad (3.37)$$

for a set of values $z_i = z_i^{(0)}$ such that, for μ (≥ 1) of the indices i, we have $|z_i^{(0)}| > 1$ while for the remaining $(v - \mu)$ indices we have $|z_i^{(0)}| = 1$. If $\mu = 1$, we have an immediate contradiction to (M). Without loss of generality, we may, in fact, suppose that

$$|z_1^{(0)}| > 1, |z_2^{(0)}| > 1, \cdots, |z_\mu^{(0)}| > 1 \quad \text{and} \quad |z_i^{(0)}| = 1 \quad \text{for} \quad i > \mu. \quad (3.38)$$

We will prove, following the technique of Lee and Yang, that there is another solution, $z_i = z_i^{(1)}$, of (2.37) with $A_1 \neq 0$ such that

$$|z_1^{(1)}| > 1, |z_2^{(1)}| > 1, \cdots, |z_{\mu-1}^{(1)}| > 1 \quad \text{and} \quad |z_i^{(1)}| = 1 \quad \text{for} \quad i > \mu - 1. \quad (3.39)$$

To see this, hold fixed

$$z_2^{(0)} = z_2^{(1)}, \cdots, z_{\mu-1}^{(0)} = z_{\mu-1}^{(1)}$$
$$\text{and} \quad z_i^{(0)} = z_i^{(1)} \quad \text{for} \quad i > \mu, \qquad (3.40)$$

and consider a root $z_\mu = \zeta_\mu$ of (3.37) as a function of z_1. When $z_1 = z_1^{(0)}$, we have $\zeta_\mu = z_\mu^{(0)}$ with $|z_\mu^{(0)}| > 1$. When $z_1 \to \infty$, the root ζ_μ must approach continuously a root ζ_μ' of $A_1(z_\mu) = 0$. But we conclude from (N) above that any such root satisfies $|\zeta_\mu'| < 1$. By continuity it follows that there is some finite value $z_1 = z_1^{(1)}$ such that $\zeta_\mu = z_\mu^{(1)}$ has modulus unity and that $A_1(\zeta_\mu) \neq 0$. This establishes the new root set (3.39). By repeating the process, we ultimately obtain a root set of (3.37) such that $|z_1^{(\lambda)}| > 1$, $|z_i^{(\lambda)}| = 1$ (for all $i \geq 2$), and $A_1(\{z_i^{(\lambda)}\}) \neq 0$; but this contradicts the statement (M).

This completes the proof of Theorem 1 under condition $(B)_{\neq}$, that is, condition (B) with a strict inequality. However, since the roots of $Z_N(z)$ are continuous functions of the $q_{\nu,r}(\sigma_1, \cdots, \sigma_\nu)$, we may relax $(B)_{\neq}$ to (B).

ACKNOWLEDGMENTS

We are indebted to Dr. T. Asano, Dr. J. Ginibre, Dr. G. Gallavotti, and Professor E. H. Lieb for sending us preprints of their work. The support of the National Science Foundation is gratefully acknowledged.

* On leave of absence from the Institute for Solid State Physics, University of Tokyo, Tokyo.
[1] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[2] T. Asano, Progr. Theoret. Phys. (Kyoto) **40**, 1328 (1968); J. Phys. Soc. Japan **25**, 1220 (1968).
[3] M. Suzuki, J. Math. Phys. **9**, 2064 (1968); Progr. Theoret. Phys. (Kyoto) **40**, 1246 (1968).
[4] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).
[5] O. J. Heilmann and E. H. Lieb, Phys. Rev. Letters **24**, 1412 (1970).
[6] P. W. Kasteleyn, in *Graph Theory and Theoretical Physics*, F. Harary, Ed. (Academic, New York, 1967), p. 43; M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).
[7] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **41**, 1438 (1969).
[8] C. Kawabata and M. Suzuki, J. Phys. Soc. Japan **28**, 16 (1970).
[9] T. Asano, J. Phys. Soc. Japan **29**, 350 (1970) and Phys. Rev. Letters **24**, 1409 (1970).
[10] Our definitions differ somewhat from Asano's.
[11] J. Ginibre, in lectures given at the Cargèse Summer School in Statistical Mechanics, July, 1969.
[12] H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).
[13] See, e.g., L. V. Ahlfors, *Complex Analysis* (McGraw-Hill, New York, 1966), 2nd ed., p. 176; E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., Oxford, 1939), 2nd ed., p. 119.
[14] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967); Commun. Math. Phys. **6**, 121 (1967); D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968); J. Ginibre, Phys. Rev. Letters **23**, 828 (1969).
[15] G. Gallavotti, The Rockefeller University, Preprint, 1970; see also Ref. 11.
[16] J. C. Slater, J. Chem. Phys. **9**, 16 (1941); H. Takahashi, Proc. Phys. Math. Soc. (Japan) **23**, 1069 (1941).
[17] J. F. Nagle, J. Math. Phys. **7**, 1484, 1492 (1966).
[18] E. H. Lieb, Phys. Rev. Letters **18**, 692, 1046; **19**, 108 (1967); B. Sutherland, Phys. Rev. Letters **19**, 103 (1967).
[19] See, e.g., F. Y. Wu, Phys. Rev. Letters **22**, 1174 (1969); **24**, 1476 (1970).
[20] S. Katsura, Y. Abe, and K. Ohkouchi, J. Phys. Soc. Japan **29**, 845 (1970).
[21] B. Sutherland, J. Math. Phys. **11**, 3183 (1970).
[22] H. A. Kramers and G. H. Wannier, Phys. Rev. **66**, 252 (1941).
[23] See, e.g., M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).
[24] T. Utiyama, Progr. Theoret. Phys. (Kyoto) **6**, 907 (1951); see also C. Domb, Advan. Phys. **34**, 150 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Quantum Many-Body Problem in One Dimension: Ground State*

BILL SUTHERLAND
Physics Department, University of California, Berkeley, California 94720

(Received 12 April 1970)

We investigate the ground state of a system of either fermions or bosons interacting in one dimension by a 2-body potential $V(r) = g/r^2$. In the thermodynamic limit, we determine the ground state energy and pair correlation function.

INTRODUCTION

We wish to investigate the properties of a 1-dimensional N-body system interacting by the 2-body potential $V(r) = g/r^2$. In particular, we shall be concerned in this paper with the ground state in the thermodynamic limit: $N \to \infty$ and volume $L \to \infty$, with the density $d = N/L$ kept finite. The discussion shall be divided into sections as follows: Section I discusses the history and peculiarities of the g/r^2 potential; Sec. II derives the ground state wavefunction for an N-body system contained in a weak harmonic well; Sec. III identifies the square of this wavefunction as identical with a probability distribution function familiar in the theory of random matrices, enabling many properties to be determined immediately by correspondence.

To see this, hold fixed

$$z_2^{(0)} = z_2^{(1)}, \cdots, z_{\mu-1}^{(0)} = z_{\mu-1}^{(1)}$$
$$\text{and} \quad z_i^{(0)} = z_i^{(1)} \quad \text{for} \quad i > \mu, \qquad (3.40)$$

and consider a root $z_\mu = \zeta_\mu$ of (3.37) as a function of z_1. When $z_1 = z_1^{(0)}$, we have $\zeta_\mu = z_\mu^{(0)}$ with $|z_\mu^{(0)}| > 1$. When $z_1 \to \infty$, the root ζ_μ must approach continuously a root ζ_μ' of $A_1(z_\mu) = 0$. But we conclude from (N) above that any such root satisfies $|\zeta_\mu'| < 1$. By continuity it follows that there is some finite value $z_1 = z_1^{(1)}$ such that $\zeta_\mu = z_\mu^{(1)}$ has modulus unity and that $A_1(\zeta_\mu) \neq 0$. This establishes the new root set (3.39). By repeating the process, we ultimately obtain a root set of (3.37) such that $|z_1^{(\lambda)}| > 1$, $|z_i^{(\lambda)}| = 1$ (for all $i \geq 2$), and $A_1(\{z_i^{(\lambda)}\}) \neq 0$; but this contradicts the statement (M).

This completes the proof of Theorem 1 under condition $(B)_{\neq}$, that is, condition (B) with a strict inequality. However, since the roots of $Z_N(z)$ are continuous functions of the $q_{\nu,r}(\sigma_1, \cdots, \sigma_\nu)$, we may relax $(B)_{\neq}$ to (B).

ACKNOWLEDGMENTS

We are indebted to Dr. T. Asano, Dr. J. Ginibre, Dr. G. Gallavotti, and Professor E. H. Lieb for sending us preprints of their work. The support of the National Science Foundation is gratefully acknowledged.

* On leave of absence from the Institute for Solid State Physics, University of Tokyo, Tokyo.
[1] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[2] T. Asano, Progr. Theoret. Phys. (Kyoto) **40**, 1328 (1968); J. Phys. Soc. Japan **25**, 1220 (1968).
[3] M. Suzuki, J. Math. Phys. **9**, 2064 (1968); Progr. Theoret. Phys. (Kyoto) **40**, 1246 (1968).
[4] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).
[5] O. J. Heilmann and E. H. Lieb, Phys. Rev. Letters **24**, 1412 (1970).
[6] P. W. Kasteleyn, in *Graph Theory and Theoretical Physics*, F. Harary, Ed. (Academic, New York, 1967), p. 43; M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).
[7] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **41**, 1438 (1969).
[8] C. Kawabata and M. Suzuki, J. Phys. Soc. Japan **28**, 16 (1970).
[9] T. Asano, J. Phys. Soc. Japan **29**, 350 (1970) and Phys. Rev. Letters **24**, 1409 (1970).
[10] Our definitions differ somewhat from Asano's.
[11] J. Ginibre, in lectures given at the Cargèse Summer School in Statistical Mechanics, July, 1969.
[12] H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).
[13] See, e.g., L. V. Ahlfors, *Complex Analysis* (McGraw-Hill, New York, 1966), 2nd ed., p. 176; E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., Oxford, 1939), 2nd ed., p. 119.
[14] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967); Commun. Math. Phys. **6**, 121 (1967); D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968); J. Ginibre, Phys. Rev. Letters **23**, 828 (1969).
[15] G. Gallavotti, The Rockefeller University, Preprint, 1970; see also Ref. 11.
[16] J. C. Slater, J. Chem. Phys. **9**, 16 (1941); H. Takahashi, Proc. Phys. Math. Soc. (Japan) **23**, 1069 (1941).
[17] J. F. Nagle, J. Math. Phys. **7**, 1484, 1492 (1966).
[18] E. H. Lieb, Phys. Rev. Letters **18**, 692, 1046; **19**, 108 (1967); B. Sutherland, Phys. Rev. Letters **19**, 103 (1967).
[19] See, e.g., F. Y. Wu, Phys. Rev. Letters **22**, 1174 (1969); **24**, 1476 (1970).
[20] S. Katsura, Y. Abe, and K. Ohkouchi, J. Phys. Soc. Japan **29**, 845 (1970).
[21] B. Sutherland, J. Math. Phys. **11**, 3183 (1970).
[22] H. A. Kramers and G. H. Wannier, Phys. Rev. **66**, 252 (1941).
[23] See, e.g., M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).
[24] T. Utiyama, Progr. Theoret. Phys. (Kyoto) **6**, 907 (1951); see also C. Domb, Advan. Phys. **34**, 150 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Quantum Many-Body Problem in One Dimension: Ground State*

BILL SUTHERLAND
Physics Department, University of California, Berkeley, California 94720

(Received 12 April 1970)

We investigate the ground state of a system of either fermions or bosons interacting in one dimension by a 2-body potential $V(r) = g/r^2$. In the thermodynamic limit, we determine the ground state energy and pair correlation function.

INTRODUCTION

We wish to investigate the properties of a 1-dimensional N-body system interacting by the 2-body potential $V(r) = g/r^2$. In particular, we shall be concerned in this paper with the ground state in the thermodynamic limit: $N \to \infty$ and volume $L \to \infty$, with the density $d = N/L$ kept finite. The discussion shall be divided into sections as follows: Section I discusses the history and peculiarities of the g/r^2 potential; Sec. II derives the ground state wavefunction for an N-body system contained in a weak harmonic well; Sec. III identifies the square of this wavefunction as identical with a probability distribution function familiar in the theory of random matrices, enabling many properties to be determined immediately by correspondence.

I. HISTORY AND PECULIARITIES OF THE g/r^2 POTENTIAL

Several recent papers have been concerned with systems of particles interacting in one dimension by a two-body potential $V(r) = g/r^2$. One line of investigation proceeds from Dyson's work[1] on phase transitions of classical systems in one dimension, which indicates that systems with attractive potentials falling off slower than $1/r^2$ have phase transitions, while attractive potentials falling off faster than $1/r^2$ do not. Thus one is particularly interested in whether an attractive $1/r^2$ potential has a phase transition. As this paper deals only with the zero temperature problem, it throws no light on the occurrence of phase transitions; finite temperature properties are to be discussed in a later paper. Dyson's work is recalled to emphasize that g/r^2 in one dimension is a particularly interesting choice.

Second, Calogero,[2] treating the corresponding quantum system, solved exactly the 3-body problem and N-body ground state problem in both free space and with mutual harmonic interactions. Calogero's work does not allow one to determine the behavior of the g/r^2 system in the thermodynamic limit. However, our ground state wavefunction is very similar to that of Calogero; the difference in this investigation and Calogero's is best described as one of viewpoint. It is this shift in viewpoint which allows us to proceed further than Calogero, to discuss the particle density, pair correlation function, and ground state energy at finite density.

The potential g/r^2 is very singular at the origin and requires some care to make physical sense. Classically, the attractive potential is too strong at the origin, requiring, for instance, a finite hard core to prevent "fall to the origin." However, for the quantum system, the zero-point motion acts to keep the particle from the origin, so that no cutoff is required, provided that the potential is not too attractive. Thus, for the 2-body Hamiltonian

$$H = -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} + \frac{g}{(x-y)^2}, \tag{1}$$

we are led to the restriction $g \geq -\frac{1}{4}$. This point is discussed by Landau and Lifshitz[3] and is reproduced as an appendix at the end of this paper. Within this range of g, there are no bound states. The unnormalized solutions for the 2-body Hamiltonian of Eq. (1) are

$$\begin{aligned} \psi_\pm &= e^{iKR} r^{\frac{1}{2}} J_{\pm a}(kr) \\ &\sim e^{iKR} r^{\frac{1}{2}\pm a}, \quad r \to 0, \\ &\sim e^{iKR} \cos(kr - \tfrac{1}{4}\pi \mp \tfrac{1}{2}\pi a), \quad r \to \infty, \\ E &= \tfrac{1}{2}(k^2 + K^2) \end{aligned} \tag{2}$$

with

$$\begin{aligned} a &= \tfrac{1}{2}(1+2g)^{\frac{1}{2}} \geq 0, \\ x &= R - r, \quad k \geq 0, \\ y &= R + r, \quad r \geq 0. \end{aligned}$$

$J_a(x)$ is a Bessel function. From the considerations of Landau and Lifshitz and the Appendix, we select the unique solution $\psi \equiv \psi_+$ corresponding to the upper sign in Eq. (2). The wavefunction for $r \leq 0$ is given by

$$\psi(-|r|) = \pm\psi(|r|), \tag{3}$$

corresponding to bosons or fermions, respectively. Defining a new parameter $\lambda = \frac{1}{2} + a = \frac{1}{2}[1 + (1+2g)^{\frac{1}{2}}]$, we see that the "physically reasonable" solutions are $\lambda \geq \frac{1}{2}$. However, as explained in the Appendix, one can also treat free bosons by including the "unreasonable" solutions $\frac{1}{2} > \lambda \geq 0$, corresponding to the lower sign in Eq. (2). With this understanding, in the expressions which follow, we may take λ to range from 0 to $+\infty$.

II. THE GROUND STATE WAVEFUNCTION IN A HARMONIC WELL

We shall now consider the ground state of a system governed by the following Hamiltonian:

$$H = -\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + \prod_{i<j} \frac{g}{(x_i - x_j)^2} + \omega^2 \sum_{i=1}^{N} x_i^2. \tag{4}$$

The last term is an external potential to contain the system, replacing the usual box of volume L. Since we will be concerned with the limit $N \to \infty$, we will let $\omega \to 0$ in order to produce a finite density of the system in the interior.

We write the wavefunction ψ as $\psi = \phi\varphi$. We verify that the choice

$$\phi = \prod_{i<j} |x_i - x_j|^\lambda, \tag{5}$$

$$\varphi = \exp\left(-\tfrac{1}{2}\omega \sum_{i=1}^{N} x_i^2\right), \tag{6}$$

$$\lambda^2 - \lambda = \tfrac{1}{2}g \quad \text{or} \quad \lambda = \tfrac{1}{2} + a, \tag{7}$$

is a solution, with

$$E = \omega N[1 + \lambda(N-1)]. \tag{8}$$

Since, for each ordering of particles, ψ is nodeless, it is the ground state. This is very much like Calogero's ground state wavefunction; however, our Hamiltonian contains an external harmonic well instead of mutual harmonic interactions.

First, let the kinetic term of the Hamiltonian act

on ψ:

$$-\sum_i \frac{\partial^2}{\partial x_i^2}(\phi\varphi) = -2\phi\varphi(\lambda^2-\lambda)\sum_{i<j}\frac{1}{(x_i-x_j)^2} - 2\lambda\phi\sum_i\left(\sum_{j(\neq i)}\frac{1}{x_i-x_j}\right)\frac{\partial\varphi}{\partial x_i} - \phi\sum_i\frac{\partial^2\varphi}{\partial x_i^2}. \quad (9)$$

Thus, choosing λ according to Eq. (7), we eliminate the mutual interaction term. We may rewrite the second term of Eq. (9) to give

$$-\sum_i\frac{\partial^2\varphi}{\partial x_i^2} - 2\lambda\sum_{i<j}\frac{1}{x_i-x_j}\left(\frac{\partial}{\partial x_i}-\frac{\partial}{\partial x_j}\right)\varphi + \omega^2\left(\sum_i x_i^2\right)\varphi = E\varphi. \quad (10)$$

Now

$$\frac{\partial\varphi}{\partial x_i} = -\omega x_i\varphi, \quad \frac{\partial^2\varphi}{\partial x_i^2} = [\omega^2 x_i^2 - \omega]\varphi.$$

Substituting into Eq. (10), we verify the eigenvalue equation with E as in Eq. (8).

Let us rewrite ψ in terms of the variables

$$y_i = (\omega/\lambda)^{\frac{1}{2}}x_i; \quad (11)$$

then

$$\psi = C^{\frac{1}{2}}\exp\left(-\tfrac{1}{2}\lambda\sum_i y_i^2\right)\prod_{i<j}|y_i-y_j|^{\lambda}, \quad (12)$$

$$\psi^2 = C\exp\left(-\tfrac{1}{2}\beta\sum_i y_i^2\right)\prod_{i<j}|y_i-y_j|^{\beta}, \quad (13)$$

with

$$\beta = 2\lambda = 1 + (1+2g)^{\frac{1}{2}}. \quad (14)$$

C is a normalization constant, so that

$$C^{-1} = \int_{-\infty}^{+\infty}\cdots\int dy_1\cdots dy_N\psi^2.$$

III. INTERPRETATION OF ψ^2

It is at this point that we recognize the expression in Eq. (13) to be identical to the joint probability density function for the eigenvalues of matrices from a Gaussian ensemble. Choosing β equal to 1, 2, and 4 corresponds to orthogonal, unitary, and symplectic ensembles, respectively. In our case, the most attractive situation, $g = -\frac{1}{2}$, corresponds to $\beta = 1$; free fermions corresponds to $\beta = 2$.

We may immediately go to the literature[4,5] and find the normalization constant C and the 1- and 2-particle correlation functions.

First, it is conjectured[6] that the normalization constant C is given by

$$C^{-1} = (2\pi)^{\frac{1}{2}N}\beta^{-\frac{1}{2}N-\frac{1}{4}\beta N(N-1)} \times [\Gamma(1+\tfrac{1}{2}\beta)]^{-N}\prod_{j=1}^{N}\Gamma(1+\tfrac{1}{2}\beta_j). \quad (15)$$

This conjecture is verified for $\beta = 1, 2, 4$.

The particle density is given by

$$\sigma(y) = N\int_{-\infty}^{+\infty}\cdots\int\psi^2\,dy_2\cdots dy_N, \quad (16)$$

normalized so that

$$\int_{-\infty}^{+\infty}\sigma(y)\,dy = N. \quad (17)$$

One finds[7]

$$\sigma(y) = \begin{cases}\pi^{-1}(2N-y^2)^{\frac{1}{2}}, & y^2 < 2N,\\ 0, & y^2 > 2N.\end{cases} \quad (18)$$

This is true for all β's. The density of x's, normalized in the same way, is

$$d(x) = \begin{cases}\frac{1}{\pi}\left(\frac{2N\omega}{\lambda}-\frac{\omega^2x^2}{\lambda^2}\right)^{\frac{1}{2}}, & x^2 < \frac{2N\lambda}{\omega},\\ 0, & x^2 > \frac{2N\lambda}{\omega}.\end{cases} \quad (19)$$

We plot $d(x)$ in Fig. 1, comparing it with the corresponding density for the δ-function interaction problem of Lieb and Liniger.[8] Denoting the density at $x = 0$ by d, we find

$$d = \pi^{-1}(2N\omega/\lambda)^{\frac{1}{2}} \quad (20)$$

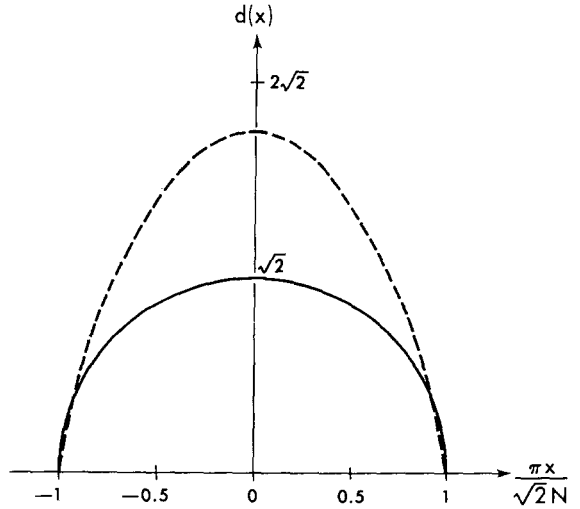

FIG. 1. Particle density $d(x)$, for $\lambda = \frac{1}{2}$ in a well $\omega = \pi^2/2N$, is shown by the solid curve. For comparison, the density of a boson system, interacting by $V(r) = 2\delta(r)$ in one dimension at the same chemical potential in the same well, is shown by the dashed curve.

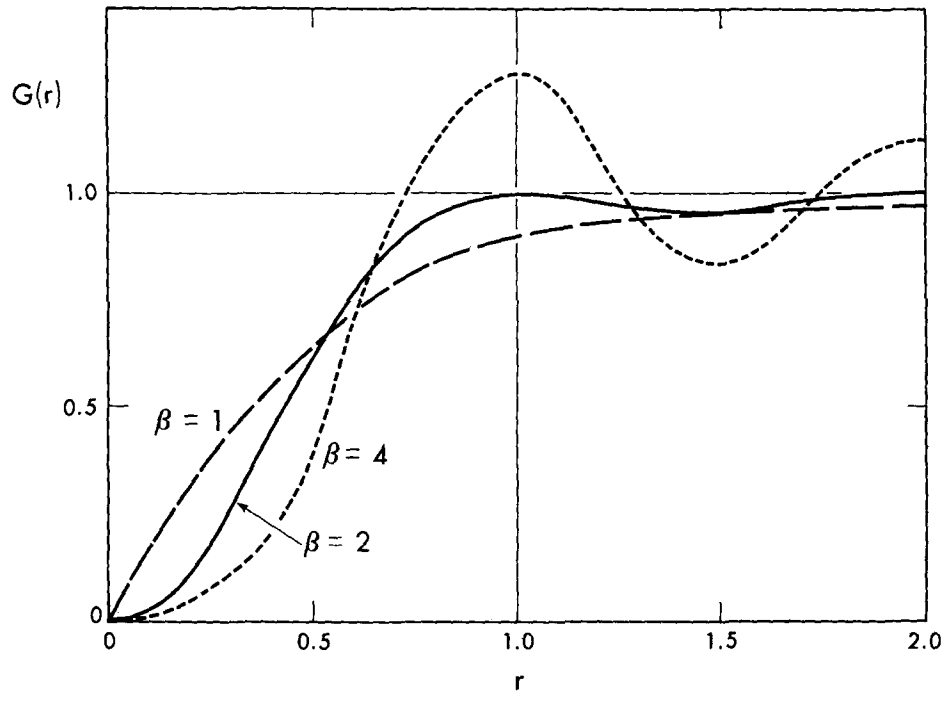

FIG. 2. The pair correlation function is shown for the three values $\beta = 1, 2, 4$; we have taken $d = 1$.

or

$$\omega = \pi^2 d^2 \lambda / 2N. \tag{20'}$$

Thus, we see that, for fixed d, letting $N \to \infty$ means that $\omega \to 0$ as N^{-1}.

We may extract from Eq. (19) the ground state energy for a system at density d in the more conventional box of volume L, instead of a harmonic well. The energy density must be intensive and have dimensions L^{-3}; thus the only choice is

$$E/L = d^3 e(\lambda), \tag{21}$$

where $e(\lambda)$ is a dimensionless function of the coupling constant. The chemical potential is then

$$\mu = 3d^2 e(\lambda). \tag{22}$$

Placing this system in a harmonic well, the chemical potential becomes

$$\mu = 3d^2(x)e(\lambda) + \omega^2 x^2$$

or

$$d(x) = \begin{cases} [(\mu - \omega^2 x^2)/3e(\lambda)]^{\frac{1}{2}}, & x^2 < \mu/\omega^2, \\ 0, & x^2 > \mu/\omega^2. \end{cases} \tag{23}$$

Comparing Eq. (23) with the previous equation (19), we find

$$e(\lambda) = \tfrac{1}{3}\lambda^2\pi^2,$$
$$E/L = \tfrac{1}{3}d^3\lambda^2\pi^2. \tag{24}$$

Likewise, we may determine the thermodynamic properties at zero temperature. Considering pressure P as a function of temperature T, fugacity Z, and coupling constant λ, we find that P must be of the form

$$P = T^{\frac{3}{2}} p(Z, \lambda). \tag{25}$$

Thus the energy density is given by

$$\frac{E}{L} = \frac{T^2 \partial}{\partial T}\left(\frac{P}{T}\right)$$
$$= \tfrac{1}{2}T^{\frac{3}{2}} p(Z, \lambda) = \tfrac{1}{2}P. \tag{26}$$

The equation of state at $T = 0$ is therefore found to be

$$P = \tfrac{1}{3}2d^3\lambda^2\pi^2. \tag{27}$$

We now consider the pair correlation function $G(r)$. Since the density factors out trivially, we take $d = 1$ in the following equations. Then $G(r) \to 1$ as $r \to \infty$, and we write

$$G(r) = 1 - Y(r) \tag{28}$$

and

$$b(k) = \int_{-\infty}^{+\infty} dr Y(r) e^{2\pi i k r}. \tag{29}$$

Then we find the following:

(a) For $\beta = 2$,

$$Y(r) = [s(r)]^2 \equiv [(\sin \pi r)/\pi r]^2, \tag{30}$$

$$b(k) = \begin{cases} 1 - |k|, & |k| \leq 1, \\ 0, & |k| \geq 1. \end{cases} \tag{31}$$

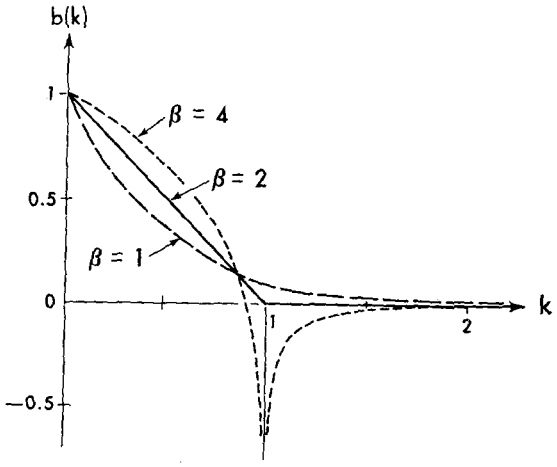

FIG. 3. The negative Fourier transform of the pair correlation function is shown for the same three values of β; again $d = 1$.

(b) For[9] $\beta = 1$,

$$Y(r) = \left(\int_r^\infty s(Z)\,dZ\right)\left(\frac{ds(r)}{dr}\right) + [s(r)]^2, \quad (32)$$

$$b(k) = \begin{cases} 1 - 2|k| + |k| \ln(1 + 2|k|), & |k| \leq 1, \\ -1 + |k| \ln\left(\dfrac{2|k| + 1}{2|k| - 1}\right), & |k| \geq 1. \end{cases} \quad (33)$$

(c) For $\beta = 4$, it is conjectured[10] that

$$Y(r) = [s(2r)]^2 - \left(\int_0^{2r} s(Z)\,dZ\right)\left(\frac{1}{2}\frac{ds(2r)}{dr}\right), \quad (34)$$

$$b(k) = \begin{cases} 1 - \frac{1}{2}|k| + \frac{1}{4}|k| \ln[|(|k| - 1)|], & |k| \leq 2, \\ 0, & |k| \geq 2. \end{cases} \quad (35)$$

These are the only cases for which the pair correlation function can be evaluated. We notice the system becomes more nearly ordered on a lattice as β increases or, equivalently, as the interaction becomes more repulsive. Figures 2 and 3 show $G(r)$ and $b(k)$ for these three values of β.

APPENDIX

We wish to cut off the g/r^2 potential for $r < b$, replacing it by a less singular potential. Then, letting $b \to 0$, we hope for a unique limiting solution ψ. Two unbiased choices for the potential V when $r < b$ are:

(i) $V = g/b^2$, $r < b$—flattening out the singularity;

(ii) $V = +\infty$, $r < b$—a hard core.

The solution for $r \geq b$ will be

$$\psi = \psi_+ + A\psi_- \quad (A1)$$

$\psi_\pm$ are defined in Eq. (2). We take kb to be small, and assume $V = c^2/b^2$, $r < b$, with c real or imaginary. This includes both cases (i) and (ii).

We match logarithmic derivatives at b, obtaining the equation

$$T = \frac{(\frac{1}{2} + a)b^a k^a + A(\frac{1}{2} - a)b^{-a}k^{-a}}{k^a b^a + Ab^{-a}k^{-a}} \quad (A2)$$

with

$$T = \begin{cases} c \tanh(c), & \text{bosons}, \\ c \coth(c), & \text{fermions}. \end{cases} \quad (A3)$$

T is simply a number, equal to $(g^{\frac{1}{2}} \tanh(g^{\frac{1}{2}}), g^{\frac{1}{2}} \coth(g^{\frac{1}{2}}))$ for cutoff (i) and $(+\infty, +\infty)$ for cutoff (ii). Solving for A, we obtain

$$A = -\frac{\frac{1}{2} + a - T}{\frac{1}{2} - a - T} b^{2a}k^{2a}. \quad (A4)$$

Thus if $g \geq \frac{1}{2}$, $a > 0$ real, then, as $b \to 0$, $A \to 0$, with the exception of the case

$$T = \tfrac{1}{2} - a. \quad (A5)$$

This is a resonance condition and is not fulfilled for cutoffs (i) and (ii).

For $g < \frac{1}{2}$, a is pure imaginary, and A does not approach a limit. Upon closer examination, one sees that there is no lowest energy eigenstate for the Hamiltonian in this case.

There is some precedent[11] for using the solutions ψ_-, although as we have seen, this is surely artificial. For instance, one wishes to have two independent solutions in the scattering of nonidentical particles by a g/r^2 potential. We take the viewpoint that the mathematical equations are easily extended to the "unphysical" ψ-solutions with no extra effort; thus these cases are included in this paper.

* Work supported in part by National Science Foundation Grant GP 13632.

[1] F. J. Dyson, Commun. Math. Phys. **12**, 91 (1969); **12**, 212 (1969).

[2] F. Calogero, J. Math. Phys. **10**, 2191, 2197 (1969).

[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958).

[4] C. Porter, *Statistical Theories of Spectra* (Academic, New York, 1965). This reference reprints the original papers of the field.

[5] M. L. Mehta, *Random Matrices* (Academic, New York, 1967). All mathematical techniques used in the present paper are presented in this reference.

[6] F. J. Dyson, J. Math. Phys. **4**, 713 (1963).

[7] E. P. Wigner, Ann. Math. **62**, 548 (1955).

[8] E. Lieb and W. Liniger, Phys. Rev. **130**, 1605 (1963).

[9] M. L. Mehta, Nucl. Phys. **18**, 395 (1960); M. L. Mehta and M. Gaudin, *ibid.*, 420 (1960); M. Gaudin, *ibid.* **25**, 447 (1960); F. J. Dyson, J. Math. Phys. **3**, 166 (1962).

[10] This is identical to the theorem for circular ensembles of F. J. Dyson, J. Math. Phys. **4**, 714 (1963).

[11] F. L. Scarf, Phys. Rev. **112**, 1137 (1958).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Quantum Many-Body Problem in One Dimension: Thermodynamics*

BILL SUTHERLAND
Physics Department, University of California, Berkeley, California 94720

(Received 21 June 1970)

We continue our investigation of a system of either fermions or bosons interacting in one dimension by a 2-body potential $V(r) = g/r^2$. We first present an approximation for the eigenstates of a general 1-dimensional quantum many-body system. We then apply this approximation to the g/r^2 potential, allowing complete determination of the thermodynamic properties. Finally, comparing the results with those properties known exactly, we conjecture that the approximation is, in fact, exact for the g/r^2 potential.

INTRODUCTION

This paper continues an investigation of the properties of a 1-dimensional quantum N-body system interacting by the 2-body potential $V(r) = g/r^2$. As before, we shall be concerned with the thermodynamic limit: $N \to \infty$ and volume $L \to \infty$, with density $d = N/L$ kept finite. The first paper[1] treated the ground state exactly; this paper will develop an approximation for the excited states, and hence the thermodynamics. Thus, the first section of the present paper introduces this approximation, while the second section applies it to the g/r^2 potential. Finally, the third section compares the results with those properties known exactly: zero temperature thermodynamics, second and third virial coefficients, and the limit of zero interaction. It is found that in all cases our approximation reproduces the exact results. Thus we are finally led to conjecture that the approximation is, in fact, exact for the g/r^2 potential.

I. APPROXIMATE EIGENVALUES OF A 1-DIMENSIONAL N-BODY SYSTEM

We now present an approximation for the energy eigenvalues and thermodynamics of a system of N fermions or bosons interacting in one dimension. Later we shall apply the method specifically to the g/r^2 potential, but we expect it to apply much more generally in one dimension whenever a potential has a phase shift and no bound states. This restriction to potentials with no bound states results because the approximation uses an asymptotic wavefunction which neglects diffraction effects from the simultaneous interaction of three or more particles. The results will then depend on the 2-body Hamiltonian through the exact 2-body phase shift. The scheme is interesting, for it gives the thermodynamics in an approximate but closed form, in contrast to a systematic series expansion. On the other hand, it is difficult to estimate the error. The approximation is modeled on the exact solution of the 1-dimensional δ-function boson problem.[2,3]

Consider the 2-body problem governed by a Hamiltonian:

$$H = -\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} + V(|x_1 - x_2|). \tag{1}$$

For $x_1 \ll x_2$, the asymptotic wavefunction is

$$\psi(x_1, x_2) = e^{i(k_1x_1+k_2x_2)} - e^{i\theta(k_1-k_2)}e^{i(k_2x_1+k_1x_2)}. \tag{2}$$

We assume that $k_1 > k_2$. Then $S(k) = -\exp[-i\theta(k)]$ is the S matrix and $\theta(k)$ is the 2-body phase shift; $\theta(k)$ is odd in k. For bosons or fermions, respectively, we have

$$\psi(x_1 \gg x_2) = \pm\psi(x_1 \ll x_2).$$

The energy is given by $E = k_1^2 + k_2^2$. In the center of mass coordinates [Eq. (I.2)], the asymptotic wavefunction becomes

$$\psi = (2/i)e^{iKR}\sin[kr - \tfrac{1}{2}\theta(k)]. \tag{3}$$

For later reference, we list $\theta(k)$ for the following examples:

$$\delta \text{ function, } V(r) = 2c\delta(r)\text{: } \theta(k) = -2\tan^{-1}(k/c), \tag{4a}$$

$$\text{hard rod, } v(r) = +\infty,\ r \le b\text{: } \theta(k) = kb, \tag{4b}$$

$$V(r) = g/r^2\text{: } \theta(k) = (\pi k/2|k|)[(1 + 2g)^{\frac{1}{2}} - 1], \tag{4c}$$

$$\text{free fermions: } \theta(k) = 0, \tag{4d}$$

$$\text{free bosons: } \theta(k) = -k\pi/|k|. \tag{4e}$$

Our basic approximation will be to assume that the asymptotic N-body wavefunction is that given by 2-body scattering alone. Therefore, there is no diffraction, and momenta are exchanged in pairs. The asymptotic wavefunction for the region $x_1 \ll x_2 \ll \cdots \ll x_N$ is a sum of $N!$ terms corresponding to the $N!$ permutations P of the N k's, $k_1 > k_2 > \cdots > k_N$:

$$\psi = \sum_P A(P)\exp(i\sum k_{Pj}x_j). \tag{5}$$

The coefficients $A(P)$ are related by 2-body scatterings:

$$A(\cdots k', k \cdots)/A(\cdots k, k' \cdots) = -e^{-i\theta(k-k')}. \tag{6}$$

For other orderings of the particles, we simply use either Fermi or Bose statistics. This wavefunction is of the form known as Bethe's ansatz.[4]

We are now ready to apply periodic boundary conditions to the wavefunction, determining a unique set of k's for each quantum state. The energy is given by

$$E = \sum k_i^2. \tag{7}$$

For a ring of circumference L, we find the following equation for the k's:

$$e^{-ikL} = (-1)^{N-1} \exp\left(i \sum_{k'} \theta(k' - k)\right). \tag{8}$$

Upon taking the logarithm, we obtain

$$kL = 2\pi I(k) + \sum_{k'} \theta(k - k'). \tag{9}$$

The $I(k)$'s are either integers or half odd integers which come from $\log(\pm 1)/2\pi$ and, in fact, serve as quantum numbers for the problem. They may be taken as free fermion k vectors. Thus, for example, the ground state is given by the $I(k)$'s densely packed about zero.

We now adopt Yang and Yang's derivation for the thermodynamics of the δ-function Bose gas,[3] without rederiving. The pressure is given as a function of temperature T and chemical potential μ by

$$P(\mu, T) = \frac{T}{2\pi} \int_{-\infty}^{\infty} dk \ln(1 + e^{-\epsilon(k)/T}). \tag{10}$$

$\epsilon(k)$ depends upon μ and T and satisfies the integral equation

$$\epsilon(k) = -\mu + k^2 + \frac{T}{2\pi} \int_{-\infty}^{\infty} dk' \theta'(k - k') \times \ln(1 + e^{-\epsilon(k')/T}), \tag{11}$$

with $\theta'(k) = d\theta(k)/dk$, $\theta(k)$ being the 2-body phase shift. As usual, the density d is given by

$$d = \frac{\partial P}{\partial \mu}. \tag{12}$$

We now consider simple examples.

(a) δ-function potential: As our approximation is modeled on this problem, substituting the phase shift (4a) into the above equations gives the exact equations of Yang and Yang.

(b) Hard rod of radius b: The phase shift is given by (4b). Equation (9) for the k's is

$$kL(1 - db) = 2\pi I(k) - bK, \tag{13}$$

$$K = \sum k = \frac{2\pi}{L} \sum I(k). \tag{14}$$

Thus the k's for $K = 0$ are like free fermions in a volume reduced by the hard cores.

We have

$$\epsilon(k) = -\mu + k^2 + bP, \tag{15}$$

$$P = \frac{T}{2\pi} \int_{-\infty}^{\infty} dk \ln\left[1 + \exp\left(\frac{\mu - bP - k^2}{T}\right)\right]. \tag{16}$$

If we write $P_0(\mu, T)$ as the pressure for free fermions, then $P(\mu, T) = P_0(\mu - bP, T)$. We may invert to find $\mu(P) = \mu_0(P) + bP$. Differentiating with respect to P gives the specific volume $v \equiv 1/d$,

$$v(P, T) = v_0(P, T) + b. \tag{17}$$

This clearly shows that the effect of the potential is only to create an excluded volume, and the system otherwise behaves as a system of free fermions. These results are exact.

II. APPLICATION OF THE APPROXIMATION TO THE g/r^2 POTENTIAL

We now apply the approximation of the previous section to the g/r^2 potential. Using the phase shift of Eq. (4c), we find the kernel of the integral equation (11) to be

$$\theta'(k) = 2\pi\gamma\delta(k), \tag{18}$$

with the following definitions:

$$a = \tfrac{1}{2}(1 + 2g)^{\frac{1}{2}}, \quad \lambda = a + \tfrac{1}{2}, \quad \gamma = a - \tfrac{1}{2}. \tag{19}$$

Thus the integral equation simply reduces to the transcendental equation

$$\epsilon = -\mu + k^2 + T\gamma \ln(1 + e^{-\epsilon/T}). \tag{20}$$

Before proceeding with the finite T results, let us first investigate the ground state and low lying excitations.

A. Ground State

In the case of the ground state, we have the k's distributed with a density $\rho(k)$,

$$2\pi\rho(k) = \begin{cases} 1 - 2\pi\gamma\rho(k), & |k| < k_0, \\ 0, & |k| > k_0, \end{cases} \tag{21}$$

where k_0 is a Fermi momentum determined by

$$\int_{-k_0}^{k_0} \rho(k)\, dk = d. \tag{22}$$

The ground state energy density $E_0/L \equiv u_0$ is given by

$$u_0 = \int_{-k_0}^{k_0} \rho(k) k^2\, dk. \tag{23}$$

These equations are easily solved to yield

$$\rho(k) = 1/2\pi\lambda, \tag{21'}$$

$$d = k_0/\pi\lambda, \tag{22'}$$

$$u_0 = k_0^3/3\pi\lambda = \tfrac{1}{3}\pi^2\lambda^2 d^3. \tag{23'}$$

Differentiating u_0 with respect to d gives the chemical potential

$$\mu = (\pi\lambda d)^2. \tag{24}$$

We recall that these $T = 0$ results agree with the exact results of Paper I.

B. Excitations near the Ground State

Excitations near the ground state are finite numbers of hole–particle pairs obtained by taking particles from states k_α below the Fermi surface to states k'_α above the Fermi surface. Then the energy and momentum are given by

$$E - E_0 = \sum [\epsilon(k'_\alpha) - \epsilon(k_\alpha)],$$
$$P = \sum [p(k'_\alpha) - p(k_\alpha)]. \tag{25}$$

We easily find

$$\epsilon(k) = \begin{cases} (k^2 - k_0^2)/\lambda + k_0^2, & |k| < k_0, \\ k^2, & |k| > k_0, \end{cases} \tag{26a}$$

$$p(k) = \begin{cases} k/\lambda, & |k| < k_0, \\ k - k_0 + k_0/\lambda, & |k| > k_0. \end{cases} \tag{26b}$$

Defining $p_0 = k_0/\lambda = \pi d$, we find the dispersion curve $\epsilon(p)$ of the excitations to be

$$\epsilon(p) = \begin{cases} \lambda[p^2 + \gamma p_0^2], & |p| < p_0, \\ [p + \gamma p_0]^2, & |p| > p_0. \end{cases} \tag{27}$$

Note that the derivative is continuous across p_0 and equal to $2\lambda p_0 = 2k_0$.

C. Thermodynamics

Let us now return to Eq. (20). In terms of the following quantities,

$$Z = \text{fugacity} = e^{\mu/T},$$
$$\alpha = \exp[(\mu - k^2)/T] = Ze^{-k^2/T}, \tag{28}$$
$$\zeta = e^w = 1 + e^{-\epsilon/T},$$

Eq. (20) becomes

$$\alpha = \zeta^\lambda - \zeta^\gamma \equiv 2e^{aw} \sinh \tfrac{1}{2}w, \tag{29}$$

with pressure given by

$$P = \frac{T}{2\pi}\int_{-\infty}^{\infty} dk \ln \zeta = \frac{T}{2\pi}\int_{-\infty}^{\infty} dk\, w(k). \tag{30}$$

Writing P as a function of Z, T, and a, we note the symmetry $P(T, -Z, -a) = P(T, Z, a)$. Further, P as a function of Z has the following singularities:

(i) $\lambda \geq 1$, a branch cut along the negative real axis beginning at the branch point

$$Z_0 = -\frac{2}{(4a^2 - 1)^{\frac{1}{2}}}\left(\frac{2a - 1}{2a + 1}\right)^a = \frac{1}{(\lambda\gamma)^{\frac{1}{2}}}\left(\frac{\gamma}{\lambda}\right)^a; \tag{31a}$$

(ii) $1 > \lambda > 0$, two branch cuts extending to infinity from the following two branch points:

$$Z_0 = e^{\pm i\pi\lambda}\frac{2}{(1 - 4a^2)^{\frac{1}{2}}}\left(\frac{1 - 2a}{1 + 2a}\right)^a$$
$$= e^{\pm i\pi\lambda}\frac{1}{(-\lambda\gamma)^{\frac{1}{2}}}\left(-\frac{\gamma}{\lambda}\right)^a. \tag{31b}$$

The location of these branch points in the complex Z plane as a function of λ is shown in Fig. 1. $\lambda = 0$ is the case of free bosons. We therefore conclude from the above that there are no singularities on the positive real Z axis for $\lambda > 0$ and hence no phase transitions in these models.

D. Power Series Expansion for P

We wish to find a power series expansion for P in Z of the form

$$\frac{P}{T} = \frac{1}{2}\left(\frac{T}{\pi}\right)^{\frac{1}{2}}\sum_{n=1}^{\infty} Z^n B_n. \tag{32}$$

We first invert Eq. (29) to give

$$w = \ln \zeta = \sum_{n=1}^{\infty} \alpha^n C_n. \tag{33}$$

Then we perform the k integration on Eq. (30) to give

$$B_n = C_n/n^{\frac{1}{2}}, \tag{34}$$

independent of T. Therefore, the energy per volume u is given by the expression $u = \frac{1}{2}P$, as found in Eq. (I.26).

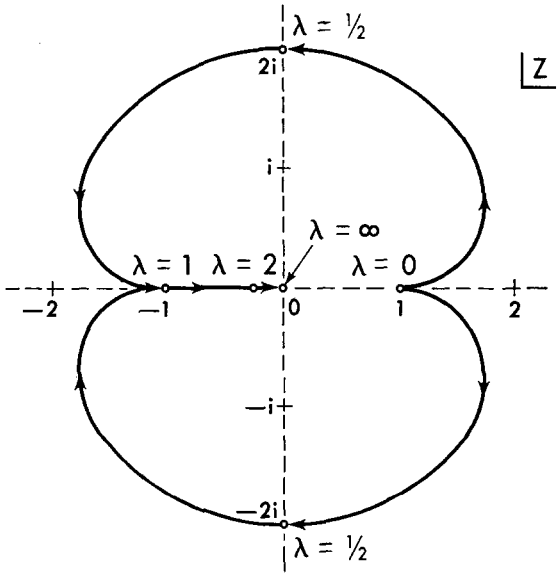

FIG. 1. Solid curve indicates location of branch points of pressure in the complex Z plane as a function of $\lambda \equiv \frac{1}{2}[1 + (1 + 2g)^{\frac{1}{2}}]$.

Upon inverting Eq. (29), we find the coefficients in the expansion of Eq. (33) to be

$$C_n = \frac{(-1)^{n+1}}{n!}(n\gamma+1)(n\gamma+2)\cdots(n\gamma+n-1)$$
$$= \frac{(-1)^{n+1}}{n(\gamma+1)}\frac{[n(\gamma+1)]!}{n!\,(n\gamma)!}. \tag{35}$$

Thus

$$B_n = \frac{(-1)^{n+1}}{n^{\frac{3}{2}}(\gamma+1)}\frac{[n(\gamma+1)]!}{n!\,(n\gamma)!}, \tag{36}$$

and the radius of convergence of the power series of Eq. (32) is

$$R = |\gamma^\gamma/\lambda^\lambda|, \tag{37}$$

in agreement with Eq. (31). For $Z \geq R$, P is given by the analytic continuation of the power series of Eq. (32).

E. Classical Limit

The classical limit exists only for the repulsive case, for which Eq. (32) becomes

$$\frac{P}{T} = -\left(\frac{T}{2\pi g}\right)^{\frac{1}{2}}\sum_{n=1}^{\infty}\left(\frac{n^{n-1}}{n!\,n^{\frac{1}{2}}}\right)[-Z(\tfrac{1}{2}g)^{\frac{1}{2}}]^n. \tag{38}$$

This expression as a function of Z has a branch cut along the negative real axis beginning at

$$Z_0 = -(1/e)(2/g)^{\frac{1}{2}}. \tag{39}$$

F. Special Cases

We may now consider as examples the special cases investigated in Paper I, with $\beta = 1, 2, 4$. These may all be expressed in closed form.

(a) $\beta = 2, g = 0, \gamma = 0$: $\zeta = 1 + \alpha$,

$$\frac{P}{T} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk \ln[1 + Ze^{-k^2/T}]. \tag{40}$$

This is the familiar expression for free fermions in one dimension.

(b) $\beta = 1, g = -\frac{1}{2}, \gamma = -\frac{1}{2}$:

$$\ln\zeta = w = 2\sinh^{-1}\tfrac{1}{2}\alpha,$$
$$\frac{P}{T} = \pi^{-1}\int_{-\infty}^{\infty} dk \sinh^{-1}(\tfrac{1}{2}Ze^{-k^2/T}), \tag{41}$$
$$d = \pi^{-1}\int_{-\infty}^{\infty} dk(1 + 4Z^{-2}e^{2k^2/T})^{-\frac{1}{2}}. \tag{41'}$$

(c) $\beta = 4, g = 4, \gamma = 1$:

$$\zeta^2 - \zeta = \alpha \quad \text{or} \quad \zeta = \tfrac{1}{2}[1 + (1+4\alpha)^{\frac{1}{2}}],$$
$$\frac{P}{T} = (2\pi)^{-1}\int_{-\infty}^{\infty} dk \ln[\tfrac{1}{2}(1 + 4Ze^{-k^2/T})^{\frac{1}{2}} + \tfrac{1}{2}], \tag{42}$$
$$d = \pi^{-1}\int_{-\infty}^{\infty} dk\, Ze^{-k^2/T}[1 + 4Ze^{-k^2/T} + (1 + 4Ze^{-k^2/T})^{\frac{1}{2}}]^{-1}. \tag{42'}$$

III. COMPARISON OF RESULTS WITH KNOWN THERMODYNAMIC PROPERTIES

In this final section, we shall test the approximation of the first section as applied to the g/r^2 potential in the second section. As has been remarked, this potential presents a unique opportunity to test, since the 3-body problem is separable and hence exactly soluble. This enables us to calculate the third virial coefficient, or equivalently the third coefficient B_3 in the fugacity expansion of Eq. (32).

Preliminary to evaluating B_3 exactly, let us first modify Calogero's solution[5] for the 3-body Hamiltonian

$$H = -\sum_{i=1}^{3}\frac{\partial^2}{\partial x_i^2} + \sum_{i<j=1}^{3}\frac{g}{(x_i - x_j)^2} + \omega^2\sum_{i=1}^{3}x_i^2. \tag{43}$$

We first transform to the variables

$$R = \tfrac{1}{3}(x_1 + x_2 + x_3),$$
$$x = 2^{-\frac{1}{2}}(x_1 - x_2),$$
$$y = 6^{-\frac{1}{2}}(x_1 + x_2 - 2x_3), \tag{44}$$
$$r^2 = x^2 + y^2,$$
$$\tan\phi = x/y.$$

The Schrödinger equation becomes

$$H_R\psi + \left(-\frac{\partial^2}{\partial r^2} - \frac{1}{r}\frac{\partial}{\partial r} + \omega^2 r^2 + \frac{M_\phi}{r^2}\right)\psi = E\psi, \tag{45}$$

$$H_R = -\frac{1}{3}\frac{\partial^2}{\partial R^2} + 3\omega^2 R^2, \tag{46}$$

$$M_\phi = -\frac{\partial^2}{\partial\phi^2} + \frac{g}{2}\left(\frac{1}{\sin^2} + \frac{1}{\sin^2(\phi + \frac{2}{3}\pi)} + \frac{1}{\sin^2(\phi + \frac{4}{3}\pi)}\right). \tag{47}$$

Thus the problem is separable.

Only the operator H_R of Eq. (46) differs from Calogero's problem. We seek a solution $\psi = \psi_\phi(\phi)\psi_r(r)\psi_R(R)$. First we solve $M_\phi\psi_\phi = m\psi_\phi$ to find

$$\psi_\phi(n) = (\sin 3\phi)^\lambda C_n^\lambda(\cos\phi),$$
$$m = [3(n+\lambda)]^2, \quad n = 0, 1, 2, \cdots. \tag{48}$$

C_n^λ is a Gegenbauer polynomial. We then solve for ψ_r finding

$$\psi_r^{(l,n)} = r^{3(l+\lambda)}e^{-\frac{1}{2}\omega r^2}L_n^{3(l+\lambda)}(\omega r^2),$$
$$E_{n,l} = 2\omega[2n + 1 + 3(l+\lambda)], \quad n = 0, 1, 2, \cdots. \tag{49}$$

L_n^j is a Laguerre polynomial. Finally, the center of mass equation (46) has a solution

$$\psi_R^{(m)} = e^{-\frac{3}{2}\omega R^2}\,\mathrm{Hm}[(3\omega)^{\frac{1}{2}}R],$$
$$Em = 2\omega(m + \tfrac{1}{2}), \quad m = 0, 1, 2, \cdots. \tag{50}$$

Hm is a Hermite polynomial. The total energy E is given by

$$E = E_{n,l} + E_m = 2\omega[3l + 2n + m + \tfrac{3}{2}] + 6\omega\gamma. \quad (51)$$

The first term in this expression is simply the free fermion expression, while the last term displaces the entire free fermion spectrum by $6\omega\gamma$. This is a surprisingly simple result. Generally, Calogero made a conjecture, equivalent to the conjecture in our case, that for any N, the spectrum is that of free fermions in a harmonic well ω displaced by $\gamma\omega N(N-1)$. This is verified for $N \leq 3$.

We now wish to compare these exact results with the calculation of the previous section. We do this by directly calculating a fugacity expansion for a g/r^2 system in a macroscopic harmonic well. At the same time, we may easily place the system of Sec. 2 in the same harmonic well, so that we may compare the first three coefficients in the fugacity expansion.

Let us write the grand canonical ensemble as

$$Q(Z, \omega, T) = e^{\Omega} = \sum_{N=1}^{\infty} Z^N Q_N(\omega, T). \quad (52)$$

Ω is an extensive variable. We find the coefficients of a fugacity expansion

$$\Omega = \omega^{-1} \sum_{n=1}^{\infty} B_n'(T) Z^n, \quad (53)$$

by the usual derivation of cluster expansions, to be

$$B_1' = \lim_{\omega \to 0} \omega Q_1,$$
$$B_2' = \lim_{\omega \to 0} \omega[Q_2 - Q_1^2/2], \quad (54)$$
$$B_3' = \lim_{\omega \to 0} \omega[Q_3 - Q_1 Q_2 + \tfrac{1}{3}Q_1^3], \quad \text{etc.}$$

If q_N denotes the partition function for the free fermions in a harmonic well ω, then Calogero's conjecture gives

$$Q_N = e^{-\beta\omega\gamma N(N-1)} q_N. \quad (55)$$

This is correct for $N \leq 3$. Using the expressions for q_N, we find it an easy task to evaluate B_1', B_2', and B_3' from Eq. (54). We leave this task for the reader, quoting the values

$$B_1' = \tfrac{1}{2}T, \quad B_2' = \tfrac{1}{8}T(2\gamma + 1), \quad B_3' = \tfrac{1}{36}T(3\gamma + 1)(3\gamma + 2). \quad (54')$$

On the other hand, the expression of Eq. (32) allows us to calculate Ω as

$$\Omega = \int_{-\infty}^{\infty} \frac{P(x)}{T} dx = \frac{1}{2}\left(\frac{T}{\pi}\right)^{\frac{1}{2}} \sum_{n=1}^{\infty} B_n Z^n \int_{-\infty}^{\infty} e^{-n\omega^2 x^2/T} dx = \frac{T}{2\omega} \sum_{n=1}^{\infty} \frac{B_n Z^n}{n^{\frac{1}{2}}}, \quad (56)$$

where B_n is given by Eq. (36).

Comparing Eq. (54′) and (56), we see that they are the same; our approximation gives the exact second and third virial coefficients.

In all instances where exact results exist, our approximation agrees. We collect these cases:

(i) free fermions or bosons when $g = 0$;
(ii) second and third virial coefficients;
(iii) zero temperature thermodynamics.

The last two points indicate agreement in opposite regimes. We are thus led to the following.

Conjecture: All results of the second section are exact.

It is very likely, although not proven, that Calogero's conjecture implies our conjecture through Eqs. (53) and (56); the converse cannot be true. However, it is believed that the approach of this paper reveals the physical basis of the conjecture to be the absence of diffraction effects in many-body scattering.

We include as an appendix another amusing instance where an approximation applied to the g/r^2 potential yields the exact solution.

APPENDIX

In this appendix, we wish to present a very simple example where again an approximation applied to the g/r^2 potential problem yields an exact result. The example, being a 1-body problem, has little to do with the more complicated many-body problem, yet it gives insight into the peculiarities of the g/r^2 potential and supports the conjecture made in this paper.

We consider a single particle interacting with a 1-dimensional rigid lattice of g/r^2 potentials; the lattice constant is b. However, we first present an approximation, reminiscent of the approximation in the text of the paper, valid for any potential in one dimension which has a phase shift. This approximation is well known in solid state physics.

Let the Schrödinger equation be

$$\left(-\frac{1}{2}\frac{d^2}{dx^2}+U(x)\right)\psi = E\psi, \tag{A1}$$

with

$$U(x)=\sum_{n=-\infty}^{+\infty} V(x-nb). \tag{A2}$$

$V(x)$ is the potential between one lattice point and the particle; we assume it to be symmetrical. $U(x)$ is periodic with period b.

We now divide the lattice into equal cells, the nth being $(n-\frac{1}{2})b$ to $(n+\frac{1}{2})b$. Let b be so large that there is no overlap of potentials. We write the solution in the nth cell as

$$\psi(x)=\psi_1^{(n)}\phi_1^{(n)}(kx)+\psi_2^{(n)}\phi_2^{(n)}(kx). \tag{A3}$$

$\phi_1^{(n)}(kx)$ is a solution in the nth cell corresponding to a plane wave of momentum k, incident from the right on a scatterer at the origin, with amplitude chosen to make the amplitude of the outgoing wave on the left unity. $\phi_2^{(n)}(kx)$ is the time reversal of $\phi_1^{(n)}(kx)$. We may then define a transfer matrix M, acting on the vector

$$\psi^{(n)} \equiv (\psi_1^{(n)}, \psi_2^{(n)}), \quad \text{so that} \quad M\psi^{(n)} = \psi^{(n+1)}. \tag{A4}$$

In terms of the transmission and reflection amplitudes T and R, we find

$$M = \begin{vmatrix} e^{ibk}/T & R^*/T^* \\ R/T & e^{-ikb}/T^* \end{vmatrix}. \tag{A5}$$

We shall parametrize the transmission amplitude by

$$T = \exp[i\delta(k)]\cos\theta(k). \tag{A6}$$

To obtain a band structure, we require the solution to be periodic after N sites. This in turn requires $M^N = I$, and, with $\det M = 1$, we find

$$\operatorname{Tr} M = 2\cos(2\pi n/N), \quad n = 1, 2, \cdots, N,$$

or

$$\cos(bk-\delta)/\cos\theta = \cos(2\pi n/N). \tag{A7}$$

We solve this transcendental equation for k as a function of n and a band index, and calculate the energy by

$$E = \tfrac{1}{2}k^2. \tag{A8}$$

Our approximation consists of assuming Eqs. (A7) and (A8) to hold for *all* b. Obviously, for the δ-function potential, our results will be exact since point interactions cannot overlap.

We next apply the previous approximation to the g/r^2 potential, when $0 \geq g \geq -\frac{1}{4}$. First, we see that Eq. (A2) gives

$$U(x) = g\sum_{n=-\infty}^{\infty}(x-nb)^{-2} = g\pi^2 b^{-2}\sin^{-2}(\pi x/b). \tag{A9}$$

The resulting differential Eq. (A1) is identical with Eq. (47) of the text; this band problem has been solved exactly by Scarf.[6] We shall use Scarf's matching condition through the singular potential, where the most singular solution ψ_- is taken symmetrical and the least singular solution ψ_+ is taken antisymmetrical. As discussed in Paper I, this is an artificial choice, yet necessary for a nontrivial band problem. We easily find the transmission and reflection amplitudes to be

$$T = \sin(\pi a), \quad R = -i\cos(\pi a). \tag{A10}$$

They are independent of k. Equation (A7) gives

$$\cos(kb) = \sin(\pi a)\cos(2\pi n/N)$$

or

$$k = \frac{\cos^{-1}[\sin(\pi a)\cos(2\pi n/N)]}{b}. \tag{A11}$$

Thus the allowed bands of k values are of equal width and equally spaced. The energy is given by Eq. (A8) to be

$$E = \frac{\{\cos^{-1}[\sin(\pi a)\cos(2\pi n/N)]\}^2}{2b^2}. \tag{A12}$$

Upon comparing this result with Scarf's exact solution, we find them to be identical for *all* b, whether the potentials overlap or not. Thus, once again, an approximation applied to the g/r^2 potential, where it would clearly seem to be inappropriate, has in fact yielded exact results.

Addendum: After this paper was submitted, two preprints have appeared which may be used to supply a proof of our final conjecture. In the first preprint, F. Calogero[7] proves his own conjecture, our Eq. (55). One may then easily evaluate the grand canonical ensemble of Eqs. (52) and (53) by saddle point methods, as done by C. Marchioro and E. Presutti[8] in the second preprint. The result is then seen to be identical to our Eq. (56); actually it is easiest to compare the average number of particles $\bar{N} = \partial\Omega/\partial \ln Z$.

* Work supported in part by National Science Foundation Grant GP 13632.

[1] B. Sutherland, J. Math. Phys. **12**, 246 (1971). This paper will be referred to as I, and equation numbers preceded by I.

[2] E. Lieb and W. Liniger, Phys. Rev. **130**, 1605 (1963); E. Lieb, *ibid.*, 1616 (1963).

[3] C. N. Yang and C. P. Yang, J. Math. Phys. **10**, 1115 (1969).

[4] H. A. Bethe, Z. Physik **71**, 205 (1931).

[5] F. Calogero, J. Math. Phys. **10**, 2191, 2197 (1969).

[6] F. L. Scarf, Phys. Rev. **112**, 1137 (1958).

[7] F. Calogero, J. Math. Phys. **12**, March, 1971 (to be published).

[8] C. Marchioro and E. Presutti, "Exact Statistical Model in One Dimension," preprint.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Some Properties of Cylindrically Symmetric Einstein–Maxwell Fields

JOHN L. SAFKO
University of South Carolina, Columbia, South Carolina 29208
AND
LOUIS WITTEN
University of Cincinnati, Cincinnati, Ohio 45221

(Received 27 August 1970)

Some properties of static cylindrically symmetric universes consisting of gravitational and electromagnetic fields with a central axial mass, charge, or current density are discussed. Of three permissible distinct configurations, general solutions of the appropriate equations are given for two. The properties of the solutions of these two are then examined in greater detail. The equations of motion of test particles (both charged and uncharged) are described for some simple trajectories. The qualitative effect of the magnetic field on uncharged test particle behavior in some instances is noted as well as the qualitative distinction in the behavior of charged and uncharged test particles. The algebraically independent invariants of the Riemann tensor are calculated. A stability analysis is carried out for the two solutions when the system is subjected to radial perturbations. It turns out that both solutions are stable. The stability analysis is carried out by using the equations of the already unified field theory of Rainich, Misner, and Wheeler.

1. INTRODUCTION

This paper reviews some aspects of static cylindrically symmetric solutions of the combined Einstein–Maxwell gravitational and electromagnetic fields. When there is a central axial mass density and no electromagnetic field, Einstein's theory introduces two integration constants into the solutions. One is related to the mass density; the other quite possibly arises from the fact that a central mass density is a limiting case of a mass distribution of finite extent in the plane perpendicular to the central axis. Hence the second constant is related to the radius over which the mass distribution extends. This paper shows that the solutions for the situation with vanishing electromagnetic field is implicitly contained in the solutions with electromagnetic fields present in the limit that the latter fields vanish. The existence of electromagnetic fields introduces a third constant of integration.

When a central mass density and an electromagnetic field are both permitted to be present, three possible configurations exist. The first is the central mass together with an axial current producing a magnetic field whose field lines are circles in the plane perpendicular to the axis and centered about the axis. The second configuration is described by a central mass density together with magnetic and electric fields parallel to the axis and no currents or charges anywhere. A special case of this permits the central mass density to vanish and leave only the magnetic and electric fields as sources of the gravitational fields. The magnetic field is held together by its own gravitational attraction. It has been said that the same possibility exists in the Newton–Maxwell theory also—namely, that the repulsion due to the gradient of the magnetic pressure (assuming the case of no electric field as well as no central mass density) is balanced by the gravitational attraction of the different parts of the field for each other. However, if, in addition, the magnetic field must have vanishing curl, it must be a constant in the Newton–Maxwell theory since it is always in the axial direction. This is not the same type of possibility that is permitted by the Einstein–Maxwell field, where the field has different values for different distances from the axis.

The third configuration possible is when there is a central mass density and a central charge density existing together with a radial electric field. The most general solutions possible have been found for the first two configurations; only a restricted range of exact solutions have been found for the last. Hence this paper deals in detail only with the first two configurations.

In the next section the three configurations are described in somewhat greater detail and the solutions are explicitly stated. In Sec. 3 the equations of motion of test charged and uncharged particles are given together with the first integrals. Some specific geodesic and charged particle motion is described. It is found that the electromagnetic field makes some unexpected qualitative changes in the motion of uncharged test particles and that charged particle motion differs unexpectedly from uncharged particle motion.

Section 4 calculates the algebraically independent invariants that can be formed from the Riemann tensor. There are three of these corresponding to the three integration constants that are required by the equations. In Sec. 5, an analysis is made of the stability of the solutions of the first two configurations described above to radial perturbations. The first configuration is stable whether the central mass density is

positive or negative. The second configuration is stable whether the central mass density is positive, zero, or negative. However, when the mass is negative and large in magnitude, the analysis is not quite clear. The prediction of stability or not depends on the choice of boundary condition. Both solutions admit undamped, oscillatory behavior as perturbed solutions to the linearized equations.

2. STATIC CYLINDRICALLY SYMMETRIC SOLUTIONS OF THE EINSTEIN-MAXWELL FIELD

The Einstein–Maxwell equations for a combined gravitational and electromagnetic field in the absence of electromagnetic sources are[1]

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -\tfrac{1}{2}T_{\mu\nu} \equiv -\tfrac{1}{4}\omega_{\mu\alpha}\bar{\omega}^{\alpha}_{\nu}, \quad (2.1)$$

$$\omega_{\alpha\beta,\gamma} + \omega_{\beta\gamma,\alpha} + \omega_{\gamma\alpha,\beta} = 0, \quad (2.2)$$

$$\omega_{\mu\nu} \equiv f_{\mu\nu} + \tfrac{1}{2}(-g)^{\frac{1}{2}}\epsilon_{\mu\nu\alpha\beta}f^{\alpha\beta}. \quad (2.3)$$

$f_{\mu\nu}$ is the antisymmetric electromagnetic field which in a local Minkowski frame is determined by $f_{01} = E_x$, $f_{12} = H_z$, etc.; ϵ is the completely antisymmetric Levi-Civita tensor density and g is det $g_{\mu\nu}$. In the presence of an electromagnetic charge-current vector j^{μ}, Eq. (2.2) can be written in the integral form

$$\iint_{\mu<\nu} \omega_{\mu\nu}d(x^{(\mu)}, x^{(\nu)}) = i\iiint_{\mu<\sigma<\tau} (-g)^{\frac{1}{2}}\epsilon_{\mu\nu\sigma\tau}j^{\mu}d(x^{(\nu)}, x^{(\sigma)}, x^{(\tau)}). \quad (2.4)$$

The left-hand integral is a surface integral taken over a closed 2-dimensional surface. If the surface is described by the parameters μ_1 and μ_2, we have

$$d(x^{(\mu)}, x^{(\nu)}) = \frac{\partial(x^{(\mu)}, x^{(\nu)})}{\partial(\mu_1, \mu_2)}\,d\mu_1\,d\mu_2. \quad (2.5)$$

The right-hand integral of (2.4) is the volume integral over the 3-dimensional volume enclosed by the surface over which the left side was integrated.

The theory described by (2.1), (2.2), and (2.3) is equivalent to a purely geometrical theory for which the following equations are valid[2]:

$$R \equiv R_{\mu}{}^{\mu} = 0, \quad (2.6)$$

$$R_{\mu}{}^{\alpha}R_{\alpha}{}^{\nu} = \tfrac{1}{4}R_{\alpha\beta}R^{\alpha\beta}\delta_{\mu}{}^{\nu}, \quad (2.7)$$

$$\alpha_{\mu,\nu} - \alpha_{\nu,\mu} = 0, \quad (2.8)$$

$$\alpha_{\mu} \equiv (-g)^{\frac{1}{2}}\epsilon_{\sigma\nu\lambda\mu}R^{\lambda\gamma;\mu}R_{\gamma}{}^{\nu}/R_{\alpha\beta}R^{\alpha\beta}, \quad (2.9)$$

$$R_{\mu\nu}v^{\mu}v^{\nu} \geq 0. \quad (2.10)$$

Equation (2.10) must be true for any timelike vector v^{μ}. Actually, the geometrical theory (2.6)–(2.10) is equivalent to the usual theory (2.1)–(2.3) only for nonnull fields, i.e., fields for which $R_{\alpha\beta}R^{\alpha\beta}$ do not vanish. We shall only be dealing with nonnull fields in this paper.

If Eqs. (2.6) and (2.7) are valid, it can be shown that there are two null eigenvectors, k_{μ} and l_{μ} of $R_{\mu\nu}$ normalized such that $k_{\mu}l^{\mu} = \frac{1}{4}R_{\alpha\beta}R^{\alpha\beta}$. $\omega_{\mu\nu}$ can be determined from k_{μ} and l_{μ} as follows:

$$R_{\mu}{}^{\nu}k_{\nu} = \tfrac{1}{2}R_{\alpha\beta}R^{\alpha\beta}k_{\mu}, \quad R_{\mu}{}^{\nu}l_{\nu} = \tfrac{1}{2}R_{\alpha\beta}R^{\alpha\beta}l_{\mu}, \quad (2.11)$$

$$\omega_{\mu\nu} = \frac{4}{(R_{\alpha\beta}R^{\alpha\beta})^{\frac{1}{2}}}[l_{\mu}k_{\nu} - k_{\mu}l_{\nu} + i(-g)^{\frac{1}{2}}\epsilon_{\mu\lambda\nu\sigma}k^{\lambda}k^{\sigma}]e^{i\theta}. \quad (2.12)$$

With ω chosen in this way and θ an arbitrary function of the coordinates, (2.1) is satisfied. If and only if (2.8) is satisfied, θ can be chosen up to a constant by a line integral[2] of α_{μ} so that the resultant ω satisfies (2.2). Our paper is concerned with a coordinate system in which α_{μ} vanishes identically; hence (2.8) is certainly true, and θ is an arbitrary constant.

The most general metric with cylindrical symmetry can be written in the form[3]

$$ds^2 = e^{2\gamma-2\psi}(-dt^2 + d\rho^2) + \rho^2e^{-2\psi}\,d\phi^2 + e^{2\psi+2\mu}\,dz^2,$$
$$-\infty < t = x^0 < \infty, \quad 0 \leq \phi = x^2 < 2\pi,$$
$$-\infty < z = x^3 < \infty, \quad 0 \leq \rho = x^1 < \infty, \quad (2.13)$$

where γ, ψ, and μ are functions of ρ alone. We shall review the solutions to the Einstein–Maxwell field equations for this metric.[4] The components of the Ricci tensor for the metric are written in Refs. 2–4. When $\mu = 0$ and γ and ψ depend on both ρ and τ, the components are given in Eqs. (5.3)–(5.6). For $\mu \neq 0$ and all variables dependent on ρ alone, $R_0{}^0$, to which we often refer, is

$$R_0{}^0 = -e^{2\psi-2\gamma}[\gamma_{\rho\rho} - \psi_{\rho\rho} - (\psi_{\rho}/\rho) + (\gamma_{\rho}/\rho) + \gamma_{\rho}\mu_{\rho} - \psi_{\rho}\mu_{\rho}]. \quad (2.14)$$

Equation (2.8) is satisfied for the metric since $\alpha_{\mu} \equiv 0$. Equations (2.6) and (2.7) become

$$R_0{}^0 + R_1{}^1 + R_2{}^2 + R_3{}^3 = 0, \quad (2.15)$$

$$(R_0{}^0)^2 = (R_1{}^1)^2 = (R_2{}^2)^2 = (R_3{}^3)^2. \quad (2.16)$$

It is easy to see that there are three qualitatively different possibilities:

$$\text{Case I: } R_0{}^0 = -R_1{}^1 = R_2{}^2 = -R_3{}^3;$$
$$\text{Case II: } R_0{}^0 = -R_1{}^1 = -R_2{}^2 = R_3{}^3; \quad (2.16')$$
$$\text{Case III: } R_0{}^0 = R_1{}^1 = -R_2{}^2 = -R_3{}^3.$$

For both Cases I and II, $R_0{}^0 + R_1{}^1 = 0$. Under these conditions, one can consider[3] $\mu \equiv 0$. In Case III the

character of the electromagnetic field is completely determined by μ. Let us consider each case in turn.

Case I: $R_0^0 = -R_1^1 = R_2^2 = -R_3^3$

In this case the most general solution[5] can be described by the line element

$$ds^2 = -\rho^{2c^2+2c}(1 + k\rho^{-2c})^2 e^{2a}(dt^2 - d\rho^2) + \rho^{2c+2}(1 + k\rho^{-2c})^2\, d\phi^2 + \rho^{-2c}(1 + k\rho^{-2c})^{-2}\, dz^2. \tag{2.17}$$

Also

$$R_{00} = -4c^2k/g_{22}. \tag{2.18}$$

c, k, and a are integration constants, with $k \geq 0$ to satisfy (2.10). $\omega_{\mu\nu}$ is found by the prescription given. Its phase is chosen so that an integration of (2.4) over a cylinder of unit length including the central axis yields a real value for the current I along the axis. The axial current and the field are

$$I = 4\pi c(k)^{\frac{1}{2}}, \tag{2.19}$$

$$f_{\rho z} = 2c(k)^{\frac{1}{2}} g_{33}/\rho. \tag{2.20}$$

All other components of the electromagnetic field vanish. In the terminology of flat space, a current in the z direction is producing a magnetic field in the ϕ direction. If $k = 0$, we recover the most general solution for an Einstein space with cylindrical symmetry.

To comment for a moment on the solution with $k = 0$, we see that the line element contains two arbitrary constants c and a. We shall show later that for weak fields $\frac{1}{2}c$ can be identified as the mass per unit length of the Newtonian theory under certain conditions. Whereas the linear mass density is the only parameter in the Newtonian theory, it is not so in the Einstein theory. The parameter a plays a role. We shall later show that the values of the scalar invariants that can be constructed from the Riemann tensor in the Einstein and Einstein–Maxwell cases depend on both c and a in the former case and on c, k, and a in the latter. If k and c both vanish, the space is not Euclidean unless $a = 0$; the limit $(\rho \to 0)$ of the ratio of the circumference to radius of a small circle perpendicular to the axis is $2\pi \exp(-a)$. Marder[6] has discussed several possible physical reasons why a second parameter is necessary in the purely gravitational case. A real physical situation has mass distributed from the axis to a distance ρ_0; a is needed to match the line elements properly at ρ_0. Also, a is needed as a measure of some parameter of a physical system which is stationary, undergoes radial motions of some kind, and then is stationary again; a is thus somehow related to the fact that gravitational radiation occurred. Demanding that space becomes flat if $c = 0$ (no mass and no current) requires that a vanishes whenever c does.

It is interesting that if k vanishes, the electromagnetic field vanishes, but that the mass density and the gravitational field do not. However, if c vanishes, the mass density and the electromagnetic field both disappear. Of course, we know that nature permits masses without currents but not currents without masses, consistent with the results obtained in this case. This result may be contrasted with the situation in Case II, where the solution will permit an electromagnetic field without masses. However, for Case II, there will be no currents either; presumably both currents and masses have been moved to infinite values of z, and the solution cannot elucidate the limiting physical situation regarding the sources present there.

Let $\lambda^\mu{}_{(\alpha)}$ be an orthonormal tetrad with $\lambda^\mu{}_{(0)}$ timelike and pointing to the future. From $f_{\alpha\beta}$ we can form the invariants

$$f_{(\alpha\beta)} = f_{\mu\nu}\lambda^\mu{}_{(\alpha)}\lambda^\nu{}_{(\beta)}. \tag{2.21}$$

Sometimes invariants formed in this way are called the "physical" components of the corresponding tensor. If we choose

$$\lambda_{(0)} = ((-g^{00})^{\frac{1}{2}}, 0, 0, 0), \tag{2.22}$$

$$\lambda_{(1)} = (0, (g^{11})^{\frac{1}{2}}, 0, 0), \tag{2.23}$$

$$\lambda_{(2)} = (0, 0, (g^{22})^{\frac{1}{2}}, 0), \tag{2.24}$$

$$\lambda_{(3)} = (0, 0, 0, (g^{33})^{\frac{1}{2}}), \tag{2.25}$$

the physical component of the electromagnetic field $f_{\rho z}$ [Eq. (2.20)] is

$$B_\phi \equiv f_{\rho z}(g^{11}g^{33})^{\frac{1}{2}} = \frac{2c(k)^{\frac{1}{2}}e^a}{\rho^{c^2+2c+1}(1 + k\rho^{-2c})^2}. \tag{2.26}$$

Case II: $R_0^0 = -R_1^1 = -R_2^2 = R_3^3$

The most general solution for this case is described by the line element

$$ds^2 = -\rho^{2c^2+2c}(1 + k\rho^{-2c})^2 e^{2a}(dt^2 - d\rho^2) + \rho^{-2c}(1 + k\rho^{-2c})^{-2}\, d\phi^2 + \rho^{2c+2}(1 + k\rho^{-2c})^2\, dz^2, \tag{2.27}$$

with

$$R_{00} = -4c^2k/g_{33}. \tag{2.28}$$

As before, c, k, and a are integration constants with $k \geq 0$. The space becomes flat for $c = -1$, $k = 0$, and $a = 0$. We shall show later that $\delta \equiv -(1 + c)$ can, under certain conditions, be interpreted as twice the mass m per unit length along the axis, $c = -1$ corresponding to vanishing mass density. For $c = -1$

and $k = 0$, the limit $\rho \to 0$ of the ratio of circumference to the radius of a circular perpendicular to the axis is again $2\pi \exp(-a)$. When $k \neq 0$, the line element (2.7) is invariant to the following transformation of constants and coordinates:

$$c = -c', \quad k = 1/k', \quad e^{2a} = k'^2 e^{2a'},$$
$$z = k'^2 z', \quad \phi = \phi'/k', \quad \rho = \rho', \quad t = t'. \quad (2.29)$$

Of course, the range of ϕ' should be readjusted such that $0 \leq \phi' < 2\pi$. It can be shown that such a readjustment is equivalent to a change in the constant a'. Because of the invariance, we shall henceforth, if $k \neq 0$, discuss (2.27) only under the consideration $c \leq 0$.

Finding $f_{\mu\nu}$ from $\omega_{\mu\nu}$ by the prescription given yields

$$f_{tz} = 2c(k)^{\frac{1}{2}} \cos\theta, \quad (2.30)$$

$$f_{\rho\phi} = -2c(k)^{\frac{1}{2}} g_{22} (\sin\theta)/\rho, \quad (2.31)$$

where θ is the arbitrary phase factor of $\omega_{\mu\nu}$. There are no charges or currents along the axis. Using the tetrad (2.22)–(2.25), we see that the physical components of the field are

$$E_z \equiv f_{tz}(-g^{00}g^{33})^{\frac{1}{2}} = -\frac{2c(k)^{\frac{1}{2}} e^{-a} \cos\theta}{\rho^{c^2+2c+1}(1 + k\rho^{-2c})^2}, \quad (2.32)$$

$$B_z = f_{\rho\phi}(g^{11}g^{22})^{\frac{1}{2}} = -\frac{2c(k)^{\frac{1}{2}} e^{-a} \sin\theta}{\rho^{c^2+2c+1}(1 + k\rho^{-2c})^2}. \quad (2.33)$$

For $c = -1$ and $a = 0$, the mass density vanishes, and the metric and electromagnetic field intensities are everywhere regular. This special choice with $\theta = \frac{1}{2}\pi$ gives a purely magnetic universe and has been studied in detail by Melvin[7,8] and Thorne.[9,10] If $\delta = -(1 + c) = 2m$ (central mass density) and $c = -1$ corresponds to flat space, a small positive mass density requires that c be somewhat less than -1. Suppose $\theta = \frac{1}{2}\pi$ and $a = 0$; then

$$B_z = 2(1 + \delta)k^{\frac{1}{2}} \rho^{-\delta^2}(1 + k\rho^{2+2\delta})^{-2}. \quad (2.34)$$

It is remarkable that, when there is no axial mass, $\delta = 0$, the magnetic field is regular along the axis; when there is a central mass, $2\delta = m$, no matter how small, the magnetic field is infinite along the axis—even though there is no charge or current distribution there as determined by Eq. (2.4). Again as in Case I, there is an intimate and surprising relationship between the existence of an electrically neutral mass density and the behavior of the electromagnetic field. In Case I, an electromagnetic field is impossible without an axial mass; in Case II, an axial mass leads to an infinite axial field.

Case III: $R_0^0 = R_1^1 = -R_2^2 = -R_3^3$

Since $R_0^0 + R_1^1 \neq 0$, we must retain μ in the general metric. Equations (2.7) to be solved in this case reduce to

$$2\psi_{\rho\rho} + \mu_{\rho\rho} + \mu_\rho^2 + 2\psi_\rho\mu_\rho + \frac{2\psi_\rho}{\rho} = 0, \quad (2.35)$$

$$\mu_\rho^2 + \mu_{\rho\rho} - 2\gamma_\rho\mu_\rho + 4\mu_\rho\psi_\rho + 2\psi_\rho^2 - \frac{2\gamma_\rho}{\rho} = 0, \quad (2.36)$$

$$2\psi_{\rho\rho} - \gamma_{\rho\rho} + \frac{2\psi_\rho}{\rho} - \frac{\gamma_\rho}{\rho} - \frac{\mu_\rho}{\rho} + 2\psi_\rho\mu_\rho - \gamma_\rho\mu_\rho = 0. \quad (2.37)$$

It can be shown[4] that the general solution to these equations can be found provided that a single integro-differential equation can be solved. We have not succeeded in finding the general solution to these equations. We can, however, determine some characteristic of the general situation without an explicit solution. Since Case I involved a ϕ component of the electromagnetic field and Case II a z component, we would expect Case III to involve a ρ component. This is indeed the case:

$$R_{00} = -[(\mu_\rho/\rho) + \tfrac{1}{2}(\mu_\rho^2 + \mu_{\rho\rho})]. \quad (2.38)$$

Finding $\omega_{\mu\nu}$, using (2.4), and defining the real-valued expression

$$Q = 2\pi \int_0^\rho \int_0^{2\pi} (-g)^{\frac{1}{2}} j^0 (g_{11}g_{22})^{\frac{1}{2}} \, d\rho \, d\phi \quad (2.39)$$

yields

$$Q = 2\pi(-R_{00})^{\frac{1}{2}} \rho^2 e^{\mu-\gamma}, \quad (2.40)$$

$$f_{01} = (-R_{00})^{\frac{1}{2}} e^{\gamma-\psi}. \quad (2.41)$$

All other components of $f_{\mu\nu}$ vanish. We have

$$E_\rho = f_{0\rho}(-g^{00}g^{11})^{\frac{1}{2}} = (-R_{00})^{\frac{1}{2}} e^{\psi-\gamma} = Qe^{\psi-\mu}\rho^{-1}/2\pi. \quad (2.42)$$

As expected, there is in this case a central charge density together with a ρ component of the electric field. In spite of the appearance of (2.40), Q is independent of ρ, as follows from its definition and Maxwell's equations (2.2) and (2.4).

Because of our inability to produce the general expression for the metric in Case III, we shall restrict the discussion in the remainder of the paper to Cases I and II. A discussion of approximate solutions for Case III has been given by Safko.[11]

3. MOTION OF TEST PARTICLES

In this section the first integrals of the equation of motion of test particles for both charged and neutral

test particles are derived. In the case of neutral test particles, the motion is along geodesics. Of course, the test particles are considered as passive objects—ignored are radiation and other electromagnetic and gravitational effects produced by the charge and mass of the test particles. The equations of motion of test particles of charge e and mass m are given by

$$m\frac{D^2x^\mu}{Ds^2} = ie\frac{dx^\nu}{ds}f_\nu{}^\mu. \tag{3.1}$$

The right-hand side is purely imaginary since $(ds)^2 = -(d\tau)^2$, where τ is the proper time of the particle.

Case I: $R_0{}^0 = -R_1{}^1 = R_2{}^2 = -R_3{}^3$

For Case I, the equations of motion (3.1) become

$$\frac{d^2t}{ds^2} + \frac{g_{00,\rho}}{g_{00}}\frac{dt}{ds}\frac{d\rho}{ds} = 0, \tag{3.2}$$

$$\frac{d^2\phi}{ds^2} + \frac{g_{22,\rho}}{g_{22}}\frac{d\psi}{ds}\frac{d\rho}{ds} = 0, \tag{3.3}$$

$$\frac{d^2z}{ds^2} + \frac{g_{33,\rho}}{g_{33}}\frac{dz}{ds}\frac{d\rho}{ds} = i\frac{e}{m}\frac{2c(k)^{\frac{1}{2}}e^a}{\rho}\frac{d\rho}{ds}, \tag{3.4}$$

$$\frac{d^2\rho}{ds^2} + \frac{1}{2}\frac{g_{11,\rho}}{g_{11}}\left(\frac{d\rho}{ds}\right)^2 - \frac{1}{2}\frac{g_{00,\rho}}{g_{11}}\left(\frac{dt}{ds}\right)^2 - \frac{1}{2}\frac{g_{22,\rho}}{g_{11}}\left(\frac{d\phi}{ds}\right)^2 - \frac{1}{2}\frac{g_{33,\rho}}{g_{11}}\left(\frac{dz}{ds}\right)^2 = i\frac{e}{m}\frac{2c(k)^{\frac{1}{2}}e^a}{\rho}\frac{g_{33}}{g_{00}}\frac{dz}{ds}. \tag{3.5}$$

The first three of these equations can be integrated to give (E, L, and J being integration constants)

$$g_{00}\frac{dt}{ds} \equiv iE, \tag{3.6}$$

$$g_{22}\frac{d\psi}{ds} \equiv iJ, \tag{3.7}$$

$$g_{33}\frac{dz}{ds} \equiv i\left[L + 2\left(\frac{e}{m}\right)\frac{e^a}{k^{\frac{1}{2}}(1 + k\rho^{-2c})}\psi(k, c)\right] \equiv i\mathfrak{L}(\rho). \tag{3.8}$$

ψ is a function which is equal to zero when either k or c equals zero and is equal to one otherwise. A first integral of Eq. (3.5) is given by the line element

$$1 = g_{00}\left(\frac{dt}{ds}\right)^2 + g_{11}\left(\frac{d\rho}{ds}\right)^2 + g_{22}\left(\frac{d\phi}{ds}\right)^2 + g_{33}\left(\frac{dz}{ds}\right)^2. \tag{3.9}$$

For timelike motion, E^2, J^2, and $\mathfrak{L}^2$ must all be positive quantities or zero so that E, J, and $\mathfrak{L}$ are all real. For spacelike motion, E^2, J^2, and $\mathfrak{L}^2$ must all be negative quantities. The motion is geodesic if $e = 0$. The geodesic motion is null if, in addition, the left-hand side of (3.9) vanishes; s is then interpreted as a parameter not related to the proper line interval, and E^2, J^2, and $\mathfrak{L}^2$ are either all positive or all negative.

The physical significance of E, J, and L can be deduced in a "weak-field, low-velocity" limit for a neutral particle. Define the velocity v by

$$ds^2 \equiv g_{00}\,dt^2(1 - v^2). \tag{3.10}$$

This together with (3.6) yields

$$E^2 = -\frac{g_{00}}{(1 - v^2)} = \frac{\rho^{2c^2+2c}(1 + k\rho^{-2c})^2e^{2a}}{(1 - v^2)}, \tag{3.11}$$

for small c, k, a, and v^2, to first order in these quantities,

$$E = 1 + c\ln\rho + \tfrac{1}{2}v^2 + a + k. \tag{3.12}$$

Compare this with the total energy U of a particle of mass M and velocity v moving in the gravitational field of an axial distribution of mass m, the velocity of light and the gravitational constant G being taken equal to unity:

$$U = M + 2Mm\ln\rho + \tfrac{1}{2}Mv^2. \tag{3.13}$$

Clearly this suggests the interpretation to first order,

$$U/M = E - (a + k), \tag{3.14}$$

$$2m = c, \tag{3.15}$$

for the constant of integration E and for the identification with the linear mass density. The interpretation can only be made for ρ sufficiently small that second-order terms are ignorable compared to first-order terms. For large ρ (say, $\ln\rho > 1$) the interpretation can no longer be made. In ordinary units it turns out that (3.15) gives $M = 2.2 \times 10^{18}c$ kg/m.

To interpret J,

$$\left(\frac{d\phi}{dt}\right)^2 = \left(\frac{d\phi}{ds}\frac{ds}{dt}\right)^2 = -J^2\frac{g_{00}}{(g_{22})^2}(1 - v^2) = \rho^{-2c^2-6c-4}\frac{e^{2a}}{(1 + k\rho^{-2c})^2}J^2(1 - v^2) \tag{3.16}$$

yields, to zeroth order in c, a, k, and v^2,

$$\rho^2\frac{d\phi}{dt} = J, \tag{3.17}$$

which identifies J as the ϕ component of angular momentum of the test particle divided by its mass.

Similarly,

$$\left(\frac{dz}{dt}\right)^2 = -\frac{g_{00}}{(g_{33})^2}(1 - v^2)L^2 \tag{3.18}$$

yields, to zeroth order,

$$\left(\frac{dz}{dt}\right) = L, \tag{3.19}$$

which identifies L as the z component of linear momentum divided by the mass of the test particle.

A complete qualitative discussion of the motion of charged, or even of neutral, particles for a large range of values of c and k is a big task, which we do not undertake at this time. Instead, we shall deal with only a few special cases and consider that the central mass density is positive. The primary purpose of the discussion is to point out that the electromagnetic field can make qualitative changes in behavior of the motion of test particles, even of neutral test particles.

Consider first motion of neutral test particles for which ρ and ϕ are fixed. In this case $J = 0$, $\mathfrak{L} = L$, and Eq. (3.5) yields

$$\frac{E^2}{\mathfrak{L}^2} = -\frac{g_{33,\rho}\,(g_{00})^2}{g_{00,\rho}\,(g_{33})^3} = -\frac{\rho^{2c^2}e^{2a}}{(g_{33})^2}\left(\frac{c\rho^{2c}(1 + k\rho^{-2c})}{2k - (c+1)(1+k\rho^{-2c})\rho^{2c}} + 1\right). \tag{3.20}$$

For the motion to be consistently spacelike or timelike, E^2 and $\mathfrak{L}^2$ must have the same sign. Hence

$$\frac{k - \rho^{2c}}{k(1-c) - \rho^{2c}(c+1)} < 0. \tag{3.21}$$

Suppose first that no electromagnetic field exists; then $k = 0$. Equation (3.21) reduces to

$$1/(c+1) < 0 \tag{3.22}$$

which cannot be satisfied for positive c. This gives the entirely expected result that, for a small central mass density (positive or negative), no motion parallel to the z axis is possible because of the gravitational attraction of the central mass. Inequality (3.22) does not have to be satisfied if the central mass disappears ($c = 0$), for then Eq. (2.5) is satisfied directly because $g_{33,\rho}$ and $g_{00,\rho}$ both vanish.

Suppose now there is a current flowing along the axis, $k > 0$, in addition to a mass, $c > 0$. Inequality (3.21) is now satisfied in the range of ρ given by

$$\begin{aligned} k(1-c)/(1+c) \equiv \rho_1^{2c} \le \rho^{2c} \le \rho_2^{2c} \equiv k, \quad & c < 1, \\ 0 \le \rho^{2c} < k, \quad & c \ge 1. \end{aligned} \tag{3.23}$$

This range shows where geodesic motion is possible; to see whether it is timelike, spacelike, or null, use (3.9):

$$1 = \left(\frac{g_{33,\rho}}{g_{00,\rho}}\frac{g_{00}}{(g_{33})^2} - \frac{1}{g_{33}}\right)\mathfrak{L}^2. \tag{3.24}$$

For timelike motion, $\mathfrak{L}^2 > 0$; for spacelike motion, $\mathfrak{L}^2 < 0$; for null motion, $\mathfrak{L}^2$ can have either sign, but its coefficient must be equal to zero. Hence

$$\frac{g_{33,\rho}}{g_{00,\rho}}\frac{g_{00}}{g_{33}} - 1 \begin{cases} > 0 & \text{timelike,} \\ < 0 & \text{spacelike,} \\ = 0 & \text{null,} \end{cases} \tag{3.25}$$

$$\frac{g_{33,\rho}}{g_{00,\rho}}\frac{g_{00}}{g_{33}} = \frac{k - \rho^{2c}}{(c+1)\rho^{2c} - (1-c)k}. \tag{3.26}$$

The analysis shows that

$$\left.\begin{aligned} \rho_1^{2c} \le \rho^{2c} < \rho_n^{2c} \equiv k\frac{(2-c)}{(2+c)}, \quad & c < 1 \\ 0 \le \rho^{2c} \le \rho_n^{2c}, \quad & c > 1 \end{aligned}\right\} \text{timelike,} \tag{3.27}$$

$$\rho = \rho_n \quad \text{null,} \tag{3.28}$$

$$\rho_n < \rho \le \rho_2 \quad \text{spacelike.}$$

All three types of geodesics are possible. The somewhat surprising result is that a circular magnetic field produced by an axial current apparently exerts a gravitational repulsion on a particle traveling along the Z axis.

Table I summarizes the results for the various simple trajectories discussed here for neutral test particles for Cases I and II. Table II summarizes the possibilities for charged test particles.

Consider now motion of neutral test particles for which ρ and z are fixed: $\mathfrak{L} = 0$, and Eq. (3.5) yields

$$\frac{E^2}{J^2} = -\frac{g_{22,\rho}\,(g_{00})^2}{g_{00,\rho}\,(g_{22})^2} = e^{2a}\frac{\rho^{2c^2-2}}{c}\left(\frac{2c(c+1) + k(c-1)}{2c(c+1) + k(c-1)}\right). \tag{3.29}$$

This ratio must be positive. For $k = 0$,

$$c^{-1} > 0. \tag{3.30}$$

All motion is allowed for positive c; motion is impossible for negative c (negative mass density). From Eq. (3.32) below we deduce that for positive $c < 1$ all the motion is timelike; for $c > 1$, all motion is spacelike. We are not surprised at the fact that all circular geodesic motion is possible in the absence of a magnetic field. The centrifugal and gravitational influences are in equilibrium.

TABLE I. Summary of permissible motion (geodesics) for uncharged test particles for positive central mass density.

Case I: $R_0{}^0 = -R_1{}^1 = R_2{}^2 = -R_3{}^3$ ($c \sim$ positive central mass density)

ρ, ϕ fixed	ρ, z fixed	ϕ, z fixed
$k = 0, c < 0$: all trajectories forbidden	$k = 0, c > 0$: all trajectories allowed, timelike if $c < 1$ and spacelike if $c > 1$	$k = 0, c > 0$: all trajectories are allowed and may be timelike, null, or spacelike
$k > 0, c > 0$:	$k > 0, c > 0$:	$k > 0, c > 0$: all trajectories allowed and may be timelike, null, or spacelike
$0 < \rho^{2c} < k\frac{(1-c)}{(1+c)}$ forbidden if $c < 1$	$0 < \rho^{2c} < \frac{k(1-c)}{c+1}$ forbidden if $c < 1$	
$k\frac{(1-c)}{(1+c)} \leq \rho^{2c} < k\frac{2-c}{2+c}$ timelike if $c < 1$	$0 < \rho^{2c} < \frac{k(c-1)}{c+1}$ forbidden if $c > 1$	
$0 \leq \rho^{2c} < k\frac{2-c}{2+c}$ timelike if $c > 1$	$\rho^{2c} \geq \frac{k(1-c)}{c+1}$ timelike if $c < 1$	
$\rho^{2c} = k\frac{2-c}{2+c}$ null	$\rho^{2c} \geq \frac{k(c-1)}{c+1}$ spacelike if $c > 1$	
$k\frac{2-c}{2+c} < \rho^{2c} \leq k$ spacelike		
$\rho^{2c} > k$ forbidden		

Case II: $R_0{}^0 = -R_1{}^1 = -R_2{}^2 = R_3{}^3$ ($\delta = -1 - c \sim$ positive central mass density)

$k = 0, \delta > 0$: all trajectories forbidden	$k = 0, \delta > 0$: all trajectories allowed and timelike	$k \geq 0, \delta \geq 0$: all trajectories are allowed and may be timelike, null, or spacelike
$k \neq 0, \delta > 0$: $0 < \rho^{2(1+\delta)} < \frac{\delta}{k(2+\delta)}$ forbidden	$k > 0, \delta \geq 0$:	
$\rho^{2(1+\delta)} > \frac{\delta}{k(2+\delta)}$ spacelike	$0 < \rho^{2(1+\delta)} < \frac{1-\delta}{k(3+\delta)}$ timelike	
$k \neq 0, \delta = 0$: all trajectories are allowed and null	$\rho^{2(1+\delta)} = \frac{1-\delta}{k(3+\delta)}$ null	
	$\frac{1-\delta}{k(3+\delta)} < \rho^{2(1+\delta)} < k^{-1}$ spacelike	
	$k^{-1} < \rho^{2(1+\delta)}$ forbidden	

For $k \neq 0$, $E^2/J^2 > 0$ requires

$$\rho^{2c} \geq \rho_m{}^{2c} = \begin{cases} k(1-c)/(1+c), & c < 1, \\ k(c-1)/(c+1), & c > 1. \end{cases} \tag{3.31}$$

There is a region near the axis which is not allowed. From

$$1 = \left(\frac{g_{22,\rho}}{g_{00,\rho}}\frac{g_{00}}{(g_{22})^2} - \frac{1}{g_{22}}\right)J^2 = \frac{(\rho^{2c}+k)(1-c^2)}{c[\rho^{2c}(c+1)+k(c-1)]}J^2, \tag{3.32}$$

we deduce that all motion allowed by (3.30) and (3.31) is timelike for $c < 1$ and spacelike for $c > 1$. At $\rho = \rho_m$ the bracket of (3.32) is infinite and J^2 vanishes. Apparently the presence of the magnetic field produced by the axial current inhibits the gravitational attraction of a circularly moving neutral test particle to the point that at $\rho = \rho_m$ it takes no centrifugal force to maintain a balanced motion (stationary) of the test particle. Closer circular motion is impossible.

Suppose that now ϕ and z are fixed. We then have $J = 0$ and $\mathfrak{L} = 0$. From $\mathfrak{L} = 0$, we must conclude that if an axial mass and current are both present and the test particle is charged, no radial motion is possible. The conclusion follows from the ρ dependence of $\mathfrak{L}$ for $e \neq 0$. This was expected; in the presence of the magnetic field, radial motion of a charged particle would produce an acceleration in the z direction. Radial motion is possible only for a neutral particle. Equations (3.5), (3.6), and (3.9) imply

$$\frac{d\rho}{ds} = \left[-\frac{1}{g}\left(\frac{E^2}{g_{00}}+1\right)\right]^{\frac{1}{2}}. \tag{3.33}$$

Timelike motion is possible if E^2 is positive and $d\rho/ds$ is purely imaginary, $E^2 > -g_{00}$. Spacelike

TABLE II. Summary of permissible motion for charged test particles (e/m) for positive central mass density.

Case I: $R_0{}^0 = -R_1{}^1 = R_2{}^2 = -R_3{}^3$ ($c \sim$ positive central mass density)

ρ, ϕ fixed	ρ, z fixed	ϕ, z fixed
$k = 0, c > 0$: all trajectories forbidden	$k = 0, 1 > c > 0$: all trajectories allowed and timelike	$k = 0, c > 0$: all trajectories are allowed and may be timelike, null, or spacelike
$k = 0, c > 0$: no spacelike motion is possible; for any ρ, a value of e/m can be found for which timelike motion is possible	$k = 0, c > 1$: all trajectories allowed and spacelike	$k > 0, c > 0$: all trajectories forbidden
	$k > 0, c > 0$	
	$0 < \rho^{2c} < \frac{k(1-c)}{c+1}$ forbidden if $c < 1$	
	$0 < \rho^{2c} < \frac{k(c-1)}{c+1}$ forbidden if $c > 1$	
	$\rho^{2c} \geq \frac{k(1-c)}{c+1}$ timelike if $c < 1$	
	$\rho^{2c} \geq \frac{k(c-1)}{c+1}$ spacelike if $c > 1$	

Case II: $R_0{}^0 = -R_1{}^1 = -R_2{}^2 = R_3{}^3$ ($\delta = -1 - c \sim$ positive central mass density)

$k = 0, \delta > 0$: all trajectories forbidden	$k = 0, \delta > 0$: all trajectories allowed and timelike	$k = 0, \delta > 0$: all trajectories are allowed and may be timelike, null, or spacelike
$k \neq 0, \delta > 0$: $0 < \rho^{2(1+\delta)} < \frac{\delta}{k(2+\delta)}$ forbidden; $\rho^{2(1+\delta)} > \frac{\delta}{k(2+\delta)}$ spacelike	$k > 0, \delta \geq 0$: all trajectories forbidden if electric field is present; if field is purely magnetic, only timelike motion is permissible; for any ρ, a value of e/m can be found for which timelike motion is possible	$k > 0, \delta \geq 0$: all trajectories forbidden
$k \neq 0; \delta = 0$: all trajectories forbidden if electric field is present; if field is purely magnetic, all trajectories are allowed and null		

motion prevails whenever E^2 is negative. Null geodesics are characterized by

$$\frac{d\rho}{dt} = 1. \tag{3.34}$$

To see the possibilities for modification of the motion for charged test particles, look again at the situation of fixed ρ and ϕ for such particles. Instead of Eq. (3.20), Eq. (3.5) yields

$$\frac{E^2}{\mathfrak{L}^2} = -\frac{g_{33,\rho}}{g_{00,\rho}}\frac{(g_{00})^2}{(g_{33})^2} + \frac{e}{m}4c(k)^{\frac{1}{2}}e^a\frac{1}{\rho\mathfrak{L}}\frac{(g_{00})^2}{g_{00,\rho}}. \tag{3.35}$$

This shows immediately (c, k, and e all being nonzero) that spacelike motion is always impossible; the left-hand side and the first term on the right-hand side are real, and the second term on the right-hand side cannot be purely imaginary as it must be for spacelike motion ($\mathfrak{L}$ purely imaginary). In addition to (3.35), motion is allowed if (3.9) is satisfied:

$$1 = \left(-\frac{E^2}{\mathfrak{L}^2 g_{00}} - \frac{1}{g_{33}}\right)\mathfrak{L}^2, \tag{3.36}$$

for positive $\mathfrak{L}^2$. Equations (3.36) and (3.35) must be solved for E^2 and $\mathfrak{L}^2$ and must yield positive quantities for both. It is clear that for any ρ such a solution is possible by a judicious choice of the sign and magnitude of e/m. Hence motion along the z axis is possible for any radius depending on the appropriate choice of e/m. This result is not surprising, since such motion in the presence of B_ϕ produces a radial repulsion which can balance the gravitational attraction.

As expected, the qualitative nature of the motion for fixed ρ and z is independent of whether the particle is charged or not since it is moving parallel to the field. It has already been pointed out that motion with ϕ and z fixed is impossible for charged particles.

Case II: $R_0^0 = -R_1^1 = -R_2^2 = R_3^3$

For Case II, the equations of motion (3.1) become

$$\frac{d^2t}{ds^2} + \frac{g_{00,\rho}}{g_{00}}\frac{dt}{ds}\frac{d\rho}{ds} = i\frac{e}{m}\frac{dz}{ds}\frac{2c(k)^{\frac{1}{2}}\cos\theta}{g_{00}}, \tag{3.37}$$

$$\frac{d^2\phi}{ds^2} + \frac{g_{22,\rho}}{g_{22}}\frac{d\phi}{ds}\frac{d\rho}{ds} = -i\frac{e}{m}\frac{d\rho}{ds}\frac{2c(k)^{\frac{1}{2}}}{\rho}\sin\theta, \tag{3.38}$$

$$\frac{d^2z}{ds^2} + \frac{g_{33,\rho}}{g_{33}}\frac{dz}{ds}\frac{d\rho}{ds} = i\frac{e}{m}\frac{dt}{ds}\frac{2c(k)^{\frac{1}{2}}}{g_{33}}\cos\theta, \tag{3.39}$$

$$\frac{d^2\rho}{ds^2} + \frac{1}{2}\frac{g_{11,\rho}}{g_{11}}\left(\frac{d\rho}{ds}\right)^2 - \frac{1}{2}\frac{g_{00,\rho}}{g_{11}}\left(\frac{dt}{ds}\right)^2 - \frac{1}{2}\frac{g_{22,\rho}}{g_{11}}\left(\frac{d\phi}{ds}\right)^2 - \frac{1}{2}\frac{g_{33,\rho}}{g_{11}}\left(\frac{dz}{ds}\right)^2 = -i\frac{e}{m}\frac{d\phi}{ds}\frac{2c(k)^{\frac{1}{2}}}{\rho}\sin\theta\frac{g_{22}}{g_{11}}. \tag{3.40}$$

The first integrals of the first three equations are

$$g_{00}\frac{dt}{ds} = i\left(E + \frac{e}{m}Z2c(k)^{\frac{1}{2}}\cos\theta\right) = i\mathcal{E}(Z), \tag{3.41}$$

$$g_{22}\frac{d\phi}{ds} = i\left(J - \frac{e}{m(k)^{\frac{1}{2}}}\frac{\sin\theta}{(1 + k\rho^{-2c})}\psi(k, c)\right) \equiv i\mathcal{J}(\rho), \tag{3.42}$$

$$g_{33}\frac{dz}{ds} = i\left(L + \frac{e}{m}t2c(k)^{\frac{1}{2}}\cos\theta\right) \equiv i\mathcal{L}(t). \tag{3.43}$$

Equation (3.9) is also a first integral of the set of equations. If the test particle is not charged, its motion will be independent of the phase factor θ. As far as strictly gravitational effects are concerned, we cannot distinguish an electric field from a magnetic field. Again for timelike motion, $\mathcal{E}$, $\mathcal{J}$, and $\mathcal{L}$ are all real; for spacelike motion, they are all purely imaginary.

To determine the physical significance of E, J, and L, proceed as before. Expand (3.11) for small values of the quantities k, a, and v^2 and $\delta \equiv -(1 + c)$ with $e = 0$:

$$E = 1 + \delta \ln\rho + k\rho^2 + a + \tfrac{1}{2}v^2. \tag{3.44}$$

If $k = 0$, the obvious identification to make is [compare (3.13)]

$$U/M = E - a, \tag{3.45}$$

$$2m = \delta. \tag{3.46}$$

We can still imagine this interpretation to first order and consider $k\rho^2$ to be an extra gravitational potential produced by the electromagnetic field. J and L have the same interpretation as they had in Case I.

For Case II, there is the interesting possibility that $\delta = 0, k \neq 0$, and $a = 0$, which would correspond to a universe with only electromagnetic and gravitational fields present. When $\theta = \frac{1}{2}\pi$, the field is purely magnetic; geodesics for this case have been examined by Thorne[10] and in greater detail by Melvin and Wallingford.[12]

Consider now the case of neutral test particles for which ρ and ϕ are fixed; $\mathcal{J} = J = 0$, $\mathcal{L} = L$, and $\mathcal{E} = E$ and from (3.40)–(3.43)

$$\frac{\mathcal{E}^2}{\mathcal{L}^2} = -\frac{g_{33,\rho}(g_{00})^2}{g_{00,\rho}(g_{33})^2} = \rho^{2c^2-2}e^{2a}\left(\frac{\rho^{2c}(1 + c) + k(1 - c)}{c[\rho^{2c}(1 + c) + k(c - 1)]}\right). \tag{3.47}$$

For $\mathcal{E}^2$ and $\mathcal{L}^2$ to have the same sign, the bracket must be positive. For $k = 0$, the magnetic field vanishes; in terms of δ the required inequality is

$$c^{-1} = -(1 + \delta)^{-1} > 0. \tag{3.48}$$

This is the same as inequality (3.22), as it must be since we are dealing with the same physical situation. Equation (3.48) cannot be satisfied for positive δ. If $k \neq 0$, the bracket in (3.47) is positive if

$$\rho^{2(1+\delta)} > \delta/k(2 + \delta). \tag{3.49}$$

The criterion for timelike, spacelike, or null motion is given by (3.25):

$$\frac{(1 - c^2)}{c}\frac{\rho^{2c} + k}{\rho^{2c}(c + 1) + k(c - 1)} \begin{cases} > 0 & \text{timelike,} \\ < 0 & \text{spacelike,} \\ = 0 & \text{null.} \end{cases} \tag{3.50}$$

If mass is present, $\delta > 0$, all allowed motion is spacelike. If no mass is present, $\delta = 0$, all allowed motion is null, and, from (3.49), all ρ are possible.

Suppose ρ and z are fixed (neutral particles). Then $\mathcal{L} = L = 0$, $\mathcal{J} = J$, and $\mathcal{E} = E$:

$$\frac{\mathcal{E}^2}{\mathcal{J}^2} = -\frac{g_{22,\rho}(g_{00})^2}{g_{00,\rho}(g_{33})^2} = -\frac{e^{2a}\rho^{2c^2}}{(g_{22})^2}\frac{k - \rho^{2c}}{k(1 - c) - \rho^{2c}(1 + c)}. \tag{3.51}$$

The ratio is positive if

$$\rho^{2(1+\delta)} < k^{-1}. \tag{3.52}$$

The criterion for timelike, spacelike, or null motion becomes (3.25) and (3.26). (The form of g_{33} for Case I is the same as g_{22} for Case II.) We are, of course, now interested in c somewhat less than -1. For small δ, the result is

$$\begin{aligned} 0 < \rho^{2(1+\delta)} &< (1 - \delta)/k(3 + \delta) && \text{timelike,} \\ \rho^{2(1+\delta)} &= (1 - \delta)/k(3 + \delta) && \text{null,} \\ (1 - \delta)/k(3 + \delta) &< \rho^{2(1+\delta)} < k^{-1} && \text{spacelike.} \end{aligned} \tag{3.53}$$

Case II seems to contrast dramatically with Case I. For fixed ρ and z the effect of the circular magnetic field in Case I was repulsive; in Case II, the z-directed magnetic field seems to be attractive. Similarly, for fixed ρ and ϕ, in Case I the magnetic field seems to give a repulsive effect, and in Case II it seems to give an attractive effect.

The possibilities for fixed ϕ and z are the same for Cases I and II. No motion is possible if the test particle is charged. Geodesics are given by (3.33) and (3.41) with $E^2 > -g_{00}$ for timelike motion and $E^2 < g_{00}$ for spacelike. Null geodesics satisfy (3.34).

Consider the possibility of charged test particle motion for fixed ρ and z. Because of the time-dependent term in $\mathfrak{L}$, such motion is impossible unless the time-dependent term is made to vanish by the choice $\cos\theta = 0$. Hence the motion is impossible if the field has an electric component. For a purely magnetic field it can readily be shown, analogously to Case I, for fixed ρ and ϕ, that no spacelike motion is possible. For any ρ, a value of e/m can be found for which timelike motion of the charged test particle with constant ρ and z is possible.

A charged test particle moving parallel to the z axis (fixed ρ and ϕ) has the same type of motion as an uncharged test particle if the field is purely magnetic. If the field is partially electric ($\cos\theta \neq 0$), such motion is impossible because of the time dependence of $\mathfrak{L}$. Equation (3.47) cannot be satisfied for all $\mathfrak{L}$. Apparently the charged particle increases energy as it is accelerated by the electric field, and hence the gravitational attraction to the axis increases.

4. INVARIANTS

In general, from the Riemann curvature tensor and the metric tensor, 14 algebraically independent scalar invariants can be constructed. If the Riemann tensor is decomposed into what is essentially the conformal curvature tensor and the Ricci tensor, four invariants can be constructed from the conformal tensor without using the Ricci tensor, four can be constructed from the Ricci tensor without using the conformal tensor, and six require the use of both the conformal and Ricci tensors. In this section the invariants of the space-time under study are calculated. The invariants can be constructed by taking the physical components of the Riemann tensor using the orthonormal tetrad given in (2.22)–(2.25). The only nonvanishing physical components are given by the invariants

$$I_1 \equiv {}^{P}R_{0101} = e^{2\psi-2\gamma}(\psi_{\rho\rho} - \gamma_{\rho\rho}), \tag{4.1}$$

$$I_2 \equiv {}^{P}R_{0202} = e^{2\psi-2\gamma}\left(\psi_\rho\gamma_\rho - \psi_\rho{}^2 - \frac{\gamma_\rho}{\rho} + \frac{\psi_\rho}{\rho}\right), \tag{4.2}$$

$$I_3 \equiv {}^{P}R_{0303} = e^{2\psi-2\gamma}(\psi_\rho{}^2 - \gamma_\rho\psi_\rho), \tag{4.3}$$

$$I_4 \equiv {}^{P}R_{1212} = e^{2\psi-2\gamma}\left(\gamma_\rho\psi_\rho - \psi_{\rho\rho} - \frac{\gamma_\rho}{\rho} - \frac{\psi_\rho}{\rho}\right), \tag{4.4}$$

$$I_5 \equiv {}^{P}R_{1313} = e^{2\psi-2\gamma}(\psi_{\rho\rho} + 2\psi_\rho{}^2 - \gamma_\rho\psi_\rho), \tag{4.5}$$

$$I_6 \equiv {}^{P}R_{2323} = e^{2\psi-2\gamma}\left(-\psi_\rho{}^2 + \frac{\psi_\rho}{\rho}\right). \tag{4.6}$$

If the calculation of the invariants is made, it turns out that they can all be expressed in the form

$$I = ce^{-2a}[\alpha + \beta\rho^{-2c} + \gamma\rho^{-4c}]/\rho^{2c^2+2c+2}(1 + k\rho^{-2c})^4, \tag{4.7}$$

where α, β, and γ are suitably chosen constants. It is obvious that all invariants vanish for $c = 0$ and that $\exp(-2a)$ is a factor multiplying all the invariants and determines a scale of some kind. For Case I, the constants α, β, and γ are given in Table III.

It is perhaps not surprising that there are only three independent invariants because the metric contains three arbitrary constants a, c, and k. The relations between the invariants shown in the first column of Table III comes from direct calculation or by use of the equations which define $R_\mu{}^\nu$ together with the fundamental equation (2.16a). For Case I,

$$R_0{}^0 = -I_1 - I_2 - I_3, \tag{4.8}$$

$$R_0{}^0 = -R_1{}^1 = -I_1 + I_4 + I_5, \tag{4.9}$$

$$R_0{}^0 = R_2{}^2 = -I_2 + I_4 + I_6, \tag{4.10}$$

$$R_0{}^0 = -R_3{}^3 = -I_3 + I_5 + I_6. \tag{4.11}$$

An easy calculation now yields for the physical components of the Ricci tensor

$$\begin{aligned} {}^{P}R_0{}^0 &= -{}^{P}R_1{}^1 = {}^{P}R_2{}^2 = -{}^{P}R_3{}^3 \\ &= \frac{4kc^2\rho^{-2c}e^{-2a}}{\rho^{2c^2+4c+2}(1 + k\rho^{-2c})^4}. \end{aligned} \tag{4.12}$$

If $k = 0$, the electromagnetic field vanishes and the invariants become equal to those for the pure Einstein vacuum case. Equation (4.7) becomes

$$I = ce^{-2a}\alpha/\rho^{2c^2+2c+2}, \quad I_1 = -I_6, \quad I_2 = -I_5, \quad I_3 = -I_4, \quad \text{and} \quad I_1 + I_2 + I_3 = 0. \tag{4.13}$$

TABLE III. Constants determining physical components of Riemann tensor for Case I [see Eq. (4.7)].

	α	β	γ
$I_1 = -I_6$	$(c+1)$	$-2ck$	$k^2(c-1)$
I_2	$-(c+1)^2$	0	$k^2(c-1)^2$
$I_3 = -I_4$	$c(c+1)$	$-2kc$	$ck^2(1-c)$
$I_5 = 2I_1 + I_2 + 2I_3$	$(c+1)^2$	$-8kc$	$-k^2(c-1)^2$

The last relation shows that only two invariants are independent. They differ by a constant factor which can be measured at one space–time point.

The invariants for Case II can be written immediately. The metric for Cases I and II are the same except for an interchange of g_{22} and g_{33}. Hence

$$\begin{aligned} &I_1(\text{Case II}) = I_1(\text{Case I}), \quad I_2(\text{Case II}) = I_3(\text{Case I}),\\ &I_3(\text{Case II}) = I_2(\text{Case I}), \quad I_4(\text{Case II}) = I_5(\text{Case I}),\\ &I_5(\text{Case II}) = I_4(\text{Case I}), \quad I_6(\text{Case II}) = I_4(\text{Case I}). \end{aligned} \tag{4.14}$$

For Case II, it is of some interest to look at the invariants when $c = -1$ (no mass density). Equation (4.7) becomes

$$I = -e^{-2\alpha}(\beta + \gamma\rho^2)/(1 + k\rho^2)^4. \tag{4.15}$$

β and γ can both be obtained from Table III and relations (4.14) with $c = -1$. It is obvious in this case that the invariants are all regular, and well behaved, and vanish when $k = 0$.

5. STABILITY ANALYSIS

In this section the solutions for Cases I and II will be analyzed to see if they are stable when the system is subjected to radial perturbations of a certain type. In particular the functions ψ and γ will be represented by

$$\gamma(\rho, t) = \gamma(\rho) + \epsilon\alpha(\rho, t), \tag{5.1}$$

$$\psi(\rho, t) = \psi(\rho) + \epsilon\beta(\rho, t), \tag{5.2}$$

where ϵ is a small parameter and the equations will be considered to the first power of ϵ only. It will turn out that both cases are stable against perturbations of this type in that no solutions with positive or zero mass density exist which grow exponentially with time. For both cases, undamped, oscillatory solutions exist.

When γ and ψ depend on both ρ and t, the non-vanishing components of the Ricci tensor are given by

$$R_0{}^0 = e^{2\psi-2\gamma}\Big(\gamma_{\rho\rho} - \psi_{\rho\rho} + \psi_{tt} - \gamma_{tt} - 2\psi_t{}^2 - \frac{\psi_\rho}{\rho} + \frac{\gamma_\rho}{\rho}\Big), \tag{5.3}$$

$$R_1{}^1 = e^{2\psi-2\gamma}\Big(\gamma_{\rho\rho} - \psi_{\rho\rho} + \psi_{tt} - \gamma_{tt} + 2\psi_\rho{}^2 - \frac{\psi_\rho}{\rho} - \frac{\gamma_\rho}{\rho}\Big), \tag{5.4}$$

$$R_2{}^2 = -R_3{}^3 = e^{2\psi-2\gamma}\Big(\psi_{tt} - \psi_{\rho\rho} - \frac{\psi_\rho}{\rho}\Big), \tag{5.5}$$

$$R_0{}^1 = -R_1{}^0 = e^{2\psi-2\gamma}\Big(2\psi_\rho\psi_t - \frac{\gamma_t}{\rho}\Big). \tag{5.6}$$

Of course, $R_\mu{}^\nu$ can be expanded in powers of ϵ:

$$R_\mu{}^\nu(\rho, t) = R_\mu{}^\nu(\rho) + \epsilon\bar{R}_\mu{}^\nu(\rho, t). \tag{5.7}$$

$R_\mu{}^\nu(\rho)$ is obtained by substituting $\gamma(\rho)$ and $\psi(\rho)$ into (5.3) to (5.6). $\bar{R}_\mu{}^\nu$ can be calculated:

$$\bar{R}_0{}^0 = e^{2\psi-2\gamma}\Big(\alpha_{\rho\rho} - \beta_{\rho\rho} + \beta_{tt} - \alpha_{tt} - \frac{\beta_\rho}{\rho} + \frac{\alpha_\rho}{\rho}\Big) + [1 + 2(\beta - \alpha)]R_0{}^0, \tag{5.8}$$

$$\bar{R}_1{}^1 = e^{2\psi-2\gamma}\Big(\alpha_{\rho\rho} - \beta_{\rho\rho} + \beta_{tt} - \alpha_{tt} + 4\psi_\rho\beta_\rho - \frac{\beta_\rho}{\rho} - \frac{\alpha_\rho}{\rho}\Big) + [1 + 2(\beta - \alpha)]R_1{}^1, \tag{5.9}$$

$$\bar{R}_2{}^2 = -\bar{R}_3{}^3 = e^{2\psi-2\gamma}\Big(\beta_{tt} - \beta_{\rho\rho} - \frac{\beta_\rho}{\rho}\Big) + [1 + 2(\beta - \alpha)]R_2{}^2, \tag{5.10}$$

$$\bar{R}_0{}^1 = -\bar{R}_1{}^0 = e^{2\psi-2\gamma}\Big(2\psi_\rho\beta_t - \frac{\alpha_t}{\rho}\Big). \tag{5.11}$$

In the above expressions ψ and γ are the unperturbed functions of ρ alone.

$R_\mu{}^\nu(\rho, t)$ must satisfy the Rainich equations to the first order in ϵ. The symmetry of the problem assures that $\alpha_\mu \equiv 0$. The vanishing of the trace of the Ricci tensor requires that

$$\bar{R}_0{}^0 = -\bar{R}_1{}^1. \tag{5.12}$$

Equation (2.7) is

$$(R_0{}^0)^2 - (R_0{}^1)^2 = \tfrac{1}{4}[(R_0{}^0)^2 - (R_0{}^1)^2 + (R_1{}^1)^2 + (R_2{}^2)^2 + (R_3{}^3)^2], \tag{5.13}$$

and to first order

$$R_0{}^0\bar{R}_0{}^0 = R_2{}^2\bar{R}_2{}^2. \tag{5.14}$$

Hence for Case I

$$\bar{R}_0{}^0 = \bar{R}_2{}^2, \tag{5.15}$$

and for Case II

$$\bar{R}_0{}^0 = -\bar{R}_2{}^2. \tag{5.16}$$

The linearized equations to be solved become the following (in terms of α and β after substituting for ψ_ρ from the unperturbed solution):

Case I:

$$\alpha_{tt} - \alpha_{\rho\rho} + \beta_{\rho\rho} - \beta_{tt} + \beta_\rho\Big(\frac{2c + 1}{\rho} - \frac{4kc\rho^{-2c-1}}{1 + k\rho^{-2c}}\Big) = 0, \tag{5.17}$$

$$\alpha_{\rho\rho} - \alpha_{tt} + \frac{\alpha_\rho}{\rho} = 0; \tag{5.18}$$

Case II:

$$\alpha_{tt} - \alpha_{\rho\rho} + \beta_{\rho\rho} - \beta_{tt} - \beta_\rho\left(\frac{2c+1}{\rho}\frac{4kc\rho^{-2c-1}}{1+k\rho^{-2c}}\right) = 0, \quad (5.19)$$

$$\alpha_{tt} - \alpha_{\rho\rho} - \frac{\alpha_\rho}{\rho} - 2\beta_{tt} + 2\beta_{\rho\rho} + 2\frac{\beta_\rho}{\rho} = 0. \quad (5.20)$$

The boundary conditions must now be specified.[13] It is perhaps a good idea to specify them in terms of invariant quantities so that the results will be independent of the particular coordinate system being used. One can choose local invariants and demand that nowhere does any component of the physical perturbed Ricci tensor [with tetrad of Eqs. (2.22)–(2.25)] diverge more quickly than the most divergent of the unperturbed physical components of the Ricci tensor (actually the nonvanishing components equal each other). This condition makes sense considering the physical problem we are dealing with in that the physically important electromagnetic field is expressed in terms of the components of the Ricci tensor. In general, the boundary conditions mean that α and β are not permitted to diverge faster than γ and ψ, respectively.

Another sensible invariant to look at is a nonlocal invariant. It is the ratio of the circumference to the radius of a circle surrounding the z axis in the (ρ, ϕ) hyperplane in the limit that the radius goes to zero:

$$\kappa(\rho, t) = \frac{\int_0^{2\pi}(-g_{22})^{\frac{1}{2}}\,d\phi}{\int_0^{\rho}(-g_{11})^{\frac{1}{2}}\,d\rho} = 2\pi e^{-\gamma(0,t)}. \quad (5.21)$$

To first order in ϵ, the change of this ratio for the perturbed solutions can be characterized by

$$\kappa(\rho, t)/\kappa(\rho, 0) = e^{-\epsilon\alpha(0,t)} = 1 - \epsilon\alpha(0, t). \quad (5.22)$$

A reasonable boundary condition would be to demand that $\alpha(0, t)$ vanish for all t in order to assure that the nature of the singularity along the axis does not change.

Case I: $\bar{R}_0^0 = -\bar{R}_1^1$, $\bar{R}_0^0 = \bar{R}_2^2$

Consider solutions of the first-order equations of the form

$$\alpha(\rho, t) = \alpha(\rho)\sinh nt, \quad (5.23)$$

$$\beta(\rho, t) = \beta(\rho)\sinh nt. \quad (5.24)$$

If they exist and the constant n has a positive real part, the unperturbed solution is an unstable one since the perturbation vanishes at zero time but grows exponentially with time when the time is small. A small initial perturbation will grow at an exponential rate.

The general solution of Eq. (5.18) is

$$\alpha = FK_0(n\rho)\sinh nt + HI_0(n\rho)\sinh nt \quad (5.25)$$

when F and H are constants and K_0 and I_0 are modified Bessel functions. Since I_0 diverges exponentially for $\rho \to \infty$, the constant $H = 0$.

Making the substitution

$$\beta(\rho) = [\rho^{-c}/(1 + k\rho^{-2c})]B(\rho) \quad (5.26)$$

and considering (5.24) and (5.25) with $H = 0$, we see that Eq. (5.17) becomes

$$\frac{d}{d\rho}(\rho B_\rho) - \left(n^2\rho + \frac{c^2}{\rho}\right)B = nF\frac{1 + k\rho^{-2c}}{\rho^{-c}}K_1(n\rho). \quad (5.27)$$

The general solution of this equation is

$$B = PK_c(n\rho) + QI_c(n\rho) + F\int_0^\infty G(n\rho, n\rho')\frac{(1 + k\rho'^{-2c})}{\rho'^{-c}}K_1(n\rho')\,d\rho'. \quad (5.28)$$

P, Q, and F are constants, and the Green's function is given by

$$G(n\rho, n\rho') = I_c(n\rho')K_c(n\rho), \quad \rho' < \rho,$$
$$= K_c(n\rho')I_c(n\rho), \quad \rho' > \rho. \quad (5.29)$$

As $\rho \to 0$, the requirement that β diverge no worse than ψ (i.e., $\ln\rho$) demands $P = 0$. The analogous requirement for $\rho \to \infty$ demands $Q = 0$. The integral in (5.28) diverges; hence $F = 0$. We conclude that no unstable solution of the type being sought exists for Case I.

In considering the general solution the possibility of nontrivial homogeneous solutions of the homogeneous equation existing and of the inhomogeneous equation existing were considered at the same time. Usually when a nontrivial solution exists satisfying the boundary conditions, there is no Green's function.

It is of some interest to look for the possibility of undamped oscillatory solutions where

$$\alpha(\rho, t) = \alpha(\rho)\sin nt, \quad (5.30)$$

$$\beta(\rho, t) = \beta(\rho)\sin nt. \quad (5.31)$$

The most general solution of Eqs. (5.18) and (5.17) with the substitution (5.26) is

$$\alpha(\rho) = FJ_0(n\rho) + HY_0(n\rho),$$
$$B(\rho) = PJ_c(n\rho) + QY_c(n\rho) + \int_0^\infty G(n\rho, n\rho')\frac{1 + k\rho'^{-2c}}{\rho'^{-2c}} \times [FJ_1(n\rho') + HY_1(n\rho')]\,d\rho', \quad (5.32)$$

where the Green's function is given by

$$G(n\rho, n\rho') = \tfrac{1}{2}\pi J_c(n\rho) Y_c(n\rho'), \quad \rho < \rho', \\ = \tfrac{1}{2}\pi Y_c(n\rho) J_c(n\rho'), \quad \rho' < \rho. \tag{5.33}$$

Again this examines the possibility of a nontrivial solution to the homogeneous problem and a Green's function at the same time.

Consider the behavior as $\rho \to 0$. In this limit, $Y_c \to \rho^{-c}$ and $\beta \to \rho^{-2c}$. This diverges faster than ψ; hence the coefficient of Y_c must vanish. Moreover, F and H both must vanish since the integrals diverge. An oscillatory solution is possible where

$$\alpha(\rho) = 0,$$

$$\beta(\rho, t) = P \frac{\rho^{-c}}{1 + k\rho^{-2c}} J_c(n\rho) \sin nt. \tag{5.34}$$

It should be observed that this oscillatory solution does not at any time change the limiting ratio of the circumference to the radius of a circle [Eq. (5.22)].

Case II: $\bar{R}_0^0 = -\bar{R}_1^1$, $\bar{R}_0^0 = -\bar{R}_2^2$

With the substitutions

$$\beta = \tfrac{1}{2}\alpha + f, \tag{5.35}$$

$$\alpha = [\rho^{-c}/(1 + k\rho^{-2c})]A, \tag{5.36}$$

Eqs. (5.19) and (5.20) become

$$f_{\rho\rho} - f_{tt} + \frac{f_\rho}{\rho} = 0, \tag{5.37}$$

$$A_{\rho\rho} - A_{tt} + \frac{A_\rho}{\rho} - \frac{c^2}{\rho^2} A \\ = -4[(c + 1)\rho^{c-1} + (1 - c)k\rho^{-c-1}]f_\rho. \tag{5.38}$$

Solutions of these equations are given by

$$f = FK_0(n\rho) \sinh nt, \tag{5.39}$$

$$A = QK_{|c|}(n\rho) \sinh nt \\ + \sinh nt\, F \int_0^\infty G(n\rho, n\rho') \\ \times [(c + 1)\rho'^c + (1 - c)k\rho'^{-c}]K_1(n\rho')\, d\rho'. \tag{5.40}$$

The Green's function is

$$G(n\rho, n\rho') = K_{|c|}(n\rho')I_{|c|}(n\rho), \quad \rho < \rho', \\ = I_{|c|}(n\rho')K_{|c|}(n\rho), \quad \rho' < \rho. \tag{5.41}$$

Terms in the solutions involving $I_0(n\rho)$ and $I_{|c|}(n\rho)$ have not been included because they diverge too fast at infinity. The integral in (5.40) diverges requiring $F = 0$ and leaving as a possibly unstable solution

$$2\beta = \alpha = Q \frac{\rho^{-c}}{1 + k\rho^{-2c}} K_{|c|}(n\rho) \sinh nt. \tag{5.42}$$

α and β are well behaved both at $\rho = 0$ and $\rho = \infty$ when c is negative.

It is necessary to compare the physical components of the perturbed Ricci tensor with the unperturbed physical components. If a perturbed physical component of the Ricci tensor diverges anywhere, it must do so more slowly than some component of the unperturbed Ricci tensor. All physical components are well behaved in this sense except

$$^P\bar{R}_1^0 = e^{2\psi - 2\gamma}\left(2\psi_\rho\beta_t - \frac{\alpha_t}{\rho}\right) \\ = \frac{Qne^{-2\alpha}\cosh nt}{\rho^{2c^2+2c+1}(1 + k\rho^{-2c})^4} \\ \times (2c + 1 - kc\rho^{-2c})\rho^{-c}K_{|c|}(n\rho). \tag{5.43}$$

This is to be compared with $^PR_0^0$ [Eq. (4.12)]. For small ρ, the ratio

$$^P\bar{R}_1^0/^PR_0^0 \sim \rho^{1+2c}. \tag{5.44}$$

The ratio diverges if $c < -\frac{1}{2}$ and converges if $c > -\frac{1}{2}$. Hence, if $c < -\frac{1}{2}$, the boundary condition requires Q to vanish, and the system is stable. For $c > -\frac{1}{2}$ the system is unstable. $c > -\frac{1}{2}$ corresponds, of course, to an enormous negative mass density along the central axis.

If the demand had been that $\alpha(0, t)$ vanish as a consequence of the requirement that the limiting ratio of circumference to radius of a small circle not be changed by the perturbation [Eq. (5.22)], Q would have to vanish for all c, since $\alpha(0, t)$, although finite, is not equal to zero. Hence, the system would be stable for all c. With either boundary condition, all systems with positive masses and even with small (not huge) negative mass densities are predicted to be stable. It has been shown[14] for the case of vanishing central mass density that the stability against radial perturbation guarantees stability against nonradial perturbation as well. This can be generalized in an obvious way to show that the stability against nonradial perturbation would hold even if mass were present.

The possibility of undamped oscillatory perturbed solutions for Case II will be demonstrated by exhibiting a solution describing such behavior:

$$2\beta = \alpha = Q \frac{\rho^{-c}}{1 + k\rho^{-2c}} J_{|c|}(n\rho) \sin nt. \tag{5.45}$$

This satisfies the appropriate equations and all boundary conditions including the requirement $\alpha(0, t) = 0$. It can, in fact, be shown to be the most general solution satisfying either the boundary conditions on the physical components of the Ricci tensor or the condition requiring $\alpha(0, t) = 0$.

[1] We use the signature $(-, +, +, +)$:

$$\Gamma^{\sigma}_{\mu\nu} \equiv \tfrac{1}{2}g^{\sigma\tau}(g_{\mu\tau,\nu} + g_{\nu\tau,\mu} - g_{\mu\nu,\tau}),$$
$$R_{\mu\nu\sigma}{}^{\tau} \equiv \Gamma_{\nu\sigma,}{}^{\tau}{}_{\mu} - \Gamma_{\mu\sigma,}{}^{\tau}{}_{\nu} + \Gamma_{\nu\sigma}{}^{\rho}\Gamma_{\mu\rho}{}^{\tau} - \Gamma_{\mu\sigma}{}^{\rho}\Gamma_{\nu\rho}{}^{\tau},$$
$$R_{\mu\nu} \equiv R_{\mu\tau\nu}{}^{\tau}.$$

A comma denotes ordinary differentiation; a semicolon denotes covariant differentiation; $16\pi G = 1$, $c = 1$, G is the gravitational constant, and c is the velocity of light. All other notation is defined in the text or is standard.

[2] For a review of the geometric theory, see *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 9.

[3] J. L. Synge, *Relativity: The General Theory* (Interscience, New York, 1960), Chap. VIII; K. S. Thorne, Ph.D. thesis, University Microfilms, Ann Arbor, 1965.

[4] The general solutions to Cases I and II of the possible field configurations and a particular solution of Case III is given in Ref. 2 and also in a paper by L. Witten, Colloq. Theor. Relativity, Centre Belge Rech. Math., Univ. (Louvain, Belgium), p. 59, 1960.

[5] This differs slightly from the line element given in Ref. 4. $\mu \equiv 0$ requires that we take the limit $l \to \infty$ in the line elements of Ref. 4. The necessity that l approach ∞ can be argued for on a variety of grounds, which we shall not present here.

[6] L. Marder, Proc. Roy. Soc. (London) **A244**, 524 (1958).
[7] M. A. Melvin, Phys. Letters **8**, 65 (1964).
[8] M. A. Melvin, Phys. Rev. **139B**, 225 (1965).
[9] K. S. Thorne, Phys. Rev. **138B**, 251 (1965).
[10] K. S. Thorne, Phys. Rev. **139B**, 244 (1965).
[11] J. L. Safko, Ann. Phys. (N.Y.) **58**, 352 (1970).
[12] M. A. Melvin and J. S. Wallingford, J. Math. Phys. **7**, 333 (1966).
[13] J. L. Safko, Bull. Am. Phys. Soc. **13**, 1725 (1968).
[14] J. L. Safko, Phys. Letters **28A**, 347 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Hypergeometric Functions with Integral Parameter Differences

PER W. KARLSSON
Department of Physics, The Technical University, Lyngby, Denmark

(Received 20 July 1970)

For a generalized hypergeometric function ${}_pF_q(z)$ with positive integral differences between certain numerator and denominator parameters, a formula expressing the ${}_pF_q(z)$ as a finite sum of lower-order functions is proved. From this formula, Minton's two summation theorems for $p = q + 1$, $z = 1$ are deduced, one of these under less restrictive conditions than assumed by Minton.

This paper deals with generalized hypergeometric functions ${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z)$ having the special property that, with suitable enumeration of parameters, $a_i = b_i + m_i$, $i = 1, 2, \cdots, n$, where $m_1, \cdots, m_n$ are positive integers and $n \leq \min(p, q)$. It is assumed that $p \leq q + 1$ and that no denominator parameter b is a negative integer or zero. A function of this type may be expressed as a finite sum of ${}_{p-n}F_{q-n}$ functions in the following way:

$${}_pF_q\left[\begin{matrix} b_1 + m_1, \cdots, b_n + m_n, a_{n+1}, \cdots, a_p; \\ b_1, \cdots, b_n, b_{n+1}, \cdots, b_q \end{matrix} z\right] = \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} A(j_1, \cdots, j_n) z^{J_n} {}_{p-n}F_{q-n}\left[\begin{matrix} a_{n+1} + J_n, \cdots, a_p + J_n; \\ b_{n+1} + J_n, \cdots, b_q + J_n \end{matrix} z\right], \quad (1)$$

where

$$J_n = j_1 + \cdots + j_n, \quad (2)$$

$$A(j_1, \cdots, j_n) = \binom{m_1}{j_1} \cdots \binom{m_n}{j_n} \frac{(b_2 + m_2)_{J_1}(b_3 + m_3)_{J_2} \cdots (b_n + m_n)_{J_{n-1}}(a_{n+1})_{J_n} \cdots (a_p)_{J_n}}{(b_1)_{J_1}(b_2)_{J_2} \cdots (b_n)_{J_n}(b_{n+1})_{J_n} \cdots (b_q)_{J_n}}, \quad (3)$$

and

$$(c)_r = \Gamma(c + r)/\Gamma(c). \quad (4)$$

By the principle of analytical continuation, Eq. (1) is valid whenever the functions involved are all analytic; restrictions upon the parameters imposed in the proof may thus be removed.

The proof is based upon an Eulerian integral representation given by Erdélyi,[1] viz.,

$${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z) = \frac{\Gamma(b_1)\Gamma(1 - b_1 + a_1)}{\Gamma(a_1)\exp(i\pi(b_1 - a_1))} \frac{i}{2\pi} \times \int_0^{(1+)} (1 - t)^{b_1 - a_1 - 1} f(t)\, dt, \quad (5)$$

where

$$f(t) = t^{a_1 - 1}\, {}_{p-1}F_{q-1}(a_2, \cdots, a_p; b_2, \cdots, b_q; zt), \quad (6)$$

[1] We use the signature $(-, +, +, +)$:

$$\Gamma^{\sigma}_{\mu\nu} \equiv \tfrac{1}{2}g^{\sigma\tau}(g_{\mu\tau,\nu} + g_{\nu\tau,\mu} - g_{\mu\nu,\tau}),$$
$$R_{\mu\nu\sigma}{}^{\tau} \equiv \Gamma_{\nu\sigma,}{}^{\tau}{}_{\mu} - \Gamma_{\mu\sigma,}{}^{\tau}{}_{\nu} + \Gamma_{\nu\sigma}{}^{\rho}\Gamma_{\mu\rho}{}^{\tau} - \Gamma_{\mu\sigma}{}^{\rho}\Gamma_{\nu\rho}{}^{\tau},$$
$$R_{\mu\nu} \equiv R_{\mu\tau\nu}{}^{\tau}.$$

A comma denotes ordinary differentiation; a semicolon denotes covariant differentiation; $16\pi G = 1$, $c = 1$, G is the gravitational constant, and c is the velocity of light. All other notation is defined in the text or is standard.

[2] For a review of the geometric theory, see *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 9.

[3] J. L. Synge, *Relativity: The General Theory* (Interscience, New York, 1960), Chap. VIII; K. S. Thorne, Ph.D. thesis, University Microfilms, Ann Arbor, 1965.

[4] The general solutions to Cases I and II of the possible field configurations and a particular solution of Case III is given in Ref. 2 and also in a paper by L. Witten, Colloq. Theor. Relativity, Centre Belge Rech. Math., Univ. (Louvain, Belgium), p. 59, 1960.

[5] This differs slightly from the line element given in Ref. 4. $\mu \equiv 0$ requires that we take the limit $l \to \infty$ in the line elements of Ref. 4. The necessity that l approach ∞ can be argued for on a variety of grounds, which we shall not present here.

[6] L. Marder, Proc. Roy. Soc. (London) **A244**, 524 (1958).
[7] M. A. Melvin, Phys. Letters **8**, 65 (1964).
[8] M. A. Melvin, Phys. Rev. **139B**, 225 (1965).
[9] K. S. Thorne, Phys. Rev. **138B**, 251 (1965).
[10] K. S. Thorne, Phys. Rev. **139B**, 244 (1965).
[11] J. L. Safko, Ann. Phys. (N.Y.) **58**, 352 (1970).
[12] M. A. Melvin and J. S. Wallingford, J. Math. Phys. **7**, 333 (1966).
[13] J. L. Safko, Bull. Am. Phys. Soc. **13**, 1725 (1968).
[14] J. L. Safko, Phys. Letters **28A**, 347 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Hypergeometric Functions with Integral Parameter Differences

PER W. KARLSSON
Department of Physics, The Technical University, Lyngby, Denmark

(Received 20 July 1970)

For a generalized hypergeometric function ${}_pF_q(z)$ with positive integral differences between certain numerator and denominator parameters, a formula expressing the ${}_pF_q(z)$ as a finite sum of lower-order functions is proved. From this formula, Minton's two summation theorems for $p = q + 1$, $z = 1$ are deduced, one of these under less restrictive conditions than assumed by Minton.

This paper deals with generalized hypergeometric functions ${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z)$ having the special property that, with suitable enumeration of parameters, $a_i = b_i + m_i$, $i = 1, 2, \cdots, n$, where $m_1, \cdots, m_n$ are positive integers and $n \leq \min(p, q)$. It is assumed that $p \leq q + 1$ and that no denominator parameter b is a negative integer or zero. A function of this type may be expressed as a finite sum of ${}_{p-n}F_{q-n}$ functions in the following way:

$${}_pF_q\left[\begin{matrix} b_1 + m_1, \cdots, b_n + m_n, a_{n+1}, \cdots, a_p; \\ b_1, \cdots, b_n, b_{n+1}, \cdots, b_q \end{matrix} z\right] = \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} A(j_1, \cdots, j_n) z^{J_n} {}_{p-n}F_{q-n}\left[\begin{matrix} a_{n+1} + J_n, \cdots, a_p + J_n; \\ b_{n+1} + J_n, \cdots, b_q + J_n \end{matrix} z\right], \quad (1)$$

where

$$J_n = j_1 + \cdots + j_n, \quad (2)$$

$$A(j_1, \cdots, j_n) = \binom{m_1}{j_1} \cdots \binom{m_n}{j_n} \frac{(b_2 + m_2)_{J_1}(b_3 + m_3)_{J_2} \cdots (b_n + m_n)_{J_{n-1}}(a_{n+1})_{J_n} \cdots (a_p)_{J_n}}{(b_1)_{J_1}(b_2)_{J_2} \cdots (b_n)_{J_n}(b_{n+1})_{J_n} \cdots (b_q)_{J_n}}, \quad (3)$$

and

$$(c)_r = \Gamma(c + r)/\Gamma(c). \quad (4)$$

By the principle of analytical continuation, Eq. (1) is valid whenever the functions involved are all analytic; restrictions upon the parameters imposed in the proof may thus be removed.

The proof is based upon an Eulerian integral representation given by Erdélyi,[1] viz.,

$${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z) = \frac{\Gamma(b_1)\Gamma(1 - b_1 + a_1)}{\Gamma(a_1)\exp(i\pi(b_1 - a_1))} \frac{i}{2\pi} \times \int_0^{(1+)} (1 - t)^{b_1 - a_1 - 1} f(t)\, dt, \quad (5)$$

where

$$f(t) = t^{a_1 - 1}\, {}_{p-1}F_{q-1}(a_2, \cdots, a_p; b_2, \cdots, b_q; zt), \quad (6)$$

valid when $\mathrm{Re}\, a_1 > 0$, b_1 is not a negative integer or zero, and $|\arg(1-z)| < \pi$ if $p = q+1$. Now, as $a_1 = b_1 + m_1$, the branch point of the integrand at $t = 1$ disappears, and the integral takes the form $\int_C f(t)\,dt/(t-1)^{m_1+1}$, where C is a closed contour encircling the point $t = 1$ counterclockwise and f is analytic within and on C. From Cauchy's integral formula, we then find that Eq. (5) becomes

$$ {}_pF_q\left[\begin{matrix} b_1 + m_1, a_2, \cdots, a_p; \\ b_1, b_2, \cdots, b_q \end{matrix} z\right] = \frac{D^{m_1} f(1)}{(b_1)_{m_1}}, \quad (7) $$

D denoting differentiation with respect to t. Application of Leibniz's differentiation formula and the well-known expression for the derivative of a ${}_pF_q$ then yields

$$ {}_pF_q\left[\begin{matrix} b_1 + m_1, a_2, \cdots, a_p; \\ b_1, b_2, \cdots, b_q \end{matrix} z\right] = \sum_{j=0}^{m_1} z^j \binom{m_1}{j} \frac{(a_2)_j \cdots (a_p)_j}{(b_1)_j \cdots (b_q)_j} \times {}_{p-1}F_{q-1}\left[\begin{matrix} a_2 + j, \cdots, a_p + j; \\ b_2 + j, \cdots, b_q + j \end{matrix} z\right]. \quad (8) $$

This result can itself be applied to each member of its rhs if $a_2 = b_2 + m_2$, etc. It is easily seen that the general result (1) is obtained in this way.

A special case ($p = 3$, $q = 2$) of Eq. (8) has been derived by Rösler[2] from the series representation.

From formula (1) we now derive two summation theorems for $p = q + 1$, $z = 1$. These have been given recently by Minton,[3] the first one, however, under more restrictive conditions than those given below.

To deduce the first theorem, we take $q = p - 1 = n + 1$, $a_{n+1} = b = b_{n+1} - 1$, $z = 1$, and for brevity $a_{n+2} = a$. The hypergeometric functions in the multiple sum of Eq. (1) then become ${}_2F_1(1)$'s, which all exist provided that

$$ \mathrm{Re}\,(-a) > m_1 + \cdots + m_n - 1. \quad (9) $$

By Gauss' summation theorem we then get, after some rearrangements,

$$ {}_{n+2}F_{n+1}(b_1 + m_1, \cdots, b_n + m_n, b, a; b_1, \cdots, b_n, b+1; 1) = \frac{\Gamma(b+1)\Gamma(1-a)}{\Gamma(b+1-a)} \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} B_n(j_1, \cdots, j_n), $$

where

$$ B_n(j_1, \cdots, j_n) = \frac{(b+1)_{J_n}}{(a)_{J_n}} (-1)^{J_n} A(j_1, \cdots, j_n). $$

Next, the definition (3) is applied, the binomial coefficients being written in the form $(-1)^j(-m)_j/j!$; this leads to

$$ B_n(j_1, \cdots, j_n) = B_{n-1}(j_1, \cdots, j_{n-1}) \times \frac{(b_n + m_n)_{J_{n-1}}}{(b_n)_{J_{n-1}}} \frac{(-m_n)_{j_n} (b + J_{n-1})_{j_n}}{j_n!\,(b_n + J_{n-1})_{j_n}}. $$

The terms containing j_n obviously constitute a terminating ${}_2F_1(1)$, which is summed by Gauss' theorem. After some rearrangements we obtain (summation limits understood)

$$ \sum_{j_1, \cdots, j_n} B_n(j_1, \cdots, j_n) = \frac{(b_n + m_n)_{-b}}{(b_n)_{-b}} \sum_{j_1, \cdots, j_{n-1}} B_{n-1}(j_1, \cdots, j_{n-1}). $$

Repeating this procedure, we finally arrive at Minton's first theorem,

$$ {}_{n+2}F_{n+1}(b_1 + m_1, \cdots, b_n + m_n, b, a; b_1, \cdots, b_n, b+1; 1) = \frac{\Gamma(b+1)\Gamma(1-a)}{\Gamma(b+1-a)} \prod_{k=1}^{n} \frac{(b_k + m_k)_{-b}}{(b_k)_{-b}}, \quad (10) $$

valid under the condition (9), i.e., if the lhs of (10) exists at all. In Minton's proof,[3] a was required to be negative integral.

The particular case $n = 1$ of Eq. (10) was obtained by Mitra[4] by series manipulations.

The second summation theorem may be deduced from the first[3] by letting $b \to \infty$. It may, however, also be deduced directly from Eq. (1) by taking $q = p - 1 = n$, $a_{n+1} = -(m_1 + \cdots + m_n)$, and $z \to 1$. The hypergeometric functions in the multiple sum of Eq. (1) then reduce to power functions $(1-z)^h$, where $h = -a_{n+1} - J_n$ and $h \geq 0$ for all terms. When $z \to 1$, all terms of the multiple sum will thus tend to zero, except the one for which $-a_{n+1} = J_n$, i.e., $j_i = m_i$, $i = 1, 2, \cdots, n$; the limit of this term is $A(m_1, \cdots, m_n)$. After some reductions we find the summation formula

$$ {}_{n+1}F_n(b_1 + m_1, \cdots, b_n + m_n, -(m_1 + \cdots + m_n); b_1, \cdots, b_n; 1) = \frac{(-1)^{m_1 + \cdots + m_n}(m_1 + \cdots + m_n)!}{(b_1)_{m_1} \cdots (b_n)_{m_n}}, \quad (11) $$

which is easily transformed to the form given by Minton.[3]

It may be of interest to compare Eq. (11) with the special case of (10) obtained by taking $b_n = b$, $m_n = 1$ and then replacing $n - 1$ by n, viz.,

$$ {}_{n+1}F_n(b_1 + m_1, \cdots, b_n + m_n, a; b_1, \cdots, b_n; 1) = 0, \quad \mathrm{Re}\,(-a) > m_1 + \cdots + m_n. \quad (12) $$

[1] A. Erdélyi, Quart. J. Math. Oxford Ser. **8**, 267 (1937).
[2] R. Rösler, Z. Angew. Math. Mech. **43**, 433 (1963).
[3] B. M. Minton, J. Math. Phys. **11**, 1375 (1970).
[4] S. C. Mitra, J. Indian Math. Soc. (N.S.) **6**, 84 (1942).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Generalized Master Equations for Inhomogeneous Systems*

JAMES W. DUFTY
Department of Physics and Astronomy, University of Florida, Gainesville, Florida 32601

(Received 20 May 1970)

A generalized master equation of the Van Hove type is derived for inhomogeneous systems with arbitrary initial conditions in such a way as to display explicitly the renormalized "free" propagation, as well as the "gain–loss" terms. The corresponding Wigner function formulation of this equation is given as a representation appropriate for the study of kinetic equations for inhomogeneous systems from the point of view of master equations. The results are formally exact.

1. INTRODUCTION

The Heisenberg equations of motion for an operator, or the von Neumann equation for the density matrix, are seldom the most convenient form in which to express the time dependence of the quantities of interest in many-body systems. A number of equations, formally equivalent to the above, have been derived on the basis of rearranging a formal solution to the Heisenberg equations to group relevant terms.[1] The point of these manipulations is that a rather simple approximation to this new form yields solutions which would be difficult to obtain directly otherwise; they typically have a close similarity to the Pauli master equation.[2] The generalized master equation (GME) of Van Hove[3] is an exact rate equation for the spectral decomposition of the matrix elements of the two time-development operators, $[U^*(t)]_{ij}[U(t)]_{kn}$, in a representation which diagonalizes the unperturbed Hamiltonian. It was originally derived for the occupation density $[U^*(t)]_{ij}[U(t)]_{ji}$, and was extended by Janner[4] and Swenson[5] to include the interference term $[U^*(t)]_{ij}[U(t)]_{jn}$. These latter results are sufficient to determine the time development of an operator diagonal in the unperturbed representation.

Here, the systems of interest are assumed to be inhomogeneous, which means the operators whose time dependence is desired (in particular, the density matrix) are nondiagonal. The full four-index quantity $U^*_{ij}U_{kn}$ is required in this case, and certain questions arise in the derivation of a corresponding GME. The first concerns uniqueness. Three different GME's were derived by Peterson and Quay,[6] and it is shown below (Sec. 4) that there are in fact an infinite number of GME's possible. Therefore, as observed in Ref. 6, additional constraints are needed to fix the most appropriate form for the problems of statistical mechanics. The second problem is to obtain a form which demonstrates explicitly the contribution from unperturbed propagation or free streaming. This contribution vanishes for the case of diagonal operators, but is nonzero for inhomogeneous cases (although the forms obtained in Ref. 6 contain this only implicitly in the inhomogeneous term). Since all kinetic equations, for which the GME may be considered a prototype, have an explicit streaming term, it is clearly desirable to write a GME in this form. Finally, the effects of the interaction should be identified as contributing either to the "gain–loss" term responsible for damping or to a renormalization of the energies associated with propagation.

The problem of characterizing the effects of the interaction has been solved, at least formally, by Van Hove for systems satisfying his "diagonal singularity" conditions.[3] He first considered self-energy, or cloud effects of field theories with no metastable states,[7] and obtained, in addition to an explanation of how these effects can arise, exact expressions for the renormalized energies of propagation. The subsequent derivation of a GME for dissipative systems dealt predominantly with the damping phenomena, and, although energy renormalization effects still occur, consideration of diagonal operators did not give rise to the renormalized propagation term. Here, a GME for the four-index quantity $U^*_{ij}U_{kn}$ will be obtained which explicitly incorporates the renormalized energies in a streaming term and a generalization of the gain–loss term. The main difference in the gain–loss term comes about from the need to extend Van Hove's prescription for finding the important terms in the thermodynamic limit to apply for nondiagonal operators. The result is a different definition of the "generalized transition probabilities," which reduces to the usual definition for homogeneous systems. This affects only the gain term. The loss terms arise from collecting terms in the same U, and the nondiagonality is unimportant.

In obtaining this form for the GME certain aspects of the lack of uniqueness become clearer. Peterson and Quay's equations differ from that given here in two important ways. First, the contributions from propagation are contained only implicitly, at least in part, in their inhomogeneous terms. Second, their equations

do not all reduce to those of Van Hove, Janner, and Swenson, or even to any self-determined equations, for the homogeneous case. The one that does reduce to these equations does not have an appropriate Wigner function (or phase space distribution function) representation.[8] These differences are due largely to a freedom to transfer contributions between an inhomogeneous term and the gain–loss term. This freedom is eliminated if one requires the presence of a free streaming term, an appropriate definition of the transition probabilities based on an extension of the diagonal singularity condition, and reduction to previously obtained equations for the homogeneous case.

The derivation here is heavily indebted to that of Peterson and Quay and follows, as closely as possible, their notation. The Wigner function formulation[9] of these equations is given, and various limits are discussed.

2. INHOMOGENEOUS SYSTEMS

The expectation value of a given time-dependent operator may be represented as the trace of a time-dependent density matrix times the initial value of the operator, and applies for a pure state as well as for an ensemble of states.[10] Therefore, it is sufficient to obtain a GME for the density matrix. The many-body density matrix may be given as a formal solution to the von Neumann equation

$$\rho(t) = U(t)\rho(0)U^*(t), \tag{2.1}$$

where $U(t) = e^{-iHt}$ (in units such that $\hbar = 1$) and the Hamiltonian H is assumed to have a natural or useful separation into unperturbed and perturbation parts,

$$H = H_0 + V. \tag{2.2}$$

The usual choice for H_0 is the free N-particle Hamiltonian and is assumed to be the case here.[11] The matrix elements of $\rho(t)$ in the representation which diagonalizes H_0 are

$$\rho(t; p_1, p_2) = \sum_{p_3, p_4} U^*(t; p_3, p_2)U(t; p_1, p_4)\rho(0; p_4, p_3), \tag{2.3}$$

where $F(t; p, p') = \langle p|\, F(t)\, |p'\rangle$. Here $|p\rangle$ is a many-body eigenstate of H_0, and p denotes the set of quantum numbers specifying the state. These quantum numbers, e.g., sets of single particle momenta, become continuously distributed in the infinite-volume thermodynamic limit,[11] and we consider the system to be large enough to approximate this limit. The time dependence of a given operator A is calculated from

$$\bar{A}(t) = \operatorname{Tr} \rho(t)A(0) = \sum_{p_1, p_2} \rho(t; p_1, p_2)A(0; p_2, p_1).$$

We shall mean by an inhomogeneous system one for which the operators of interest, A, do not have vanishing matrix elements for all $p_1 \neq p_2$.

The matrix elements $U^*(t; p_3, p_2)$ and $U(t; p_1, p_4)$ may be represented in terms of the resolvent operator for the Hamiltonian[5]

$$U^*(t; p_3, p_2)U(t; p_1, p_4) = -(2\pi)^{-2}\int_{c_1}\int_{c_2} dz_1\, dz_2 \mathcal{R}^{12}(p_3, p_2, p_1, p_4)e^{i(z_1 - z_2)t}, \tag{2.4}$$

where

$$\mathcal{R}^{12}(p_3, p_2, p_1, p_4) = R^1(p_3, p_2)R^2(p_1, p_4),$$
$$R^1 = (H - z_1)^{-1}, \quad R^2 = (H - z_2)^{-1}. \tag{2.5}$$

The contours of integration, c_1 and c_2, enclose a sufficiently large portion of the real axis (H is Hermitian, and it will be assumed that the spectrum is bounded for finite volume). The GME follows from an identity for $\mathcal{R}^{12}(p_3, p_2, p_1, p_4)$.

The resolvent operator may be written as the sum of its diagonal and nondiagonal parts. This formal separation is important to the derivation because it groups terms surviving in the thermodynamic limit according to the diagonal singularity conditions of Van Hove discussed below. Here we follow the method of Swenson to obtain a GME without recourse to the infinite-order perturbation expansions of Van Hove. It is important, however, to realize that the role of the diagonal singularity conditions is implicit in the following separation into diagonal and nondiagonal parts:

$$R = D + DND. \tag{2.6}$$

D is defined to be the diagonal part of R. It is convenient to introduce another diagonal operator $G(z)$, by[12]

$$D = [H_0 + G(z) - z]^{-1}. \tag{2.7}$$

Equations for $G(z)$, $N(z)$, and $D(z)$ may be obtained in terms of H_0 and V.[6] The definitions below generalize those introduced by Peterson and Quay:

$$d^{12}(p_1, p_2) \equiv D^1(p_1) - D^2(p_2),$$
$$g^{12}(p_1, p_2) \equiv G^1(p_1) - G^2(p_2), \tag{2.8}$$
$$\mathfrak{D}^{12}(p_1, p_3) \equiv D^1(p_1)D^2(p_2).$$

In terms of these quantities, we have

$$\mathcal{R}^{12}(p_3, p_2, p_1, p_4) = \mathfrak{D}^{12}(p_2, p_1)Q^{12}(p_3, p_2, p_1, p_4), \tag{2.9}$$

where

$$Q^{12}(p_3, p_2, p_1, p_4) = [\delta_{p_3,p_2} + D^1(p_3)N^1(p_3, p_2)] \times [\delta_{p_1,p_4} + N^2(p_1, p_4)D^2(p_4)]. \quad (2.10)$$

Multiply Eq. (2.9) by $[D^2(p_1)]^{-1} - [D^1(p_2)]^{-1} = [E(p_1) - E(p_2) - g^{12}(p_2, p_1) + z_1 - z_2]$ to get

$$\{(z_1 - z_2) - [E(p_2) - E(p_1)] - g^{12}(p_2, p_1)\} \times \mathcal{R}^{12}(p_3, p_2, p_1, p_4) = d^{12}(p_2, p_1)Q^{12}(p_3, p_2, p_1, p_4). \quad (2.11)$$

Here $E(p)$ is the eigenvalue of H_0 in state $|p\rangle$. The matrix $g^{12}(p_2, p_1)$ may be written as

$$g^{12}(p_2, p_1) = \tfrac{1}{2}[g^{12}(p_1, p_1) + g^{12}(p_2, p_2)] + [\Delta^{12}(p_2) - \Delta^{12}(p_1)],$$
$$\Delta^{12}(p) = \tfrac{1}{2}[G^1(p) + G^2(p)]. \quad (2.12)$$

The reason for this separation is to group terms which will give rise to real and imaginary contributions in the time representation (see below). The real terms arise from Δ and are closely related to the renormalized energies of Van Hove. Further, g^{12} may be expressed in terms of the generalized transition probabilities of Van Hove [see Eq. (2.16)] responsible for finite damping effects, whereas the Δ's are not. Thus the point of the above separation is to pick out the renormalization terms from the damping terms in g^{12}. These comments will be borne out below.

To proceed, it is useful to briefly discuss the diagonal singularity condition. We wish to obtain an expression for the quantity R^1AR^2 which collects terms important in the thermodynamic limit. These terms are special because they are responsible for the damping effects and renormalization observed in large systems,[7] and we identify them in the following way. The matrix elements of an operator F, which is a function of diagonal operators and the potential V, are proportional to $\delta(\mathbf{p} - \mathbf{p}')$ due to conservation of momentum. The argument of this δ function may be zero in many different ways since $\mathbf{p}$ and $\mathbf{p}'$ represent $3N$ variables. In particular, there is a term

$$\prod_{\alpha=1}^{N} \delta(\mathbf{p}_\alpha - \mathbf{p}'_\alpha),$$

where $\mathbf{p}_\alpha$ and $\mathbf{p}'_\alpha$ are the single-particle momenta. Since we are actually considering a very large but finite system, the notation is

$$\delta(\mathbf{p}_\alpha - \mathbf{p}'_\alpha) = [\Omega/(2\pi)^3]\delta_{\mathrm{Kr}}(\mathbf{p}_\alpha - \mathbf{p}'_\alpha)$$

with Ω being the volume and δ_{Kr} the Kronecker delta. Then, each single-particle δ function contributes a factor Ω. Clearly, the term with $\prod_{\alpha=1}^{N} \delta(\mathbf{p}_\alpha - \mathbf{p}'_\alpha)$ has the most δ functions of all those which yield $\delta(\mathbf{p} - \mathbf{p}')$, and therefore dominates in the infinite-volume limit. If each of the single-particle momenta is the same in the initial and final states, then the states are the same, $|p\rangle = |p'\rangle$, and the dominant contribution to $\langle p|F|p'\rangle$ is the diagonal element. Consider now R^1AR^2. There are diagonal singularities coming from each R; these give rise to the contributions $g^{12}(p_1, p_1)$ and $g^{12}(p_2, p_2)$ in (2.12). If A is diagonal, there are also diagonal singularities coming from products of terms from both R^1 and R^2. These latter are the diagonal singularities of $D^1N^1D^1AD^2N^2D^2$ [see Eq. (2.6)] or, equivalently, terms with $N^1(p_3, p_1)N^2(p_1, p_3)$. The generalized transition probabilities occurring in Van Hove's gain–loss terms are defined as the irreducible parts of these contributions only, since they are precisely the ones responsible for the damping effects.

Now consider the case at hand, i.e., A nondiagonal. There are still the contributions to R^1AR^2 from the diagonal singularities of R^1 and R^2 since they do not depend on A. These are again given in the $g^{12}(p_1, p_1)$, $g^{12}(p_2, p_2)$ terms. However, the diagonal singularity condition does not, in general, give the dominant term from contributions involving terms from both R^2 and R^1. The reason for this is that

$$\langle p_3|\, D^1N^1D^1AD^2N^2D^2\,|p_4\rangle$$

may not be proportional to $\delta(\mathbf{p}_3 - \mathbf{p}_4)$. Therefore, we have to find a new prescription for picking out the dominant term in

$$N^1(p_3, p_2)A(p_2, p_2 + k)N^2(p_2 + k, p_4)$$

[here $A(p_2, p_1)$ has been written $A(p_2, p_2 + k)$]. Since N^2 and N^1 individually conserve momentum, we have

$$N^1(p_3, p_2)A(p_2, p_2 + k)N^2(p_2 + k, p_4) \propto \delta(\mathbf{p}_4 - \mathbf{p}_3 - \mathbf{k})A(p_2, p_2 + k).$$

However, it is clear that the dominant term is the one with

$$\prod_\alpha \delta(\mathbf{p}_{4\alpha} - \mathbf{p}_{3\alpha} - \mathbf{k}_\alpha)$$

for the same reasons as discussed above, i.e., it has the most factors of Ω. This implies $|p_4\rangle = |p_3 + k\rangle$, and the singularity does not occur on the diagonal. The relevant contribution is now

$$N^1(p_3, p_2)A(p_2, p_2 + k)N^2(p_2 + k, p_3 + k).$$

If the operator A conserves momentum, then

$$A(p_2, p_2 + k) \propto \delta(\mathbf{k}),$$

and the diagonal singularity condition is regained even if A is nondiagonal. (In the following for simplicity it is assumed A has no part diagonal in the momentum.) The failure of the diagonal singularity condition for inhomogeneous systems has been discussed by Fujita,[1] who circumvents it by a connected diagram prescription for choosing the proper terms.

These observations on the dominant terms in the thermodynamic limit indicate that the generalized transition probabilities should be defined in terms of $N^1(p_3, p_2)N^2(p_2 + k, p_3 + k)$ instead of

$$N^1(p_3, p_2)N^2(p_2, p_3)$$

as in the homogeneous case. Therefore, by analogy we define two functions, W and V:

$$\begin{aligned} &N^1(p_3, p_2)N^2(p_2 + k, p_3 + k) \\ &\quad = W^{12}(p_3, p_2, p_2 + k, p_3 + k) \\ &\qquad + \sum_{p'} W^{12}(p', p_2, p_2 + k, p' + k)\mathfrak{D}^{12}(p', p' + k) \\ &\qquad \times N^1(p_3, p')N^2(p' + k, p_3 + k), \end{aligned}$$

$$\begin{aligned} &V^{12}(p_3, p_2, p_2 + k, p_4) \\ &\quad = Q^{12}(p_3, p_2, p_2 + k, p_4) \\ &\qquad - \sum_{p'} N^1(p', p_2)N^2(p_2 + k, p' + k)\mathfrak{D}^{12}(p', p' + k) \\ &\qquad \times V^{12}(p_3, p', p' + k, p_4). \end{aligned} \tag{2.13}$$

These definitions, with Eq. (2.9), lead to

$$\begin{aligned} &Q^{12}(p_3, p_2, p_2 + k, p_4) \\ &\qquad = V^{12}(p_3, p_2, p_2 + k, p_4) \\ &\qquad\quad + \sum_{p'} W^{12}(p', p_2, p_2 + k, p' + k) \\ &\qquad\quad \times \mathfrak{R}^{12}(p_3, p', p' + k, p_4). \end{aligned} \tag{2.14}$$

The definitions (2.13) are the same as those of Swenson and Janner, for $k = 0$. They are the same[13] as Van Hove's if, in addition, $p_3 = p_4$. The transition probability W in (2.13) has been defined in terms of that part of N^1N^2 which dominates in the thermodynamic limit. This is not an approximation, even for finite volumes, but rather a specific choice of terms to be gathered in the gain term [second term on the right of (2.14)]. While other definitions of W and V are possible, leading to a form like (2.14), they do not have the physical significance implied by the singularity conditions; i.e., it is precisely the singular terms which are responsible for the existence of damping associated with the thermodynamic limit, and no others.[7]

Combining Eqs. (2.11), (2.12), and (2.14) gives

$$\begin{aligned} &\{(z_1 - z_2) - [E(p_2) - E(p_2 + k)] \\ &\qquad - [\Delta^{12}(p_2) - \Delta^{12}(p_2 + k)]\} \\ &\qquad \times \mathfrak{R}^{12}(p_3, p_2, p_2 + k, p_4) \\ &\quad = d^{12}(p_2, p_2 + k)V^{12}(p_3, p_2, p_2 + k, p_4) \\ &\qquad + \sum_{p'} d^{12}(p_2, p_2 + k) \\ &\qquad \times W^{12}(p', p_2, p_2 + k, p' + k) \\ &\qquad \times \mathfrak{R}^{12}(p_3, p', p' + k, p_4) \\ &\qquad + \tfrac{1}{2}[g^{12}(p_2, p_2) + g^{12}(p_2 + k, p_2 + k)] \\ &\qquad \times \mathfrak{R}^{12}(p_3, p_2, p_2 + k, p_4). \end{aligned} \tag{2.15}$$

It remains to express $g^{12}(p, p)$ in terms of W. This may be done in the same way as outlined in the Appendix of Ref. 6, with the result

$$g^{12}(p, p) = -\sum_{p'} d^{12}(p', p')W^{12}(p, p', p', p). \tag{2.16}$$

Use has been made of the fact that

$$W^{12}(p, p', p', p) = W^{21}(p', p, p, p'),$$

which follows from the definition of W. The desired identity for $\mathfrak{R}^{12}$ is now

$$\begin{aligned} &\{(z_1 - z_2) - [E(p_2) - E(p_2 + k)] \\ &\qquad - [\Delta^{12}(p_2) - \Delta^{12}(p_2 + k)]\} \\ &\qquad \times \mathfrak{R}^{12}(p_3, p_2, p_2 + k, p_4) \\ &\quad = d^{12}(p_2, p_2 + k)V^{12}(p_3, p_2, p_2 + k, p_4) \\ &\qquad + \sum_{p'} \{d^{12}(p_2, p_2 + k) \\ &\qquad \times W^{12}(p', p_2, p_2 + k, p' + k) \\ &\qquad \times \mathfrak{R}^{12}(p_3p', p' + k, p_4) \\ &\qquad - \tfrac{1}{2}d^{12}(p', p')[W^{12}(p_2, p', p', p_2) \\ &\qquad + W^{12}(p_2 + k, p', p', p_2 + k)] \\ &\qquad \times \mathfrak{R}^{12}(p_3, p_2, p_2 + k, p_4)\}. \end{aligned} \tag{2.17}$$

Before continuing, we note that for $k = 0$ the equation of Janner and Swenson results; for $k = 0$ and $p_3 = p_2$, Van Hove's result is regained.

The contours of integration in (2.4) may be deformed to give,[5] for $t > 0$,

$$\begin{aligned} &U^*(t; p_3, p_2)U(t; p_2 + k, p_3 + k) \\ &\qquad = \int_{-\infty}^{\infty} dEP_E(t; p_3, p_2, p_2 + k, p_3 + k), \end{aligned} \tag{2.18}$$

$$\begin{aligned} &P_E(t; p_3, p_2, p_2 + k, p_3 + k) \\ &\quad = \tfrac{1}{2}\pi^{-2}\int_{-\infty}^{\infty} dE'\mathfrak{R}^{12}(p_3, p_2, p_2 + k, p_3 + k)e^{2it(E' - i\eta)}, \end{aligned} \tag{2.19}$$

where now $z_1 = E + E' - i\eta$, $z_2 = E - E' + i\eta$, and η is an arbitrarily small positive constant. The

quantity P_E is then shown to satisfy

$$\frac{\partial}{\partial t} P_E(t; p_3, p_2, p_2 + k, p_4)$$
$$- iw(p_2, p_2 + k)P_E(t; p_3, p_2, p_2 + k, p_4)$$
$$- i\int_0^t dt'[\Delta_E(t - t'; p_2) - \Delta_E(t - t'; p_2 + k)]$$
$$\times P_E(t'; p_3, p_2, p_2 + k, p_4)$$
$$= f_E(t; p_3, p_2, p_2 + k, p_4)$$
$$+ 2\pi \sum_{p'} \int_0^t dt'[w_E(t - t'; p', p_2, p_2 + k, p' + k)$$
$$\times P_E(t'; p_3, p', p' + k, p_4)$$
$$- \tfrac{1}{2}[w_E(t - t'; p_2, p', p', p_2)$$
$$+ w_E(t - t'; p_2 + k, p', p', p_2 + k,)$$
$$\times P_E(t'; p_3, p_2, p_2 + k, p_4)]. \quad (2.20)$$

Here $w(p_2, p_2 + k) = E(p_2) - E(p_2 + k)$ and

$$\Delta_E(t; p) = \pi^{-1}\int_{-\infty}^{\infty} dE' \Delta^{12}(p) e^{2it(E' - i\eta)},$$

$$f_E(t; p_3, p_2, p_2 + k, p_4)$$
$$= \tfrac{1}{2} i\pi^{-2}\int_{-\infty}^{\infty} dE' d^{12}(p_2, p_2 + k)$$
$$\times V^{12}(p_3, p_2, p_2 + k, p_4) e^{2it(E' - i\eta)}, \quad (2.21)$$

$$w_E(t, p_3, p_2, p_2 + k, p_4)$$
$$= \tfrac{1}{2} i\pi^{-2}\int_{-\infty}^{\infty} dE' d^{12}(p_2, p_2 + k)$$
$$\times W^{12}(p_3, p_2, p_2 + k, p_4) e^{2it(E' - i\eta)}.$$

Finally, the spectral decomposition of $\rho(t)$ may be defined by

$$\rho_E(t; p_2 + k, p_2) = \sum_{p_3, p_4} P_E(t; p_3, p_2, p_2 + k, p_4)$$
$$\times \rho(0; p_4, p_3),$$

$$\rho(t; p_2 + k, p_2) = \int_{-\infty}^{\infty} dE \rho_E(t; p_2 + k, p_2), \quad (2.22)$$

and it satisfies the GME

$$\frac{\partial}{\partial t} \rho_E(t; p_2 + k, p_2) + iw(p_2 + k, p_2)\rho_E(t; p_2 + k, p_2)$$
$$+ i\int_0^t dt'[\Delta_E(t - t'; p_2 + k) - \Delta_E(t - t'; p_2)]$$
$$\times \rho_E(t'; p_2 + k, p_2)$$
$$= \rho'_E(t; p_2 + k, p_2)$$
$$+ 2\pi \sum_{p'} \int_0^t dt'\{w_E(t - t'; p', p_2, p_2 + k, p' + k)$$
$$\times \rho_E(t'; p' + k, p')$$
$$- \tfrac{1}{2}[w_E(t - t'; p_2, p', p', p_2)$$
$$+ w_E(t - t'; p_2 + k, p', p', p_2 + k)]$$
$$\times \rho_E(t'; p_2 + k, p_2)\}. \quad (2.23)$$

This is the desired result for the N-particle spectral density. It incorporates the renormalized energy differences in the streaming term on the left side and the proper contributions to the generalized transition probabilities w_E on the right side. It remains to show that the Δ_E should indeed be grouped with the unperturbed motion and represents renormalization of the unperturbed energies. This requires that $\Delta_E(t; p)$ be real, and may be shown as follows:

$$\Delta_E(t; p) = \pi^{-1}\int_{-\infty}^{\infty} dE' \tfrac{1}{2}[G(p; E + E' - i\eta)$$
$$+ G(p; E - E' + i\eta)]e^{2it(E' - i\eta)},$$

$$[\Delta_E(t; p)]^* = \pi^{-1}\int_{-\infty}^{\infty} dE' \tfrac{1}{2}[G^*(p; E + E' - i\eta)$$
$$+ G^*(p; E - E' + i\eta)]e^{-2it(E' + i\eta)}$$
$$= \pi^{-1}\int_{-\infty}^{\infty} dE' \tfrac{1}{2}[G^*(p; E - E' - i\eta)$$
$$+ G^*(p; E + E' + i\eta)]e^{2it(E' - i\eta)}.$$

But since D is Hermitian, G is also, and therefore $G^*(p; z) = G(p; z^*)$. Then

$$[\Delta_E(t; p)]^* = \pi^{-1}\int_{-\infty}^{\infty} dE' \tfrac{1}{2}[G(p; E - E' + i\eta)$$
$$+ G(p; E + E' - i\eta)]e^{2it(E' - i\eta)}.$$

The function $\Delta_E(t; p)$ is thus seen to be the real part of the Fourier inversion of $G(p)$. Van Hove has shown that the real part of $G(p)$, denoted by $K(p)$, determines the exact energy of the pth state $\mathcal{E}(p)$ as a solution to

$$E(p) - E + K(p; E) = 0$$

or

$$K(p, \mathcal{E}(p)) = \mathcal{E}(p) - E(p). \quad (2.24)$$

Therefore, $K(p; E)$ may be used as the generator of a perturbation expansion for the exact energy.[12] The function $\Delta_E(p; t)$ is the corresponding real function in time-dependent form [its time integral is indeed $K(p; E)$]. The time dependence represents a non-Markovian nature of the renormalization which is familiar from other many-body approaches, e.g., Green's functions. Thus $\Delta_E(t; p)$ is a renormalization suggestive of Van Hove's description of cloud effects. Equation (2.24) has been used in nuclear physics as the basis for a determination of the energies of the ground state and certain excited states.[14]

The above discussion can be made explicit in the weak coupling limit. A coupling constant λ is introduced in the interaction potential, and the weak coupling limit is defined by $\lambda \to 0$, $t \to \infty$, $\lambda^2 t$ finite. Extensive discussion of the physical meaning of this limiting process may be found in Refs. 3 and 4. It is

assumed that there are two time scales, a microscopic time defined by a parameter τ_0 independent of λ and a longer time scale $\tau_1 \sim \lambda^{-2}$. The GME may be solved to first order in λ to obtain the density matrix at the upper limit of the microscopic time scale. The GME to order λ^2 may then be written for the long time behavior using the short time result as an initial condition. Therefore, for short times we consider (to order λ)

$$\frac{\partial}{\partial t}\rho_E(t; p_2 + k, p_2) + iw(p_2 + k, p_2)\rho_E(t; p_2 + k, p_2) = \rho''_E(t; p_2 + k, p_2),$$

$$\begin{aligned}\rho''_E(t; p_2 + k, p_2) &= \tfrac{1}{2}i\pi^{-2}\int_{-\infty}^{\infty} dE' \Big(\{[E(p_2) - E - E' + i\eta]^{-1} \\ &\quad - [E(p_2 + k) - E + E' - i\eta]^{-1}\} \\ &\quad \times \rho(0; p_2 + k, p_2) \\ &\quad - \sum_{p'}\{\rho(0; p_2 + k, p')V(p', p_2) \\ &\quad \times [E(p') - E - E' + i\eta]^{-1} \\ &\quad - V(p_2 + k, p')[E(p') - E + E' - i\eta]^{-1} \\ &\quad \times \rho(0; p', p_2)\}\Big) e^{2it(E' - i\eta)}.\end{aligned}$$

This equation may be integrated to give the result for short times. The behavior at the upper limit of the short time scale may be represented by the long time limit of the solution obtained. This is found to be

$$\begin{aligned}&\rho_E(t; p_2 + k, p_2) \\ &\quad \to e^{-iw(p_2+k,p_2)t}\delta(E - \tfrac{1}{2}[E(p_2 + k) + E(p_2)]) \\ &\qquad \times \Big(\rho(0; p_2 + k, p_2) - \sum_{p'}\{\rho(0; p_2 + k, p') \\ &\qquad \times V(p', p_2)[E(p') - E(p_2)]^{-1} - V(p_2 + k, p') \\ &\qquad \times \rho(0; p', p_2)[E(p') - E(p_2 + k)]^{-1}\}\Big). \quad (2.25)\end{aligned}$$

Just as with previous results for the diagonal elements, $\rho_E(t; p_2 + k, p_2)$ is restricted to an energy shell. Here, however, there are two energies associated with the nondiagonal elements, and the solution is obtained on the arithmetic average of $E(p_2)$ and $E(p_2 + k)$.

The GME on the long time scale is obtained by noting that $\rho_E(t; p_2 + k, p_2)$ varies slowly on this scale,[15] compared with Δ_E and w_E, and may be extracted from the time integrals. For $t > \tau_0$ the upper limit on the time integrations may be set to ∞. The resulting GME is

$$\begin{aligned}&\frac{\partial}{\partial t}\rho_E(t; p_2 + k, p_2) + iw(p_2 + k, p_2)\rho_E(t; p_2 + k, p_2) \\ &\quad + i[\Delta_E(p_2 + k) - \Delta_E(p_2)]\rho_E(t; p_2 + k, p_2) \\ &\quad = \sum_{p'}\{w_E(p', p_2, p_2 + k, p' + k) \\ &\qquad \times \rho_E(t; p' + k, p') - \tfrac{1}{2}[w_E(p_2, p', p', p_2) \\ &\qquad + w_E(p_2 + k, p', p', p_2 + k)]\rho_E(t; p_2 + k, p_2)\}, \quad (2.26)\end{aligned}$$

where $\Delta_E(p)$ and $w_E(p_1, p_2, p_3, p_4)$ are the time integrals of $\Delta_E(t; p)$ and $2\pi w_E(t; p_1, p_2, p_3, p_4)$, respectively. Equation (2.26) is to be solved subject to the initial condition given by (2.25). Use of this initial condition and Eq. (2.22) allows one to write

$$\begin{aligned}&\rho_E(t; p_2 + k, p_2) \\ &\quad = \delta(E - \tfrac{1}{2}[E(p_2 + k) + E(p_2)])\rho(t; p_2 + k, p_2).\end{aligned}$$

Then Eq. (2.26) leads, after integration over E, to

$$\begin{aligned}&\frac{\partial}{\partial t}\rho(t; p_2 + k, p_2) + i\tilde{w}(p_2 + k, p_2)\rho(t; p_2 + k, p_2) \\ &\quad = 2\pi\sum_{p'}\{w^{(0)}(p', p_2, p_2 + k, p' + k)\rho(t; p' + k, p') \\ &\qquad - \tfrac{1}{2}[w^{(0)}(p_2, p', p' + k, p_2 + k) \\ &\qquad + w^{(0)}(p_2, p' + k, p' + k, p_2)]\rho(t; p_2 + k, p_2)\}. \quad (2.27)\end{aligned}$$

Here $w^{(0)}$ indicates the transition probabilities to lowest order in λ, and will be discussed in the next section. Also,

$$\begin{aligned}\tilde{w}(p_2 + k, p_2) &= w(p_2 + k, p_2) \\ &\quad + K(p_2; \tfrac{1}{2}[E(p_2 + k) + E(p_2)]) \\ &\quad - K(p_2 + k; \tfrac{1}{2}[E(p_2 + k) + E(p_2)]). \quad (2.28)\end{aligned}$$

To understand this last expression, recall from (2.24) that $E(p_2) - E + K(p_2; E) = 0$ implies $E = \mathcal{E}(p)$. Since there are two states involved, $p_2 + k$ and p_2, it is convenient to write

$$\begin{aligned}&E + \tfrac{1}{2}[\mathcal{E}(p_2) + \mathcal{E}(p_2 + k)] - E(p_2) \\ &\quad = K(p_2; E + \tfrac{1}{2}[\mathcal{E}(p_2 + k) + \mathcal{E}(p_2)]), \\ &\qquad - E + \tfrac{1}{2}[\mathcal{E}(p_2) + \mathcal{E}(p_2 + k)] - E(p_2 + k) \\ &\quad = K(p_2 + k; -E + \tfrac{1}{2}[\mathcal{E}(p_2 + k) + \mathcal{E}(p_2)]).\end{aligned}$$

Both of these equations are satisfied by

$$E = \tfrac{1}{2}[\mathcal{E}(p_2) - \mathcal{E}(p_2 + k)],$$

i.e., half the exact energy differences. Subtracting these equations gives

$$\begin{aligned}2E + w(p_2 + k, p_2) &= K(p_2; E + \tfrac{1}{2}[\mathcal{E}(p_2) \\ &\quad + \mathcal{E}(p_2 + k)]) - K(p_2 + k; -E + \tfrac{1}{2}[\mathcal{E}(p_2 + k) \\ &\quad + \mathcal{E}(p_2)]).\end{aligned}$$

A perturbation expansion for $\mathcal{E}(p_2) - \mathcal{E}(p_2 + k)$ results by iterating about $E = 0$:

$$\mathcal{E}(p_2 + k) - \mathcal{E}(p_2) = w(p_2 + k, p_2) + K(p_2 + k; \tfrac{1}{2}[E(p_2 + k) + E(p_2)]) - K(p_2; \tfrac{1}{2}[E(p_2 + k) + E(p_2)]) + \cdots.$$

The terms shown on the right side are the same as those on the right side of (2.28). Thus $\tilde{w}(p_2 + k, p_2)$ is the first term in a perturbation expansion of $\mathcal{E}(p_2 + k) - \mathcal{E}(p_2)$ about zero. Only the lowest order in λ should be used since the weak coupling limit has been taken. This expansion of the energy differences about zero probably does not converge for states $p_2 + k$ and p_2 with greatly different energies. This is because the $w(p_2 + k, p_2)$ term in the GME gives a rapid oscillation to $\rho(t; p_2 + k, p_2)$, which invalidates the assumption that it is smoothly varying. This difficulty can be avoided by considering the GME in an interaction representation defined by

$$\rho(t; p_2 + k, p_2) = e^{-iw(p_2+k,p_2)t}\rho_I(t; p_2 + k, p_2).$$

The ρ_I may now be assumed smooth and the renormalized energies are given to second order in λ, i.e., an expansion of $\mathcal{E}(p_2 + k) - \mathcal{E}(p_2)$ about $w(p_2 + k, p_2)$, rather than about zero. Furthermore, the weak coupling limit must be considered strictly as a model, since the existence of the two time scales, related in a simple way to the coupling constant, is unproved.

3. WIGNER FUNCTION FORMULATION

For many purposes, such as investigating the foundations of kinetic theory, it is convenient to use the Wigner function rather than the density matrix.[9,16] The weak coupling, Markovian limit of the corresponding GME is called a master equation. Kac has discussed[17] the derivation of kinetic equations for homogeneous systems from a model master equation. It has been conjectured[17,18] that a master equation does not exist for inhomogeneous systems, or at least cannot serve as a basis for the kinetic theory of such systems. A useful starting point for an investigation of this question is a GME for the Wigner function. The Wigner function is related to the density matrix in momentum representation by

$$F(r, p; t) = \pi^{-3N}\int dp' e^{2ip'\cdot r}\rho(p + p', p - p'). \quad (3.1)$$

Here r and p represent the $3N$ components of position and momentum, respectively, and N is the number of particles. The "integral" in (3.1) anticipates the infinite volume limit and is defined by

$$\int dp'(\cdot) \equiv \frac{(2\pi)^{3N}}{V^N}\sum_{p'}(\cdot).$$

The Wigner function may be obtained by an "inversion" of the Weyl prescription for quantizing classical phase functions. Thus the utility of the Wigner function in quantum mechanics is that it has many of the properties of the classical distribution functions defined on phase space. Other functions have been proposed as well.

To make the connection with the GME, define the spectral decomposition of the Wigner function by

$$F_E(r, p; t) = \pi^{-3N}\int dp' e^{2ip'\cdot r}\rho_E(p + p', p - p'; t),$$
$$F(r, p; t) = \int dE F_E(r, p; t). \quad (3.2)$$

Equation (2.23) now gives the corresponding GME for the Wigner function:

$$\frac{\partial}{\partial t}F_E(r, p; t) + \frac{\mathbf{p}}{m}\cdot\nabla_r F_E(r, p; t) + \int_0^t dt'\int dr'\Delta_E(r - r', p; t - t')F_E(r', p; t) = F'_E(r, p; t) + 2\pi\int_0^t dt'\int dr'\int dp'' \times [w_E(r - r', p'', p; t - t')F_E(r', p''; t') - w'_E(r - r', p, p''; t - t')F_E(r', p; t')], \quad (3.3)$$

with

$$\Delta_E(r, p; t) = i\pi^{-3N}\int dp' \times e^{2ip'\cdot r}[\Delta_E(t; p + p') - \Delta_E(t; p - p')],$$

$$w_E(r, p'', p; t) = \pi^{-3N}\frac{V^N}{(2\pi)^{3N}}\int dp' e^{2ip'\cdot r} \times w_E(t; p'' - p', p - p', p + p', p'' + p'),$$

$$w'_E(r, p, p''; t) = \pi^{-3N}\frac{V^N}{(2\pi)^{3N}}\int dp' \times e^{2ip'\cdot r}\tfrac{1}{2}[w_E(t; p - p', p'', p'', p - p') + w_E(t; p + p', p'', p'', p + p')], \quad (3.4)$$

$$F'_E(r, p; t) = \pi^{3N}\int dp' e^{2ip'\cdot r}\rho'_E(t; p + p', p - p').$$

It follows from the discussion following (2.23) that $\Delta_E(r, p; t)$ is real. It is also straightforward to show the transition probabilities,

$$w_E(r, p'', p; t) \quad \text{and} \quad w'_E(r, p, p''; t),$$

are also real, as they should be.

Equation (3.3) is nonlocal in both the space and time variables, reflecting the fact that in general there are spatial correlations as well as "memory effects." To demonstrate that these two features are quite independent, we consider first the limit of a weakly inhomogeneous system. In this case $\rho_E(t; p + p', p - p')$ is a peaked function of p' about $p' = 0$, and the integrands in (3.4) may be expanded about $p' = 0$ to give, to lowest order,

$$\Delta_E^{(0)}(r, p; t) = \frac{\partial}{\partial r}\delta(r)\frac{\partial}{\partial p}\Delta_E(t; p),$$

$$w_E^{(0)}(r, p'', p; t) = \frac{V^N}{(2\pi)^{3N}}\delta(r)w_E(t; p'', p) = w_E'^{(0)}(r, p'', p; t). \tag{3.5}$$

This gives the weak inhomogeneity approximation to (3.3):

$$\frac{\partial}{\partial t}F_E(r, p; t) + \frac{\mathbf{p}}{m}\cdot\nabla_r F_E(r, p; t) + \int_0^t dt\left(\frac{\partial}{\partial p}\cdot\Delta_E(t - t'; p)\right)\frac{\partial}{\partial r}F_E(r, p; t')$$
$$= F_E'^{(0)}(r, p; t) + 2\pi\int dt'\int dp'' \times \frac{V^N}{(2\pi)^{3N}}[w_E(t - t'; p'', p)F_E(r, p''; t') - w_E(t - t'; p, p'')F_E(r, p; t')]. \tag{3.6}$$

The gain–loss term in (3.6) is the same as that obtained by Van Hove, Janner, and Swenson for the homogeneous case. Therefore, the principle difference for the weakly inhomogeneous system is the presence of the streaming terms on the left side. The equation is still non-Markovian in time, although spatially local in this approximation.

A further simplification occurs if, in addition, the weak coupling limit is taken. From (2.27) and (3.5) we obtain

$$\frac{\partial}{\partial t}F(r, p; t) + \left[\frac{\mathbf{p}}{m} + \frac{\partial}{\partial p}K(p; E(p))\right]\cdot\nabla_r F(r, p; t) = \int dp''\sigma(p'', p)\delta_{\mathrm{Kr}}(E(p'') - E(p)) \times [F(r, p''; t) - F(r, p; t)] \tag{3.7}$$

with

$$\sigma(p'', p) = \frac{V^N}{(2\pi)^{3N}}|\langle p''|\, V\,|p\rangle|^2.$$

The "collision integral" on the right of (3.7) is now just the gain–loss term of the Pauli equation. The streaming term may be written as $\tilde{\mathbf{p}}\cdot\nabla_r F(r, p; t)$, where $\tilde{p}$ is the renormalized momentum defined by $\tilde{p} = (\partial/\partial p)\mathcal{E}(p)$. In (3.7) we have $\tilde{p}$ to first order, as follows from (2.24):

$$\frac{\mathbf{p}}{m} + \frac{\partial}{\partial p}K(p; E(p)) = \frac{\partial}{\partial p}\mathcal{E}^{(1)}(p);$$
$$\mathcal{E}^{(1)}(p) \equiv E(p) + K(p; E(p)).$$

Equivalently, the streaming term may be written in terms of an effective mass, $\mathbf{p}/m^*\cdot\nabla_r F(r, p; t)$, with $m^* \equiv m[\partial\mathcal{E}(p)/\partial E(p)]^{-1}$.

Equation (3.7) may be considered the simplest generalization of the usual master equations to inhomogeneous systems. Its validity, of course, requires detailed investigation of the two limits taken. The derivation of the Boltzmann equation from (3.7) for homogeneous systems has been discussed by Kac[17] and, in the context of a GME, by McLennan and Swenson.[19]

4. DISCUSSION

It was mentioned in the introduction that Peterson and Quay[6] found more than one identity for

$$\mathcal{R}^{12}(p_1, p_2, p_3, p_4).$$

In fact, there are an infinite number of such identities.[20] This can be seen by choosing to define V^{12} by, for instance,

$$Q^{12}(p_3, p_2, p_1, p_4) = V^{12}(p_3, p_2, p_1, p_4) + \sum_{p', p''} A(W^{12})_{p_2, p_1, p', p''}\mathcal{R}^{12}(p_3, p', p'', p_4),$$

where A is an arbitrary linear function of W^{12}. Expressing $g^{12}(p, p)$ in terms of W^{12} requires only that $W^{12}(p_3, p_2, p_1, p_4)$ reduce to Van Hove's definition for $p_3 = p_4$ and $p_2 = p_1$, and makes no use of the definition of V^{12}. Therefore, for various choices of A, one is merely transferring contributions of the gain term to, or from, the inhomogeneous term V. Indeed, if V is defined by

$$Q^{12}(p_3, p_2, p_1, p_4) = V^{12}(p_3, p_2, p_1, p_4) + \sum_{p'}\tfrac{1}{2}d^{12}(p', p') \times (W^{12}(p_2, p', p', p_2) + W^{12}(p_1, p', p', p_1)) \times \mathcal{R}^{12}(p_3, p_2, p_1, p_4),$$

the gain–loss term vanishes! In the derivation presented here, V and W are defined by choosing only those contributions for the gain–loss term which have special significance in the thermodynamic limit. The relevance of this choice is due to the usual meaning of the gain–loss term as the contribution responsible for dissipation. Thus, while other GME's may be obtained,

their "gain–loss" terms cannot be interpreted as such, and it is not clear how one would make a meaningful approximation to these equations without a reordering of terms according to the singularity conditions, or similar criteria. The separation of the resolvent operator into diagonal and nondiagonal parts is not sufficient by itself to pick out the important terms, as the above ambiguity suggests. In addition, one must choose the dominant terms from the product of resolvents (since the thermodynamic limit of the product is generally different from the product of the limits), and it was this analysis which led to the choice of definition for W and V here. A similar situation exists with other methods for obtaining GME's. More elegant and direct derivations of general rate equations may be obtained by using projection operator techniques. However, there still remains the justification of a choice of projector, and again it requires specification of terms dominating the effects to be described.

Finally, it should be mentioned that, although the above reordering is based on the behavior of large systems, the thermodynamic limit has not actually been taken in Eqs. (2.23) or (3.3). One must first multiply by the matrix elements of the operator to be averaged and sum, before actually carrying out the thermodynamic limit. Thus, while the density matrix may not strictly exist in the thermodynamic limit, it is expected that calculations for a large but finite system will lead to expectation values whose limit is well defined. For an example of this, see Ref. 19.

ACKNOWLEDGMENT

The author is pleased to acknowledge a number of useful conversations with Professor R. J. Swenson.

* Research supported by the National Aeronautics and Space Administration.

[1] An early review with references is given by G. Chester, Rept. Progr. Theoret. Phys. **26**, 411 (1963); see also, G. Emch, *Lectures in Theoretical Physics*, edited by W. Brittin (U. of Colorado Press, Boulder, 1966), Vol. VIIIa; S. Fujita, *Non-Equilibrium Quantum Statistical Mechanics* (Saunders, Philadelphia, 1966); E. Montroll, in *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1962); I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1961); R. Zwanzig, *Lectures in Theoretical Physics*, edited by W. Brittin and J. Downs (Interscience, New York, 1961), Vol. III.

[2] W. Pauli, *Festschrift zum 60. Geburtstage A. Sommerfelds* (Hirtzel, Leipzig, 1928), p. 30.

[3] L. Van Hove, Physica **23**, 441 (1957); **25**, 268 (1959); *La théorie des gaz neutres et ionisés* (Hermann, Paris, 1960), p. 149.

[4] A. Janner, Helv. Phys. Acta. **35**, 1 (1962).

[5] R. Swenson, J. Math. Phys. **3**, 1017 (1962).

[6] R. Peterson and P. Quay, J. Math. Phys. **5**, 85 (1964).

[7] L. Van Hove, Physica **21**, 901 (1955); **22**, 343 (1956).

[8] This is due to the fact that the equation for the nondiagonal elements depends on the diagonal elements of the density matrix.

[9] E. Wigner, Phys. Rev. **40**, 749 (1932).

[10] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1955).

[11] It is only important that some of the quantum numbers of H_0 becomes continuous in the infinite volume limit. There may be some, however, such as spin, which do not.

[12] G may be identified as the self-energy—see, e.g., E. P. Gross, in *Mathematical Methods in Solid State and Superfluid Theory*, edited by R. C. Clark and G. H. Derrick (Plenum, New York, 1967).

[13] Strictly speaking, Van Hove's results are only valid in the thermodynamic limit, while those of Swenson and those here are valid even for finite volumes. This leads to differences of order $1/\Omega$ in the functions G and N (Ref. 5, p. 1020). It is important to note also that the diagonal singularity condition has only been used here on R^1AR^2, and the "irreducibility conditions" of Van Hove have not been required.

[14] K. Kumar, *Perturbation Theory and the Nuclear Many-Body Problem* (North-Holland, Amsterdam, 1962).

[15] This is true only if $w(p_2 + k, p_2)$ is small; see comments below.

[16] H. Mori, I. Oppenheim, and J. Ross, in *Studies in Statistical Mechanics*, Vol. I, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962).

[17] M. Kac, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (U. of California Press, Berkeley, 1956).

[18] M. Dresden, in *Studies in Statistical Mechanics*, Vol. I, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962), p. 310.

[19] J. McLennan and R. Swenson, J. Math. Phys. **4**, 1527 (1963).

[20] This was first pointed out to me by Professor R. Swenson.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Crossing Symmetric Expansions of Physical Scattering Amplitudes; The $O(2, 1)$ Group and Lamé Functions

N. W. MACFADYEN* AND P. WINTERNITZ†
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 7 July 1970)

An explicitly symmetric 2-variable expansion, convergent in the physical regions of both the s and t channels, is obtained for 2-body relativistic scattering amplitudes. Use is made of a symmetric mapping of the variables s and t onto a 2-dimensional hyperboloid, leading to an expansion in terms of Lamé functions, which are the basis functions for an irreducible representation of the group $O(2, 1)$ in our parametrization.

INTRODUCTION

A series of previous publications[1–7] (for a more complete list of references see Refs. 6 and 7) was devoted to the development of a scattering theory based on 2-variable expansions of relativistic scattering amplitudes, just as the usual partial wave analysis is based on a single-variable expansion at a fixed energy or the Regge scattering theory is based on a single-variable expansion for a fixed momentum transfer. The aim of this approach is to make use of group representation theory to separate out as much as possible of the kinematics of the reactions involved. Indeed, all the dependence on kinematic parameters, such as energy and scattering angle (or the Mandelstam variables s, t, and u), is separated into known functions (the basis functions for group representations, the finite transformation matrix elements, or some similar objects), whereas the dynamics are transferred to the expansion coefficients, called the "Lorentz amplitudes" in Refs. 1–7.

The desire to make as much use as possible of relativistic invariance, in particular to incorporate the $O(3)$, $O(2, 1)$, E_2, and $O(3, 1)$ little group expansions (see Refs. 7–10), leads to expansions based on an $O(3, 1)$ group, isomorphic to the homogeneous Lorentz group, which underlies the kinematics of the special theory of relativity.

Two-variable expansions, based on an $O(3, 1)$ group, have so far been obtained for the 2-body scattering of spinless particles with arbitrary (positive) rest masses, for arbitrary values of s and t. The method used was to consider the scattering amplitude as a function of the relativistic velocities $v = p/m$ ($p = \{p_0, \mathbf{p}\}$ is the energy–momentum, m is the rest mass, so that $v^2 = 1$ on the mass shell) of all four particles and then to express the components of three of the velocities in terms of the fourth, by making use of the conservation laws and choosing a specific frame of reference. In such a fashion, we obtain a mapping of a physical region in the Mandelstam plane onto the hyperboloid $v^2 = 1$. The scattering amplitude $f(s, t)$ can then be considered to be a function $f(v)$ of a single point v in relativistic velocity space, and, as s, t run through the physical region of one channel, the point v runs over the whole upper sheet of the considered hyperboloid. Having such a function $f(v)$, we find it a relatively simple matter to expand this function in terms of the basis functions of the irreducible representations of the group of motions of the space $v^2 = 1$, i.e., the group $O(3, 1)$. To make the above considerations explicit, it is necessary to choose a certain frame of reference, like the center-of-mass system, the brick-wall system, etc., to choose convenient coordinates on the hyperboloid, and to choose a convenient basis for the representations of $O(3, 1)$. All these questions are linked to each other and were discussed in detail in Refs. 1–7. In particular, it was shown that different chains of reductions of the $O(3, 1)$ group to its subgroups lead to different coordinate systems, different bases for the group representations, and thus to different expansions of scattering amplitudes.

The group reduction $O(3, 1) \supset O(3) \supset O(2)$ was associated with spherical coordinates on a hyperboloid and was used to consider scattering in the center-of-mass frame. The $O(3)$ group then figured as the little group of a timelike vector—the total energy–momentum $p_1 + p_2$. The 2-variable expansion could be interpreted as an $O(3)$ little group expansion in terms of $P_l(\cos\vartheta)$ (Legendre polynomials; ϑ is the c.m.s. scattering angle), supplemented by an expansion of the partial wave amplitude $a_l(s)$ in terms of certain Legendre functions.

The reduction $O(3, 1) \supset O(2, 1) \supset O(2)$ was associated with hyperbolic coordinates on the hyperboloid and can be conveniently applied to scattering in the brick-wall frame; and the $O(2, 1)$ subgroup appears as the little group of a spacelike vector—the momentum transfer $p_1 - p_3$ [for $t = (p_1 - p_3)^2 < 0$]. The obtained 2-variable expansion can be interpreted

as a Regge integral (the background integral of Regge pole theory or an integral along the path $\mathrm{Re}\, l = \text{const} > -\frac{1}{2}$, where l is the complex angular momentum), supplemented by a definite expansion for the Reggeized partial wave amplitude $a(l, t)$.

Finally, the reduction $O(3, 1) \supset E_2 \supset O(2)$, where E_2 is the group of motions of an Euclidean plane, is related to so-called horospheric coordinates[1] and was used when considering scattering in a "light-velocity" frame of Ref. 4. The E_2 group figures as the little group of a lightlike vector $K(s, t)$, constructed out of the momenta of the four particles involved, which for unequal-mass scattering ($m_1 \neq m_3$ and/or $m_2 \neq m_4$) at $t = 0$ coincides with the momentum transfer. The obtained 2-variable expansion can be interpreted as a generalization of the E_2 little group expansions for $t = 0$, again supplemented by a further expansion for the corresponding partial wave amplitude.

The $O(3, 1)$ 2-variable expansions thus incorporate the $O(3)$, $O(2, 1)$, and E_2 little group expansions. The $O(3, 1)$ little group expansion[9] for equal-mass scattering at $t = 0$ is also contained as a special limiting case.

In the program described above, and indeed in any application of harmonic analysis on a group manifold or a homogeneous space, the choice of a specific basis for the representations is very important and determines the properties of the resulting expansion. An approach commonly used in physics is to choose a complete set of commuting operators, using operators in the enveloping algebra and possibly some others, and then to construct the basis functions as the set of common eigenfunctions of this complete set of operators. A standard way to construct such complete sets of commuting operators is to consider various possible chains of subgroups of the given group and to form the set out of the invariant operators of the group itself and of all the subgroups in the chain.[11,1,2]

In the last few years much attention has been devoted to the representation theory of the Poincaré group, the Lorentz group $O(3, 1)$, and their subgroups. Various aspects of the representations of $O(3, 1)$ have been considered both in the "canonical basis," corresponding to an $O(3, 1) \supset O(3) \supset O(2)$ reduction,[12–15] and in the "noncompact" bases, corresponding to reductions to the $O(2, 1)$ or E_2 groups[15–20] (the list of references is by no means complete).

The group $O(2, 1)$ in a compact basis, corresponding to the reduction $O(2, 1) \supset O(2)$, was first investigated by Bargmann.[21] The noncompact reduction $O(2, 1) \supset O(1, 1)$ has been considered by numerous authors (see, e.g., Refs. 19, 22, and 23).

It should be stressed, however, that reducing a group to its subgroups is not the only manner in which complete sets of operators determining basis functions of representations can be constructed. Indeed, all that is necessary is to include the Casimir operators of the group itself—the choice of the remaining operators is arbitrary. A systematic investigation of possible complete sets was initiated in Refs. 2, 3, and 24 for all the little groups of the Poincaré group. It was shown that if, for convenience, we restrict ourselves to second-order operators, then there exists only a finite number of nonequivalent complete sets. For the group $O(3)$ there are only two such sets, for E_2 there are four, for $O(2, 1)$ there are nine, and for $O(3, 1)$ there are 34. It was shown[3,24] that each independent set corresponds to one system of coordinates, allowing the separation of variables[25] in the Laplace–Beltrami equation for the space on which the corresponding group acts transitively. Those operators which were obtained as Casimir operators of subgroups correspond to the simplest types of coordinates; the other operators correspond to more complicated ones, of the elliptic type.

From the point of view of group representation theory, it is of interest to investigate representations in these new bases, not related to any subgroups. As a by-product, one might hope to obtain new group-theoretical properties of the special functions, which will appear as basis functions, matrix elements, or in some similar role (for the connection between special function theory and group representations, see, e.g., Vilenkin[26]).

From the point of view of physical applications, we are led to a consideration of group representations in such an "elliptic" basis by the desire to incorporate crossing symmetry conveniently into the 2-variable expansion theory. Indeed, consider the scattering amplitude $f(v)$, where v is a point on the hyperboloid $v^2 = 1$. We can parametrize v, using arbitrary coordinates, but we wish to obtain a convenient mapping of the Mandelstam plane onto it. In particular, if we succeed in constructing such a mapping that an interchange of, say, s and t corresponds to an interchange of two of the curvilinear coordinates of v and if the basis functions in the expansion have a simple behavior under such an interchange, then we can obtain explicitly crossing-symmetric expansions.

In this paper we show that such a program is indeed feasible and leads to crossing symmetric expansions in terms of Lamé functions. We limit ourselves to expansions based on an $O(2, 1)$ group, rather than $O(3, 1)$; but this is purely for mathematical simplicity, and indeed an extension to the group $O(3, 1)$ is in progress.

In Sec. I we discuss the hyperboloid $v_0^2 - v_1^2 - v_2^2 = 1$ on which $O(2, 1)$ acts transitively, consider coordinate systems allowing the separation of variables in the Laplace equation, and introduce the basis functions of representations of $O(2, 1)$. Section II is devoted to the mapping of s, t, and u onto this hyperboloid. In Sec. III we write an expansion of the scattering amplitude for one of the two variables fixed in terms of Lamé functions, and in Sec. IV we derive the 2-variable expansion. In Sec. V we apply the derived expansions to obtain an explicitly crossing-symmetric expansion of a physical scattering amplitude, convergent in both the s and t physical regions. Possible future applications and developments are discussed in the final section.

I. THE $O(2, 1)$ GROUP AND ELLIPTIC COORDINATES

The Lorentz group in a 3-dimensional space with two spacelike dimensions and one timelike dimension can be defined as the group of linear transformations of a 3-dimensional vector space leaving the indefinite form $v_0^2 - v_1^2 - v_2^2$ invariant. In this paper we shall concentrate on one of the homogeneous spaces of this group, namely the upper sheet of the two sheeted hyperboloid

$$v_0^2 - v_1^2 - v_2^2 = 1, \quad v_0 \geq 1. \tag{1}$$

The generators of the group will be denoted K_1, K_2 (boosts along the first and second axis), and L_3 (a rotation in the 12 plane), satisfying the commutation relations

$$[L_3K_1] = K_2, \quad [K_2L_3] = K_1, \quad [K_1K_2] = -L_3, \tag{2}$$

so that the generators are represented by anti-Hermitian operators in any unitary representation. The invariant of the group, i.e., the Casimir operator, can be written as

$$\Delta = L_3^2 - K_1^2 - K_2^2. \tag{3}$$

We shall consider the space of functions $f(v)$ square integrable with respect to the invariant measure $dv_1 dv_2/v_0$ on the hyperboloid. Any function transforming under an irreducible representation of $O(2, 1)$ must, of course, be an eigenfunction of Δ:

$$\Delta f_l(v) = -l(l + 1)f_l(v). \tag{4}$$

As stated in the introduction, basis functions $f_l(v)$ for an irreducible representation can be obtained by demanding that they should be the eigenfunctions of a further differential operator, commuting with Δ, and possibly of some (discrete) reflection operators. A second-order symmetric differential operator, commuting with Δ, can be written as

$$\begin{aligned} L = {} & aK_1^2 + b(K_1K_2 + K_2K_1) \\ & + cK_2^2 + d(K_1L_3 + L_3K_1) \\ & + e(K_2L_3 + L_3K_2) + fL_3^2 \end{aligned} \tag{5}$$

(where $a, \cdots, f$ are arbitrary constants), and it was shown in a previous paper[3] that, by an inner automorphism of the group and by taking linear combinations of the transformed operator L' with Δ, a general operator L can be reduced to one of nine inequivalent standard forms L_s. The basis functions can be taken as the common solutions of (4) and the equation

$$L_s f_{lh}(v) = h f_{lh}(v). \tag{6}$$

It was also shown[3] that Eqs. (4) and (6) allow the separation of variables so that we can write

$$f_{lh}(v) = \phi_{lh}(a)\Psi_{lh}(b), \tag{7}$$

where a and b are one of the nine types of curvilinear coordinates which do allow separation in (4). The problem of finding all such coordinate systems in 2- and 3-dimensional spaces of constant curvature has been discussed by Olevsky,[25] and there is a one-to-one correspondence[3] between the operators L_s and the separable coordinate systems. The three simplest cases are those in which L_s is simply the square of the generator of one of the three nonequivalent 1-parameter subgroups of $O(2, 1)$: $(L_3)^2$, $(K_1)^2$, and $(K_1 + L_3)^2$, corresponding, respectively, to rotations $O(2)$, boosts $O(1, 1)$, and Euclidean translations E_1, that is, to spherical, hyperbolic, and horocyclic coordinates,[3,25] respectively. The remaining six operators L_s correspond to elliptic coordinates or various degenerations thereof and have so far received very little attention.

Since our aim is to write crossing symmetric expansions of scattering amplitudes, we wish to choose separable coordinates for which the two functions $\phi_{lh}(a)$ and $\Psi_{lh}(b)$ in (7) are the same. Clearly, the subgroup type coordinates do not satisfy this condition, since one function will be an exponential and the other a Legendre or cylindrical function. However, the two nondegenerate elliptic coordinate systems prove to be very convenient.

Let us now consider one of these two coordinate systems; the other, less suitable for the two variable expansions, will be discussed in the Appendix.

These coordinates[3,25] on the hyperboloid $v^2 = 1$

can be introduced algebraically by the relations

$$v_0^2 = \frac{(\rho_1 - b)(b - \rho_2)}{(a - b)(b - c)}, \quad v_1^2 = \frac{(\rho_1 - c)(c - \rho_2)}{(a - c)(b - c)},$$
$$v_2^2 = \frac{(\rho_1 - a)(a - \rho_2)}{(a - b)(a - c)}, \tag{8}$$
$$\rho_2 < c < b < a < \rho_1,$$

where ρ_1 and ρ_2 are confocal coordinates and a, b, and c are constants. In order to uniformize these expressions, i.e., express v_0, v_1, and v_2, instead of their squares, in such a fashion as to uniquely cover the hyperboloid, it is convenient to introduce the Jacobian elliptic functions,[27] with

$$\rho_1 - b = -(b - c)\,\mathrm{cn}^2(\alpha, k),$$
$$b - \rho_2 = -(a - b)\,\mathrm{cn}^2(\beta, k').$$

In terms of the new elliptic coordinates α and β, we have

$$v_0 = -\mathrm{cn}(\alpha, k)\,\mathrm{cn}(\beta, k'),$$
$$v_1 = i\,\mathrm{sn}(\alpha, k)\,\mathrm{dn}(\beta, k'), \tag{9}$$
$$v_2 = i\,\mathrm{dn}(\alpha, k)\,\mathrm{sn}(\beta, k')$$

with

$$k^2 = (b - c)/(a - c), \quad k'^2 = (a - b)/(a - c),$$
$$k^2 + k'^2 = 1. \tag{10}$$

The properties of the Jacobian elliptic functions are discussed in Ref. 27. Let us here only remind ourselves that they are doubly periodic meromorphic functions with a real quarter-period $K(k)$ and imaginary quarter-period $iK'(k)$. In general, K and K' are given in terms of complete elliptic integrals. For us it will be convenient to put

$$k = k' = 2^{-\frac{1}{2}}, \quad b = (a + c)/2, \tag{11}$$

then

$$K = K' = K(2^{-\frac{1}{2}}) = [\Gamma(\tfrac{1}{4})]^2/4(\pi)^{\frac{1}{2}}.$$

The range for the variables α and β in (9) is

$$\alpha \in (iK, iK + 2K), \quad \beta \in (iK, iK + 2K). \tag{12}$$

It is a simple matter to check that the metric in elliptic coordinates can be written as

$$ds^2 = \tfrac{1}{2}[\mathrm{cn}^2(\alpha, k) + \mathrm{cn}^2(\beta, k)](d\alpha^2 + d\beta^2), \tag{13}$$

that the Laplace operator is

$$\Delta = \frac{2}{\mathrm{cn}^2(\alpha, k) + \mathrm{cn}^2(\beta, k)}\left(\frac{\partial}{\partial\alpha^2} + \frac{\partial}{\partial\beta^2}\right), \tag{14}$$

and that the invariant measure over the hyperboloid is

$$\frac{dv_1\,dv_2}{v_0} = -\tfrac{1}{2}(\mathrm{cn}^2\alpha + \mathrm{cn}^2\beta)\,d\alpha\,d\beta. \tag{15}$$

As was shown in Ref. 3, the operator L_s of (6) for this coordinate system can be written as

$$L_s = K_1^2 - \frac{b - c}{a - c}L_3^2 = K_1^2 - \tfrac{1}{2}L_3^2. \tag{16}$$

The generators of the group as differential operators can be obtained by considering the quasiregular representation[26] of the group $g \to T_g$, where $T_g f(v) = f(g^{-1}v)$, parametrizing v according to (9), taking the group element $g(t)$ from a 1-parameter subgroup of $O(2, 1)$, and expanding $f(g^{-1}v)$ into a Taylor series. Thus we obtain

$$K_1 = \frac{i}{\mathrm{cn}^2\alpha + \mathrm{cn}^2\beta}\left(2\,\mathrm{dn}\,\alpha\,\mathrm{cn}\,\beta\,\mathrm{dn}\,\beta\frac{\partial}{\partial\alpha} + \mathrm{sn}\,\alpha\,\mathrm{cn}\,\alpha\,\mathrm{sn}\,\beta\frac{\partial}{\partial\beta}\right),$$
$$K_2 = \frac{i}{\mathrm{cn}^2\alpha + \mathrm{cn}^2\beta}\left(\mathrm{sn}\,\alpha\,\mathrm{cn}\,\beta\,\mathrm{sn}\,\beta\frac{\partial}{\partial\alpha} + 2\,\mathrm{cn}\,\alpha\,\mathrm{dn}\,\alpha\,\mathrm{dn}\,\beta\frac{\partial}{\partial\beta}\right), \tag{17}$$
$$L_3 = \frac{2}{\mathrm{cn}^2\alpha + \mathrm{cn}^2\beta}\left(-\mathrm{cn}\,\alpha\,\mathrm{sn}\,\beta\,\mathrm{dn}\,\beta\frac{\partial}{\partial\alpha} + \mathrm{sn}\,\alpha\,\mathrm{dn}\,\alpha\,\mathrm{cn}\,\beta\frac{\partial}{\partial\beta}\right)$$

(we are dropping the modulus $k = k'$ of the elliptic functions). The operator (16) can be written in terms of elliptic coordinates as

$$L_s = \frac{1}{\mathrm{cn}^2\alpha + \mathrm{cn}^2\beta}\left((-1 - \mathrm{cn}^2\beta)\frac{\partial^2}{\partial\alpha^2} + (-1 + \mathrm{cn}^2\alpha)\frac{\partial^2}{\partial\beta^2}\right). \tag{18}$$

Finally, we obtain two equations for the basis functions of an irreducible representation of the $O(2, 1)$ group:

$$\Delta f_{lh}(\alpha, \beta) = -l(l + 1)f_{lh}(\alpha, \beta),$$
$$L_s f_{lh}(\alpha, \beta) = h f_{lh}(\alpha, \beta). \tag{19}$$

Taking suitable combinations of these two equations, we find that we can separate the variables and obtain two essentially identical ordinary differential equations. Thus

$$f_{lh}(\alpha, \beta) = \Lambda_{lh}(\alpha)\Lambda_{l\tilde{h}}(\beta), \tag{20}$$

satisfying

$$\frac{d^2\Lambda_{lh}(\alpha)}{d\alpha^2} + \left(h - \frac{l(l + 1)}{2}\mathrm{sn}^2\alpha\right)\Lambda_{lh}(\alpha) = 0,$$
$$\frac{d^2\Lambda_{l\tilde{h}}(\beta)}{d\beta^2} + \left(\tilde{h} - \frac{l(l + 1)}{2}\mathrm{sn}^2\beta\right)\Lambda_{l\tilde{h}}(\beta) = 0, \tag{21}$$

$$h + \tilde{h} = l(l + 1). \tag{22}$$

In (21) we recognize the Lamé equation in its Jacobian form.

As mentioned above, we are only interested in the region $\alpha \in (iK, iK + 2K)$, and so we have to study the Lamé equation over a finite interval with a singularity at each end point.

Making use of the Fuchsian theory[28] of ordinary differential equations, i.e., expanding $\Lambda_{lh}(\alpha)$ into a generalized power series about each singularity and obtaining the indicial equation, we find that the singularities are indeed regular and that the exponents at the singularities are equal to $-l$ and $l + 1$.

We shall, in the body of this paper, restrict ourselves to unitary representations of the principal continuous series, so that $l = -\frac{1}{2} + iq$ with q real. Since all solutions of the Lamé equation behave as some combination of

$$\Lambda_1(\alpha) = (\alpha - \alpha_0)^{-l}\phi_1(\alpha - \alpha_0),$$
$$\Lambda_2(\alpha) = (\alpha - \alpha_0)^{l+1}\phi_2(\alpha - \alpha_0), \tag{23}$$

where α_0 is any singular point and $\phi_{1,2}(\alpha)$ are regular there, we find that for

$$-\tfrac{3}{2} < \mathrm{Re}\, l < \tfrac{1}{2}, \tag{24}$$

in particular for $\mathrm{Re}\, l = -\frac{1}{2}$, both solutions are square integrable. Thus, both independent solutions of the Lamé equation may appear in the expansions. To distinguish between them, let us introduce two reflection operators:

$$Xv_1 = -v_1, \quad Xv_0 = v_0, \quad Xv_2 = v_2,$$
$$Yv_2 = -v_2, \quad Yv_0 = v_0, \quad Yv_1 = v_1. \tag{25}$$

Thus, the operator X corresponds to the reflection $\beta \to 2K + 2iK - \beta$ and Y to $\alpha \to 2K + 2iK - \alpha$. The operators Δ, L_s, X, and Y now form a complete set, and the basis functions for the representation will satisfy Eqs. (19) supplemented by

$$Xf_{lh}^{pq}(\alpha, \beta) = qf_{lh}^{pq}(\alpha, \beta),$$
$$Yf_{lh}^{pq}(\alpha, \beta) = pf_{lh}^{pq}(\alpha, \beta), \quad p, q = \pm 1, \tag{26}$$

and can be written as

$$f_{lh}^{pq}(\alpha, \beta) = \Lambda_{lh}^{p}(\alpha)\Lambda_{\tilde{l}\tilde{h}}^{q}(\beta). \tag{27}$$

Since there does not seem to exist an agreed standardization for the Lamé functions, we fix them completely by imposing the boundary conditions

$$\Lambda_{lh}^{+}(iK + K) = 1, \quad \Lambda_{lh}^{+\prime}(iK + K) = 0,$$
$$\Lambda_{lh}^{-}(iK + K) = 0, \quad \Lambda_{lh}^{-\prime}(iK + K) = -1. \tag{28}$$

Little is known about the Lamé functions of a real argument, still less of a complex argument, and, in particular, no explicit expressions or integral representations exist. It is amusing to note, however, that they can be considered to interpolate smoothly between exponentials (for $k = 0$) and associated Legendre functions (for $k = 1$).

Let us further remark that the Lamé functions do occur in physics—as the wavefunctions for an asymmetrical top[29] and also that they can be considered to be a special case of the Heun functions, which are the wavefunctions for a hydrogen molecular ion.[30] For further information we refer to Ref. 31.

II. CROSSING SYMMETRIC MAPPING OF MANDELSTAM PLANE ONTO AN $O(2, 1)$ HYPERBOLOID

Since our aim is to write crossing symmetric expansions in terms of the functions (27), we must construct a symmetric mapping of the Mandelstam variables s, t, and u onto the points v, i.e., onto the elliptic coordinates α and β. We shall do this in a similar fashion to how other $(s, t, u) \to v$ mappings were constructed in Ref. 4.

We choose the scattering plane to be the plane $(0, X, Z)$ so that the Y components of the 4-momenta of all four particles are equal to zero. Thus, although $p_1 \cdots p_4$ have four components each, we can still parametrize them in terms of the $O(2, 1)$ hyperboloid, putting

$$p_r = m_r(-\mathrm{cn}\,\alpha_r\,\mathrm{cn}\,\beta_r, i\,\mathrm{sn}\,\alpha_r\,\mathrm{dn}\,\beta_r, 0, i\,\mathrm{dn}\,\alpha_r\,\mathrm{sn}\,\beta_r), \tag{29}$$

where we again drop the symbol $k = k' = 2^{-\frac{1}{2}}$ in the Jacobian elliptic functions. In this paper we shall also simplify our task by taking all masses equal:

$$m_1 = m_2 = m_3 = m_4 = 1.$$

We start from the physical region of the s channel and wish to choose a frame of reference in which the interchange of s and t for fixed u will correspond to some simple exchange of α and β. Having this in mind, we must first get rid of the ten redundant parameters in (29). It proves most convenient to describe the scattering in a specifically oriented brick-wall frame (or Breit frame) of reference, as illustrated on Fig. 1. Notice that the momentum of particle 4 is as if particle 1 had perpendicularly rebounded from a brick wall parallel to the vector $\mathbf{p}_2 + \mathbf{p}_3$. The momenta (29) in this frame can be written as

$$\begin{aligned}
p_1 &= (-\mathrm{cn}\,\alpha_1\,\mathrm{cn}\,\beta_1, i\,\mathrm{sn}\,\alpha_1\,\mathrm{dn}\,\beta_1, 0, i\,\mathrm{dn}\,\alpha_1\,\mathrm{sn}\,\beta_1),\\
p_2 &= (-\mathrm{cn}\,\alpha\,\mathrm{cn}\,\beta, i\,\mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta, 0, i\,\mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta),\\
p_3 &= (-\mathrm{cn}\,\alpha\,\mathrm{cn}\,\beta, -i\,\mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta, 0, -i\,\mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta),\\
p_4 &= (-\mathrm{cn}\,\alpha_1\,\mathrm{cn}\,\beta_1, -i\,\mathrm{sn}\,\alpha_1\,\mathrm{dn}\,\beta_1, 0, -i\,\mathrm{dn}\,\alpha_1\,\mathrm{sn}\,\beta_1),
\end{aligned} \tag{30}$$

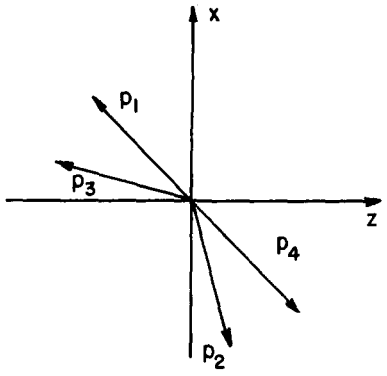

FIG. 1. The frame of reference.

with the additional constraints from the conservation laws:

$$2\,\mathrm{sn}\,\alpha_1\,\mathrm{dn}\,\beta_1 + \mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta + \mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta = 0,$$
$$2\,\mathrm{dn}\,\alpha_1\,\mathrm{sn}\,\beta_1 + \mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta + \mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta = 0. \quad (31)$$

Solving (31), we have

$$\mathrm{dn}\,\alpha_1 = \mathrm{dn}\,\beta_1 = 2^{-\frac{1}{2}}i\{[1 - \tfrac{1}{2}(\mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta + \mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta)^2]^{\frac{1}{2}} - 1\}^{\frac{1}{2}}. \quad (32)$$

Thus, all four momenta are expressed in terms of α and β only, and the scattering amplitude in the chosen frame of reference is also a function of these two kinematic variables only.

Using (30) and (32), we easily check that

$$s = (p_1 + p_2)^2 = 2 + 2[-y(1 - \tfrac{1}{2}x^2)^{\frac{1}{2}} - \tfrac{1}{2}x^2],$$
$$t = (p_1 - p_3)^2 = 2 + 2[y(1 - \tfrac{1}{2}x^2)^{\frac{1}{2}} - \tfrac{1}{2}x^2], \quad (33)$$
$$u = (p_1 - p_4)^2 = 2x^2,$$

where

$$x = \mathrm{sn}\,\alpha\,\mathrm{dn}\,\beta + \mathrm{dn}\,\alpha\,\mathrm{sn}\,\beta,$$
$$y = \mathrm{cn}\,\alpha\,\mathrm{cn}\,\beta. \quad (34)$$

Conversely, we can express the elliptic coordinates in terms of the Mandelstam variables as

$$\mathrm{cn}^4\,\alpha = \frac{(s+t)^2 + 2st(2-s-t)}{4(s+t)} + \frac{1}{2}\left(\frac{stu(s+t-st)}{s+t}\right)^{\frac{1}{2}},$$
$$\mathrm{cn}^4\,\beta = \frac{(s+t)^2 + 2st(2-s-t)}{4(s+t)} - \frac{1}{2}\left(\frac{stu(s+t-st)}{s+t}\right)^{\frac{1}{2}}. \quad (35)$$

In the physical region of the s channel, we have $4 \leq s < \infty$, $-\infty < t \leq 0$, and $-\infty < u \leq 0$, and this whole region will be described by (33) and (35) with

$$\alpha \in (iK, iK + 2K), \quad \beta \in (iK, iK + 2K). \quad (36)$$

The t channel corresponds to $4 \leq t < \infty$, $-\infty < s \leq 0$, and $-\infty < u \leq 0$ and is described by the same formulas with

$$\alpha \in (iK, iK + 2K), \quad \beta \in (-iK, -iK + 2K). \quad (37)$$

The physical region of the u channel with $4 \leq u < \infty$, $-\infty < t \leq 0$, and $-\infty < s \leq 0$ can also be described in these terms by putting

$$\alpha \in (iK, iK + 2K), \quad \beta \in (0, 2iK). \quad (38)$$

Note that the parametrization (29) with α and β in the regions (36) or (37) covers the whole upper or lower sheet of the hyperboloid $v^2 = 1$, respectively. With α and β in the u channel region (38), however, a different parametrization of the hyperboloid is necessary, namely

$$v_0 = 2^{-\frac{1}{2}}\,\mathrm{sn}\,\alpha\,(\mathrm{cn}\,\beta)/(\mathrm{dn}\,\beta),$$
$$v_1 = 2^{-\frac{1}{2}}\,\mathrm{cn}\,\alpha\,(\mathrm{sn}\,\beta)/(\mathrm{dn}\,\beta), \quad (39)$$
$$v_2 = i\,\mathrm{dn}\,\alpha/\mathrm{dn}\,\beta,$$

and the physical region of the u channel gets mapped onto one-half of the upper sheet of the hyperboloid.

Thus, each physical region gets mapped onto such a manifold that the scattering amplitude in the physical region can be expanded in terms of the basis functions of the irreducible representation of $O(2, 1)$. Our approach is so developed as to simplify crossing symmetry between the s and t channels.

Let us now proceed to write expansions of functions defined over the hyperboloid in terms of the "elliptic" basis for $O(2, 1)$ representations, after which we shall return to the problem of crossing symmetry.

III. SINGLE-VARIABLE EXPANSIONS IN TERMS OF LAMÉ FUNCTIONS

Classical Sturmian theory tells us of the existence of expansions in eigenfunctions of a differential operator $L = -d^2/dx^2 + q(x)$, defined on some closed interval of the real line. More accurately, we have to find a self-adjoint extension $T = T^*$ of L, which involves constructing a Hilbert space of functions over the interval; but this step is usually trivial in the simpler cases and is not considered in detail. In the singular Sturmian theory, however, when the interval is either infinite or has a singular point of L at one end or both, these considerations become much more involved.

The general theory of such expansions is given in Refs. 32 and 33. It is found that, when neither of the eigenfunctions of L is square integrable, then L by

itself defines a unique self-adjoint extension but that, otherwise, one or more boundary conditions have to be imposed. In the former case the spectrum of T will be continuous; in the latter, either discrete or mixed. A generally excellent account of the situation from the viewpoint or classical analysis will be found in Titchmarsh[34]; however, in this particular case, he does not apply the necessary boundary conditions to ensure that he does indeed expand in terms of a complete orthogonal set of eigenfunctions, so that his conclusions differ from those of Naimark[32] and Dunford and Schwartz,[33] with whom we are in complete agreement. Since this[34] is perhaps the best-known work on the subject, we feel that a rather more detailed presentation of our argument is called for.

We wish then to find an expansion of the scattering amplitude with one variable fixed in terms of the eigenfunctions of the formally self-adjoint formal differential operator

$$L = -\frac{d^2}{dx^2} + q(x)$$
$$= -\frac{d^2}{dx^2} + l(l+1)k^2\,\mathrm{sn}^2(x,k) \qquad (40)$$

over the interval $I = [iK' + K, iK' + 2K)$, the upper end of which is a singular point of L; the other half of the interval $(iK', iK' + K)$ will be treated by symmetry considerations. Denote by $L^2(I)$ and $C^k(I)$ the spaces of Lebesque-square-integrable and of k-times-differentiable functions over I, and define the space

$$H_L = \{f \in C^2(I) \,|\, f, Lf \in L^2(I)\}$$

and the operator T_1 in $L^2(I)$,

$$D(T_1) = H_L, \quad T_1 f = Lf.$$

Because T_1 is a closed operator, H_L becomes a Hilbert space $\mathcal{H}_L$ upon the introduction of the inner product (and, of course, completion with respect thereto)

$$(f, g)^* = (f, g) + (T_1 f, T_1 g), \qquad (41)$$

where

$$(f, g) = \int_I f(x)g(x)\,dx; \qquad (42)$$

however, T (that is, its unique extension) is not necessarily self-adjoint in this space. It can be shown indeed[32,33] that it is so automatically only if neither of the eigenfunctions of L is square integrable in a neighborhood of either end of I; otherwise, we must restrict the domain of T_1 by imposing boundary conditions. In the case at hand, we know from Lamé's differential equation that for $-\frac{3}{2} < \mathrm{Re}\, l < \frac{1}{2}$ both solutions are $L^2(I)$, and hence we need two boundary conditions.

Let us introduce H_0, which is the space of all functions of H_L which vanish in some open neighborhood of the two end points of I. Then by a boundary value for L we mean a continuous linear functional $\mathcal{A}$ on H_L that vanishes on H_0; this can always be written as

$$\mathcal{A}(f) = (Lf, A) - (f, LA) \qquad (43)$$

for some $A \in H_L$. It is convenient here to introduce the function A_+, which is defined to be any C^∞ function equal to $(iK' + 2K - x)^{l+1}$ in a neighborhood of $iK' + 2K$ and vanishing in a neighborhood of $iK' + K$, and the function $A_- = A_+\,(l \to -l - 1)$: Then these define the boundary values

$$\mathcal{A}_+(f) = \lim_{\epsilon \to 0} [\epsilon^{l+1} f'(E - \epsilon) - (l+1)\epsilon^l f(E - \epsilon)],$$
$$\mathcal{A}_-(f) = \lim_{\epsilon \to 0} [\epsilon^{-l} f'(E - \epsilon) + l\epsilon^{-l-1} f(E - \epsilon)], \qquad (44)$$

where we have set $iK' + 2K = E$. Because of the relations

$$\mathcal{A}_+(A_+) = 0 = \mathcal{A}_-(A_-),$$
$$\mathcal{A}_-(A_+) = 2l + 1 = -\mathcal{A}_+(A_-), \qquad (45)$$

these are independent, and so we can write any boundary condition at E as

$$(\sin\theta \mathcal{A}_+ + \cos\theta\, \mathcal{A}_-)(f) = 0 \qquad (46)$$

for some fixed θ independent of h. The other end of the interval $iK' + K$ is associated with no singularities, and so here we can set

$$f \sin\theta' + f' \cos\theta' = 0. \qquad (47)$$

A convenient choice of θ' is 0 or $\frac{1}{2}\pi$: This corresponds to choosing a solution of Lamé's equation which is even or odd about $iK' + K$; we have already introduced such solutions and denoted them $\Lambda^+_{lh}(x)$ or $\Lambda^-_{lh}(x)$. It is clear that we could specify these as well by imposing boundary conditions at iK' instead of $iK' + K$.

Now define the space $\mathcal{H} = \{f \in \mathcal{H}_L \,|\, f$ satisfies (46), (47).$\}$ Then the restriction $T = T_1 \,|\, \mathcal{H}$ of T_1 to $\mathcal{H}$ is the self-adjoint extension of L that we have been seeking, and the eigenfunctions of L which satisfy (46) and (47) form a basis of $\mathcal{H}$. Equation (46) is then just a transcendental equation for h that defines discrete characteristic values[35] of this parameter.

A reasonably fair summary of the procedure is therefore:

(i) Choose any value of h, say h_0;
(ii) select all values h such that $(\Lambda^+_{lh}, \Lambda^+_{lh_0}) = 0$;
(iii) these form a complete set.

The actual value of h_0 in (i) will be determined by the boundary condition (46)—or vice versa. An identical procedure can be carried through for that part of $f \in \mathcal{K}_L$ that satisfies (46) but is odd about $iK' + K$; but, of course, we do not expect the set of values of h to be that found earlier. We can then at last write

$$f(x) = \sum_{p,h} N_{lh}^{p} \Lambda_{lh}^{p}(x) \tilde{f}_{p}^{h}, \tag{48}$$

$$\tilde{f}_{p}^{h} = \int_{iK'}^{iK'+2K} f(x) \Lambda_{lh}^{p}(x) \frac{dx}{\sqrt{2}}, \tag{49}$$

where

$$(N_{lh}^{p})^{-1} = \int_{iK'}^{iK'+2K} [\Lambda_{lh}^{p}(x)]^2 \frac{dx}{\sqrt{2}}; \tag{50}$$

the label p stands for the parity ($\pm$) and the set $\{h\}$ depends upon both p and l. The factors of $\sqrt{2}$ are inserted for convenience. The expansion is convergent[33] in the topology of $C^2(I)$ defined by the norm

$$\|f\| = \sum_{k=0}^{\infty} 2^{-k} \frac{|(f \mid J_k)|}{1 + |(f \mid J_n)|},$$

where $\{J_k\}$ is an increasing sequence of closed intervals of I whose union is I and $f \mid J$ is the restriction of f to J.

Let us consider these formulas. The most striking thing about them is that despite I being an open interval with singular points at either end—corresponding to an infinite path on the hyperboloid—nonetheless, the spectrum is discrete. This contrasts sharply with the situation found when the operator L is defined by some subgroup of a symmetry group of the problem, and at first sight is somewhat surprising; nonetheless, it follows immediately from the general theory (cf. Ref. 33, Sec. XIII, 10-C-1) once we notice that both solutions of Lamé's equation are $L^2(I)$, and we have gone into such detail merely to show whence the result comes: The two key steps are (a) imposing boundary conditions (to make L symmetric) and (b) introducing the Hilbert space $\mathcal{K}$ (to make it self-adjoint). An intuitive but convincing argument is that when all the eigenfunctions are square integrable, then there is no mechanism for a divergence to occur to produce a δ function—and hence indeed the expansion must be discrete.

IV. TWO-VARIABLE EXPANSIONS

We now return to the main problem, which is to expand a function of two variables in terms of a product of Lamé functions. Since this is defined over a hyperboloid, the most convenient way to do this is to make use of the Gel'fand–Graev pair of integral transforms, which we now briefly explain.

The Radon transform in n-dimensional Euclidean space is an integral over all possible hyperplanes; the Gel'fand–Graev transform is a generalization of this integral transform to any homogeneous manifold, and the role of the planes is taken by the horospheres of the space. For the hyperboloid $v^2 = 1$ these are merely the cross sections of the hyperboloid satisfying $v \cdot k = 1$, where k is any point on the associated light cone $k^2 = 0$—that is, they are the surfaces of zero curvature in the Lobachevskii geometry. The integral transform maps any function defined on the hyperboloid into one defined on the cone, and, on the cone, it is meaningful to expand in terms of homogeneous functions (i.e., carry out a Mellin transform parallel to a generator), thus effecting a decomposition into functions transforming under a UIR of the associated group of motions. Since this is unaffected by the transform, by first moving to the cone, decomposing, and then transforming back to the hyperboloid we have a convenient method for Fourier analyzing over the group any function defined on its homogeneous space.

For the case in hand, these transforms can be written

$$f(v) = \frac{1}{8\pi^2 i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dl\, l \cot \pi l \int_{\Gamma} \Phi(k, l)(v \cdot k)^{-l-1}\, dk, \tag{51}$$

$$\Phi(k, l) = \int f(v)(v \cdot k)^{l} \frac{d^2 v}{v_0}; \tag{52}$$

the integration contour Γ is an arbitrary path on the cone which intersects every generator once; dk is the invariant measure on this path, defined by $d(tk) = dt\, dk$, where $d(tk) = dk_1\, dk_2/k_0$ is that on the cone. For further details, we cite Refs. 1, 26, 36, and 37.

Now such a path Γ is given by the quadruplet

$$\begin{aligned} \Gamma^{1\pm} &= (i \operatorname{cn} \theta,\ i \operatorname{dn} \theta,\ \pm 2^{-\frac{1}{2}} \operatorname{sn} \theta), \\ \Gamma^{2\pm} &= (i \operatorname{cn} \theta,\ \pm 2^{-\frac{1}{2}} \operatorname{sn} \theta,\ i \operatorname{dn} \theta), \\ \theta &\in (iK',\ iK' + 2K) \end{aligned} \tag{53}$$

(the modulus of the elliptic functions is, of course, $k = 2^{-\frac{1}{2}}$); a stereographic projection of this is given in Fig. 2. On each of the subpaths we expand the function $\Phi(k^{i\pm}, l)$ by the Lamé transform (48); then

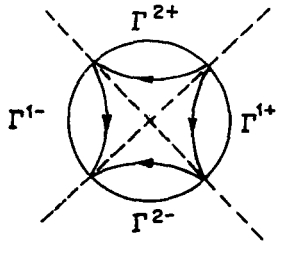

FIG. 2. The integration contour Γ on the cone in stereographic projection.

for Γ^{1+} the contour integral in (51) becomes

$$\int_{\Gamma^{1+}} \Phi(k, l)(v, k)^{-l-1}\, dk$$
$$= \sum_{p,h} N^p_{lh} A^p_{1+}(l, h) \int_I \Lambda^p_{lh}(\theta) \times \left(-ic\alpha c\beta c\theta + s\alpha\, d\beta\, d\theta - \frac{i}{\sqrt{2}} d\alpha s\beta s\theta\right)^{-l-1} \frac{d\theta}{\sqrt{2}}, \tag{54}$$

where the subscripts on A indicate that it is the relevant function for this subpath.

Consider the integral over θ in (54). It is clear that it converges uniformly for all α and β not in a neighborhood of the singular points iK' and $iK' + 2K$; so we can apply a differential operator and take it under the integral sign. After some tedious algebra we find that the kernel $J = \{\ \}^{-l-1}$ of the integral satisfies the differential equations

$$L_\beta J = L_\theta J = -[L_\alpha - l(l+1)]J,$$

where L_x is the operator (40), so that we can replace $L_\beta J$ by $L_\theta J$, carry out two partial integrations, and conclude that (54) satisfies Lamé's equation in β provided that the integral converges and

$$(\Lambda J' - \Lambda' J)(\theta)\Big|_{iK}^{iK+2K} = 0.$$

The convergence is assured, but the bilinear concomitant just oscillates if $\operatorname{Re} l = -\frac{1}{2}$. We therefore "regularize" the integral (51) by giving the exponent $-l-1$ a small negative increment.

A similar result holds in α, except that now the eigenvalue h is replaced by $\tilde{h} = l(l+1) - h$; so that we can write the integral as

$$\int \Lambda^p_{lh}(\theta) J(\alpha, \beta, \theta)\, d\theta = \sum_{p',q} \lambda^{p'}_{pq}(l, h) \Lambda^{p'}_{lh}(\beta) \Lambda^q_{l\tilde{h}}(\alpha), \tag{55}$$

where the λ's are certain coefficients. Inspection of the parity under $\beta \to 2K + 2iK' - \beta$ tells us that λ vanishes unless $p = p'$, and by using the standardization (28) of our Lamé functions we can write explicitly

$$\lambda_{++} = \frac{1}{\sqrt{2}} \int \Lambda^+_{lh}(\theta)(i\ \mathrm{cn}\ \theta)^{-l-1}\, d\theta,$$
$$\lambda_{+-} = \frac{(l+1)}{2} \int \Lambda^+_{lh}(\theta)(i\ \mathrm{cn}\ \theta)^{-l-2}\, \mathrm{sn}\ \theta\, d\theta,$$
$$\lambda_{-+} = \frac{(l+1)}{\sqrt{2}} \int \Lambda^-_{lh}(\theta)(i\ \mathrm{cn}\ \theta)^{-l-2}(i\ \mathrm{dn}\ \theta)\, d\theta, \tag{56}$$
$$\lambda_{--} = \frac{(l+1)(l+2)}{2} \int \Lambda^-_{lh}(\theta)(i\ \mathrm{cn}\ \theta)^{-l-3}(i\ \mathrm{dn}\ \theta\ \mathrm{sn}\ \theta)\, d\theta,$$

where the integrations are, of course, over $iK' < \theta < iK' + 2K$.

We carry out this program for each of the four subpaths $\Gamma^{i\pm}$; but, before writing down the expansion formula obtained, let us examine the projection formulas. These we find, by substituting (52) into (49),

$$A^p_{i\pm}(l, h) = \int \frac{d^2v}{v_0} f(v) \int \Lambda^p_{lh}(\theta)[v \cdot k^{i\pm}(\theta)]^l \frac{d\theta}{\sqrt{2}}, \tag{57}$$

and the integral over θ once again gives a product of Lamé functions, being just the complex conjugate of the corresponding integral in (54). Let us define the coefficients

$$2A^{pq}_i(l, h) = A^p_{i+}(l, h) + qA^p_{i-}(l, h); \tag{58}$$

then we find, after some straightforward manipulations,

$$A^{pq}_1(l, h) = \lambda_{pq}(l, h)^* \int f(\alpha, \beta) \Lambda^p_{lh}(\beta) \Lambda^q_{l\tilde{h}}(\alpha) \frac{d^2v}{v_0},$$
$$A^{pq}_2(l, h) = \lambda_{pq}(l, h)^* \int f(\alpha, \beta) \Lambda^p_{lh}(\alpha) \Lambda^q_{l\tilde{h}}(\beta) \frac{d^2v}{v_0}. \tag{59}$$

After some trivial redefinitions we can add together the contributions to (51) of all four subpaths, and make use of (55), to obtain at last the pair of transforms

$$A^{pq}(l, h) = -\tfrac{1}{2} \int d\alpha\, d\beta (\mathrm{cn}^2\, \alpha + \mathrm{cn}^2\, \beta) \times \Lambda^p_{lh}(\alpha) \Lambda^q_{l\tilde{h}}(\beta) f(\alpha, \beta), \tag{60}$$

$$f(\alpha, \beta) = \frac{1}{8\pi^2 i} \int dl (2l+1) \cot \pi l \times \sum_h \sum_{p,q} N^p_{lh} |\lambda_{pq}(l, h)|^2 \times [\Lambda^p_{lh}(\alpha) \Lambda^q_{l\tilde{h}}(\beta) A^{pq}(l, h) + \Lambda^q_{l\tilde{h}}(\alpha) \Lambda^p_{lh}(\beta) A^{qp}(l, \tilde{h})], \tag{61}$$

for $\alpha \in (iK, iK + 2K)$ and $\beta \in (iK, iK + 2K)$.

Quite similarly, we can derive expansions for the region $\alpha \in (iK, iK + 2K)$, $\beta \in (-iK, -iK + 2K)$. They coincide with (60) and (61), except that $\Lambda^q_{l\tilde{h}}(\beta)$ and $\Lambda^p_{lh}(\beta)$ are replaced by $\tilde{\Lambda}^q_{l\tilde{h}}(\beta)$ and $\tilde{\Lambda}^q_{lh}(\beta)$. The new functions are again solutions of the Lamé equation, this time standardized at $\beta = -iK + K$ to be

$$\tilde{\Lambda}^+_{lh}(-iK + K) = 1, \quad \tilde{\Lambda}^{+\prime}_{lh}(-iK + K) = 0,$$
$$\tilde{\Lambda}^-_{lh}(-iK + K) = 0, \quad \tilde{\Lambda}^{-\prime}_{lh}(-iK + K) = -1. \tag{62}$$

Let us note that, for $f(\alpha, \beta)$ satisfying

$$-\tfrac{1}{2} \int |f(\alpha, \beta)|^2 (\mathrm{cn}^2\, \alpha + \mathrm{cn}^2\, \beta)\, d\alpha\, d\beta < \infty,$$

only the unitary representations of the principal continuous series figure in (61), i.e., $l = -\frac{1}{2} + iq$, q real. The discrete unitary representations are absent, because they are not realized on the homogeneous space $v_0^2 - v_1^2 - v_2^2 = 1$. If we were to expand functions defined over the whole group manifold, which would be necessary for particles with spin, some representations of the discrete series would be present.

V. REMARKS ON CROSSING SYMMETRY

We shall now apply the expansion formulas (60) and (61) to expand physical scattering amplitudes. Before doing this, let us make use of some symmetries of the mapping (33) and (35) from s, t, and u to α and β.

First, this mapping is left invariant by reflecting α and β simultaneously in the center point of their region of definition:

$$\alpha \to 2K + 2iK - \alpha, \quad \beta \to 2K + 2iK - \beta, \qquad s \text{ channel},$$

$$\alpha \to 2K + 2iK - \alpha, \quad \beta \to 2K - 2iK - \beta, \qquad t \text{ channel}.$$

Thus, these points correspond to the same values of s, t, and u, and we can put

$$f(\alpha, \beta) = f(2iK + 2K - \alpha, 2iK + 2K - \beta) \quad (63)$$

or, respectively,

$$f(\alpha, \beta) = f(+2iK + 2K - \alpha, -2iK + 2K - \beta). \quad (63')$$

Using (60) and (61), we can easily see that such a symmetry of the scattering amplitude implies that the "Lorentz amplitude" $A^{pq}(l, h)$ satisfies

$$A^{pq}(l, h) = \delta_{pq} A^p(l, h). \quad (64)$$

Second, this mapping is also left invariant by the interchange

$$\alpha \to \beta, \quad \beta \to \alpha,$$

so that

$$f(\alpha, \beta) = f(\beta, \alpha), \quad (65)$$

implying that we can put

$$A^p(l, h) = A^p(l, \tilde{h}). \quad (66)$$

A complete consideration of the crossing transformation involves an analytic continuation of the scattering amplitude from one channel into the other, i.e., from (α, β) in region (36) to, say, (37). It is a difficult task to discuss the convergence problems involved in the analytic continuation of expansion (61), and in this paper we avoid the problem altogether.

Thus, let us consider two (*a priori*) different scattering amplitudes $f^s(s, t, u)$ and $f^t(s, t, u)$, defined in the s- and t-channel physical regions, respectively. Independently of whether or not they are two different pieces of one and the same analytic function, we can consider a crossing symmetric reaction in which

$$1 + 2 \to 3 + 4, \quad 1 + \bar{3} \to \bar{2} + 4$$

are identical processes, i.e., particle 3 is the antiparticle of 2. For such a reaction we must have

$$f^s(s, t, u) = \pm f^t(t, s, u) \quad (67)$$

(each amplitude is evaluated at a definite point in its own physical region).

Now expand each of the amplitudes:

$$f^s(s, t, u) = f^s(\alpha, \beta) = \frac{1}{8\pi^2 i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dl(2l + 1) \cot \pi l \times \sum_h \sum_p N^p_{lh} |\lambda_{pp}(l, h)|^2 A^p(l, h) \times [\Lambda^p_{lh}(\alpha)\Lambda^p_{l\tilde{h}}(\beta) + \Lambda^p_{l\tilde{h}}(\alpha)\Lambda^p_{lh}(\beta)], \quad (68)$$

$$\alpha \in (iK, iK + 2K), \quad \beta \in (iK, iK + 2K),$$

$$f^t(s, t, u) = f^t(\alpha, \tilde{\beta}) = \frac{1}{8\pi^2 i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dl(2l + 1) \cot \pi l \times \sum_h \sum_p N^p_{lh} |\lambda_{pp}(l, h)|^2 B^p(l, h) \times [\Lambda^p_{lh}(\alpha)\tilde{\Lambda}^p_{\tilde{h}}(\tilde{\beta}) + \Lambda^p_{l\tilde{h}}(\alpha)\tilde{\Lambda}^p_{lh}(\tilde{\beta})], \quad (69)$$

$$\alpha \in (iK, iK + 2K), \quad \tilde{\beta} \in (-iK, -iK + 2K).$$

For a crossing symmetric reaction the two amplitudes must satisfy (67). As can be seen by examining the kinematic formulas (33)–(35), crossing symmetry can be ensured by putting

$$f^s(\alpha, \beta) = \pm f^t(\beta, 2K - \alpha) \quad (70)$$

or, alternatively,

$$f^s(\alpha, \beta) = \pm f^t(\alpha, 2K - \beta). \quad (71)$$

However, we have

$$\tilde{\Lambda}^p_{lh}(\tilde{\beta}) = p\Lambda^p_{lh}(\beta), \quad \tilde{\beta} = 2K - \beta. \quad (72)$$

Combining (68)–(72), we find that crossing symmetry is implied by

$$A^p(l, h) = \pm pB^p(l, h). \quad (73)$$

The "elliptic Lorentz amplitudes" in the two channels are given by almost identical formulas

$$A^p(l, h) = -\tfrac{1}{2} \int_{iK}^{iK+2K} d\alpha \int_{iK}^{iK+2K} d\beta (\mathrm{cn}^2 \alpha + \mathrm{cn}^2 \beta) \times \Lambda^p_{lh}(\alpha)\Lambda^p_{l\tilde{h}}(\beta) f^s(\alpha, \beta), \quad (74)$$

$$B^p(l, h) = -\tfrac{1}{2} p \int_{iK}^{iK+2K} d\alpha \int_{iK}^{iK+2K} d\beta (\mathrm{cn}^2 \alpha + \mathrm{cn}^2 \beta) \times \Lambda^p_{lh}(\alpha)\Lambda^p_{l\tilde{h}}(\beta) f^t(\alpha, 2K - \beta). \quad (75)$$

Let us consider these further. By using (72) and (73), Eqs. (68) and (69) become identical, modulo possible minus signs: In particular, the arguments of the Lamé functions all lie in the same range $(iK, iK + 2K)$, where they are given in terms of s and t by (35) with a unique prescription for extracting the fourth root. Therefore, unlike conventional single- or double-variable expansions, our single formula (68) converges simultaneously in both the s- and t-channel physical regions with a single mapping of $s, t \to \alpha, \beta$. The expansions are not merely crossing symmetric—they are explicitly so, in the sense that s and t appear symmetrically throughout. This is the property we have been seeking.

Analyticity of the scattering amplitude would now imply that $f^s(\alpha, \beta)$ and $f^t(\alpha, \beta)$ are the same function $f(\alpha, \beta)$ and that its values in the two channels are connected by some sort of dispersion relation.

VI. DISCUSSION

By choosing a Breit frame of reference for the scattering process, we have succeeded in developing a mapping $s, t \to \alpha, \beta$ of such symmetry that the expansion formulas are explicitly symmetric in the Mandelstam variables s and t. The essential reason for this was that both the s- and t-channel physical regions were mapped onto hyperboloids in such a way that the order of the pair $\{s, t\}$ was irrelevant: Hence, the two manifolds were identifiable and gave identical expansion theorems, upon which crossing symmetry could be imposed term by term.

Notice, however, that the Lamé functions of β in (68) and (69), corresponding to the s and t channels, respectively, are not the same Lamé functions; but, because their arguments lie in the appropriate ranges, we could equate their numerical values by (72). This was the final step in our argument; it corresponds basically to inserting by hand a factor of (-1) in the parametrization of $\bar{p}_2^0$ and $\bar{p}_3^0$ (the energies of the antiparticles in the t channel) and keeping the ranges of α and β unaltered. For the imposition of crossing symmetry this is extremely convenient—but it has to be reconciled with analyticity, which must start with (68) alone and by analytic continuation in β reach (69).

Let us note that, unlike other 2-variable expansions (see, e.g., Refs. 38–40) also written with the purpose of incorporating crossing symmetry (and using a completely different approach), our expansions are written directly in the physical regions.

From a mathematical viewpoint, the interest of this paper lies in the beginnings of a treatment of $SO(2, 1)$ in a basis which does not correspond to any subgroup at all. As far as we are aware, no group has been so treated before. We have obtained the "spherical functions" of the group in the elliptic basis and have written down the expansion formulas. These latter contain several new features: Notice, in particular, the discrete spectrum of h, when we might have expected a continuous one, and the "essentially 2-variable" nature of the expansions, which cannot be derived by an expansion in one variable followed by an expansion of the coefficient in terms of the other.

Let us briefly mention future developments. First of all, we hope to return to a more thorough investigation of the crossing symmetric expansions (60) and (61), mainly to investigate the analyticity properties, threshold behavior, asymptotic behavior, possible kinematic constraints (especially after incorporating reactions with nonequal mass particles), etc. We also wish to apply these expansions to analyze experimental data, which can only be done if methods of truncating and approximating the sums and integrals are devised. To achieve this, a further study of the properties of Lamé functions is required.

Secondly, since the scattering amplitude is actually a function of 4-dimensional momenta, our use of the $O(2, 1)$ group is somewhat artificial and expansions based on an $O(3, 1)$ group would be of even more interest (and also more in the spirit of previous work[1–7]). The threshold and asymptotic behavior based on the $O(3, 1)$ group would be different and could be hoped to correspond more simply to physical expectations.

Finally, a further development of group representation theory in elliptic bases would be of interest, in particular, a calculation of the matrix elements of infinitesimal and finite transformations of the group.

ACKNOWLEDGMENTS

We should like to thank Professor J. Ashkin for his kind hospitality at the Physics Department of Carnegie-Mellon University and Professor I. Bialynicki-Birula, Professor C. V. Coffman, Professor R. E. Cutkosky, and Professor J. S. W. Wong for interesting discussions.

APPENDIX: ANOTHER ELLIPTIC PARAMETRIZATION

Let us set

$$\begin{aligned} v_0 &= (1/k') \operatorname{dn} \alpha \operatorname{dn} \beta, \\ v_1 &= -ik \operatorname{sn} \alpha \operatorname{sn} \beta, \\ v_2 &= (k/k') \operatorname{cn} \alpha \operatorname{cn} \beta. \end{aligned} \tag{A1}$$

This covers the entire upper sheet of $v_0^2 - v_1^2 - v_2^2 = 1$ when the variables take the range

$$\alpha \in [0, 4K], \quad \beta \in [0, iK']; \tag{A2}$$

notice that the curves of constant β are closed. The generators of the group are (using a shortened notation for the elliptic functions)

$$J_3 = \frac{i}{k'}(S^2\beta - S^2\alpha)^{-1}\left(d\alpha C\beta S\beta \frac{\partial}{\partial\alpha} - C\alpha S\alpha\, d\beta \frac{\partial}{\partial\beta}\right),$$

$$K_1 = \frac{ik'}{k}(S^2\beta - S^2\alpha)^{-1}\left(C\alpha S\beta\, d\beta \frac{\partial}{\partial\alpha} - S\alpha\, d\alpha C\beta \frac{\partial}{\partial\beta}\right),$$

$$K_2 = \frac{1}{k}(S^2\beta - S^2\alpha)^{-1}\left(S\alpha C\beta\, d\beta \frac{\partial}{\partial\alpha} - C\alpha\, d\alpha S\beta \frac{\partial}{\partial\beta}\right), \tag{A3}$$

and the Laplace operator in the homogeneous space separates to give two identical Lamé equations (21).

We look now for an expansion over the cycles $\beta = \text{const}$. The boundary condition imposed earlier is now replaced by the periodicity condition $f(\theta) = f(\theta + 4K)$, which implies that h is restricted to a set of discrete characteristic values, and so we can write

$$f(\theta) = \sum_{mp} N_m^p(l) E p_l^m(\theta) \tilde{f}_p^m(l),$$

$$\tilde{f}_p^m(l) = \int_0^{4K} f(\theta) E p_l^m(\theta) \frac{d\theta}{\sqrt{2}}. \tag{A4}$$

Here we have used the traditional notation[31] for the periodic transcendental Lamé functions of a real argument, except that the letter p now stands for the index set $\{c, s\}$ corresponding exactly to our $\{+, -\}$, the parity about the origin. The label m runs over all nonnegative integers: It signifies the number of zeros of the Lamé function in $(0, 2K)$ and determines $h(l)$ uniquely through the periodicity condition.

The Gel'fand–Graev transform is much simplified because a single contour (dn θ, $2^{-\frac{1}{2}}$ sn θ, cn θ) encircles the cone; we take $k = k'$ and proceed as before to obtain

$$A^p(l, m) = \int d\alpha\, d\beta (\mathrm{sn}^2\alpha - \mathrm{sn}^2\beta) \times E p_l^m(\alpha) E p_l^m(\beta) f(\alpha, \beta), \tag{A5}$$

$$f(\alpha, \beta) = \frac{1}{8\pi^2 i} \int dl\, l \cot \pi l \sum_{mp} N_m^p(l) |\lambda_p^m(l)|^2 \times E p_l^m(\alpha) E p_l^m(\beta) A^p(l, m), \tag{A6}$$

where the constants λ are defined much as before:

$$\lambda_c^m(l) = \frac{1}{\sqrt{2}} \int_0^{4K} E c_l^m(\theta)(\sqrt{2}\,\mathrm{dn}\,\theta - \mathrm{cn}\,\theta)^{-l-1}\, d\theta,$$

$$\lambda_s^m(l) = \frac{-i(l+1)}{2\sqrt{2}} \int_0^{4K} E s_l^m(\theta) \times (\sqrt{2}\,\mathrm{dn}\,\theta - \mathrm{cn}\,\theta)^{-l-2}\,\mathrm{sn}\,\theta\, d\theta, \tag{A7}$$

if we standardize the Lamé functions at the origin.

The greater simplicity of this coordinate system is, however, destroyed by the mapping $s, t \to \alpha, \beta$. Defining

$$2A = 2^{-\frac{1}{2}} i S\alpha S\beta - C\alpha C\beta,$$
$$B = d\alpha\, d\beta,$$

we find

$$s = 2(1 + B)(1 + 2A^2),$$
$$t = 2(1 - B)(1 + 2A^2), \tag{A8}$$
$$u = -8A^2,$$

which are simple; but the inverse mapping is not so. We find that dn $(\alpha, 2^{-\frac{1}{2}})$ is a root of

$$2x^8 - 3x^6(1 + 2B^2) + x^4[4B^4 + 9B^2 + 1 + \tfrac{1}{2}(2A^2 - B^2 + \tfrac{1}{2})^2] - 3x^2B^2(1 + 2B^2) + 2B^4 = 0, \tag{A9}$$

and we cannot solve this explicitly. We have been unable to find an alternative mapping which is satisfactory in all respects.

* Now at DAMTP, University of Cambridge, Cambridge, England.
† On leave of absence from Nuclear Research Institute, Czechoslovak Academy of Sciences, Řež, Czechoslovakia. Present address: Department of Physics, University of Pittsburgh, Pittsburgh, Pa. 15213

[1] N. Vilenkin and J. Smorodinsky, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys. JETP **19**, 1209 (1964)].
[2] P. Winternitz and I. Friš, Yad. Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)].
[3] P. Winternitz, I. Lukač, and J. Smorodinsky, Yad. Fiz. **7**, 192 (1968) [Sov. J. Nucl. Phys. **7**, 139 (1968)].
[4] P. Winternitz, J. Smorodinsky, and M. Sheftel, Yad. Fiz. **7**, 1325 (1968) [Sov. J. Nucl. Phys. **7**, 785 (1968)].
[5] P. Winternitz, J. Smorodinsky, and M. Sheftel, Yad. Fiz. **8**, 833 (1968) [Sov. J. Nucl. Phys. **8**, 485 (1969)].
[6] P. Pajas and P. Winternitz, Phys. Rev. D **1**, 1105 (1970).
[7] P. Winternitz, Lectures at the Summer School on Group Representations and Quantum Theory, Dublin, Ireland, 1969; Preprint RPP/T/3, Rutherford Laboratory, England.
[8] H. Joos, Fortschr. Physik **10**, 65 (1962).
[9] M. Toller, Nuovo Cimento **53A**, 671 (1968); **54A**, 295 (1968).
[10] J. F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, Preprint IC/67/9, Trieste, 1967.
[11] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825, 1017 (1950).
[12] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, Oxford, 1964).
[13] D. W. Duc and N. V. Hieu, Ann. Inst. Henri Poincaré **A6**, 17 (1967).
[14] S. Ström, Arkiv Fysik **29**, 467 (1967).
[15] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[16] N. Mukunda, J. Math. Phys. **9**, 50 (1968).
[17] R. Delbourgo, K. Koller, and P. Mahanta, Nuovo Cimento **52**, 1254 (1967).
[18] N. W. Macfadyen, Nuovo Cimento **62A**, 925 (1969).
[19] M. W. Macfadyen, J. Math. Phys. (to be published).
[20] M. Liberman and A. Makarov, Yad. Fiz. **9**, 1314 (1969) [Sov. J. Nucl. Phys. **9**, 766 (1969)].
[21] V. Bargmann, Ann. Math. **48**, 568 (1947).
[22] A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8**, 52 (1968).
[23] N. Mukunda, J. Math. Phys. **10**, 2086, 2092 (1969).
[24] J. Smorodinsky and I. Tugov, Zh. Eksp. Teor. Fiz. **50**, 653 (1966) [Sov. Phys. JETP **23**, 434 (1966)].
[25] M. N. Olevsky, Mat. Sb. **27** (69), 379 (1950) (In Russian).

[26] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, R.I., 1968).
[27] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II, Chap. 13.
[28] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[29] I. Lukač and J. Smorodinsky, Zh. Eksp. Teor. Fiz. **57**, 1342 (1969) [Sov. Phys. JETP **30**, 728 (1970)].
[30] A. Lemieux and A. K. Bose, Ann. Inst. Henri Poincaré **10**, 259 (1969).
[31] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. III, Chap. 15.
[32] M. A. Naimark, *Linear Differential Operators* (Moscow, 1954) [German translation: (Akademie-Verlag, Berlin, 1967)].
[33] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1964), Part II.
[34] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Differential Equations* (Clarendon, Oxford, 1946).
[35] It is actually possible to make these values even more explicit by a slight modification of Titchmarsh's integral-equation method (Ref. 34, Chap. 5.4) so that it includes his (excluded) limiting case $q'(x) = O([q(x)]^{\frac{3}{2}})$; but the highly transcendental result is not particularly illuminating, and we do not give it here. We mention it only to emphasize that such a modification does exist.
[36] I. M. Gel'fand and M. I. Graev, Tr. Mosk. Mat. Obshch. **11**, 243 (1962) (In Russian).
[37] I. M. Gel'fand, M. I. Graev, and N. J. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[38] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).
[39] A. P. Balachandran, W. J. Meggs, P. Ramond, and J. Nuyts, Phys. Rev. **187**, 2080 (1969).
[40] J. M. Charap and B. M. Minton, J. Math. Phys. **10**, 1823 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Analytic Properties of Nonstrictly Localizable Fields*

F. CONSTANTINESCU†
Department of Physics, University of Munich, Germany

(Received 2 March 1970)

A class of test functions "minimal with respect to causality" is introduced. The corresponding fields are called local. Tempered and strictly localizable fields are local, but there is a large class of fields that are local but not strictly localizable. For local fields, the analytic properties of vacuum expectation values are studied. The local fields that are not strictly localizable are characterized by an arbitrary fast increase of Wightman functions near the light cone. With an adequate definition of local commutativity, other properties of tempered and strictly localizable fields seem to remain valid.

1. INTRODUCTION

In the usual Wightman framework, great attention has been paid to the investigations of the theoretical and experimental consequences of the following basic requirements of relativistic quantum field theories[1–3]:

(a) Hilbert space of states,

(b) covariance of the fields under the inhomogeneous Lorentz group,

(c) positive energy,

(d) local commutativity of fields,

(e) particle interpretation.

On the basis of the two first requirements it follows that a field $A(x)$ will be an operator-valued generalized function.[3,4] In order to obtain an operator, A must be averaged with a smooth test function $f(x)$:

$$A(f) = \int A(x)f(x)\,dx. \tag{1}$$

At this point we have to answer the following question: How to take the test functions? It was pointed out[1–3] that some basic properties of relativistic fields can be obtained if we choose tempered test functions. The requirement of temperedness appears natural from a physical viewpoint because it reflects the symmetry between coordinate and momentum spaces. At the same time, with this requirement, one can prove that the scattering amplitude $F(s, t)$ is analytic in s (for fixed $t < 0$) in a cut plane and has a polynomial behavior.

In a series of remarkable papers, Jaffe[3] has shown that the requirements (a)–(e) can be incorporated in a theory of strictly localizable fields with test functions which allow, in momentum space, a nontempered increase of fields. The wider class of fields studied by Jaffe is physically relevant since it allows for the possibility that the off-mass-shell amplitudes can grow, at large energies, faster than any polynomial. Besides temperate fields, strictly localizable fields include a class of nonrenormalizable interactions.[5] Entire functions of free fields are other examples.

We say that a field $A(x)$ is strictly localizable in a certain region of space–time if A can be averaged with some test function $f(x)$ which vanishes outside this region. Such a notion is very convenient for the following statement of local commutativity: A field A obeys

[26] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, R.I., 1968).
[27] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II, Chap. 13.
[28] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[29] I. Lukač and J. Smorodinsky, Zh. Eksp. Teor. Fiz. **57**, 1342 (1969) [Sov. Phys. JETP **30**, 728 (1970)].
[30] A. Lemieux and A. K. Bose, Ann. Inst. Henri Poincaré **10**, 259 (1969).
[31] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. III, Chap. 15.
[32] M. A. Naimark, *Linear Differential Operators* (Moscow, 1954) [German translation: (Akademie-Verlag, Berlin, 1967)].
[33] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1964), Part II.
[34] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Differential Equations* (Clarendon, Oxford, 1946).
[35] It is actually possible to make these values even more explicit by a slight modification of Titchmarsh's integral-equation method (Ref. 34, Chap. 5.4) so that it includes his (excluded) limiting case $q'(x) = O([q(x)]^{\frac{3}{2}})$; but the highly transcendental result is not particularly illuminating, and we do not give it here. We mention it only to emphasize that such a modification does exist.
[36] I. M. Gel'fand and M. I. Graev, Tr. Mosk. Mat. Obshch. **11**, 243 (1962) (In Russian).
[37] I. M. Gel'fand, M. I. Graev, and N. J. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[38] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).
[39] A. P. Balachandran, W. J. Meggs, P. Ramond, and J. Nuyts, Phys. Rev. **187**, 2080 (1969).
[40] J. M. Charap and B. M. Minton, J. Math. Phys. **10**, 1823 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Analytic Properties of Nonstrictly Localizable Fields*

F. CONSTANTINESCU†
Department of Physics, University of Munich, Germany

(Received 2 March 1970)

A class of test functions "minimal with respect to causality" is introduced. The corresponding fields are called local. Tempered and strictly localizable fields are local, but there is a large class of fields that are local but not strictly localizable. For local fields, the analytic properties of vacuum expectation values are studied. The local fields that are not strictly localizable are characterized by an arbitrary fast increase of Wightman functions near the light cone. With an adequate definition of local commutativity, other properties of tempered and strictly localizable fields seem to remain valid.

1. INTRODUCTION

In the usual Wightman framework, great attention has been paid to the investigations of the theoretical and experimental consequences of the following basic requirements of relativistic quantum field theories[1–3]:

(a) Hilbert space of states,

(b) covariance of the fields under the inhomogeneous Lorentz group,

(c) positive energy,

(d) local commutativity of fields,

(e) particle interpretation.

On the basis of the two first requirements it follows that a field $A(x)$ will be an operator-valued generalized function.[3,4] In order to obtain an operator, A must be averaged with a smooth test function $f(x)$:

$$A(f) = \int A(x)f(x)\,dx. \tag{1}$$

At this point we have to answer the following question: How to take the test functions? It was pointed out[1–3] that some basic properties of relativistic fields can be obtained if we choose tempered test functions. The requirement of temperedness appears natural from a physical viewpoint because it reflects the symmetry between coordinate and momentum spaces. At the same time, with this requirement, one can prove that the scattering amplitude $F(s, t)$ is analytic in s (for fixed $t < 0$) in a cut plane and has a polynomial behavior.

In a series of remarkable papers, Jaffe[3] has shown that the requirements (a)–(e) can be incorporated in a theory of strictly localizable fields with test functions which allow, in momentum space, a nontempered increase of fields. The wider class of fields studied by Jaffe is physically relevant since it allows for the possibility that the off-mass-shell amplitudes can grow, at large energies, faster than any polynomial. Besides temperate fields, strictly localizable fields include a class of nonrenormalizable interactions.[5] Entire functions of free fields are other examples.

We say that a field $A(x)$ is strictly localizable in a certain region of space–time if A can be averaged with some test function $f(x)$ which vanishes outside this region. Such a notion is very convenient for the following statement of local commutativity: A field A obeys

the requirement of local commutativity if $A(f)$ commutes, or anticommutes, with $A(g)$ whenever the test functions $f(x)$ and $g(x)$ vanish outside spacelike separated regions.

From the mathematical point of view, the property of a field to be strictly localizable is closely connected with the existence of sufficiently many test functions with compact support in configuration space. On the other hand, the existence of test functions with compact support in configuration space requires test functions in momentum space which have a decrease at infinity like $\exp(-\|p\|^{\mu})$, $\mu < 1$, where $\|p\|$ is the Euclidean norm.[6,3]

The main results of Jaffe are the following:

(i) The momentum space test functions suitable for a strictly local field theory belong to a class $\mathfrak{M}(\mathbb{R}^4) \subset \mathfrak{D}(\mathbb{R}^4)$.

(ii) The convergence in $\mathfrak{M}(\mathbb{R}^4)$ is defined by the following family of norms:

$$\|f\|_{n,m,N} = \sup_{p\in\mathbb{R}^4} g(N\|p\|^2)(1 + \|p\|)^n |D^m f(p)|. \quad (2)$$

Here n and N are integers and $g(t)$ is an entire function which characterizes the momentum space growth of the off-mass-shell amplitudes. The condition $\mu < 1$, where μ is the order of growth of $g(t)$, is a sufficient condition for strict localizability. [A necessary and sufficient condition for strict localizability is

$$\int_1^{\infty} (1 + t^2)^{-1} \log g(t^2)\, dt < +\infty.\Big]$$

(iii) The vacuum expectation values are generalized functions over the Fourier transform of $\mathfrak{M}(\mathbb{R}^{4n})$, and are boundary values of analytic functions in the forward tube with essential singularities near the real points. The general principles of relativistic field theory can be worked out in this frame. The only difference with tempered fields is a faster than polynomial bound of the off-mass-shell amplitude at high energy. However, the on-mass-shell scattering amplitudes have the property of polynomial boundedness. This last result was recently obtained in all theories with local observables.[7] In spite of the important results obtained for a strictly local field theory, the choice of the test functions remains without direct physical justification. It is merely motivated by the technical point of view to have at hand a familiar formulation of local commutativity.

In this work we consider a category of fields larger than the strictly localizable ones. For the order of growth of the Jaffe indicatrix, we admit the values $\mu \leq 1$. The choice of the test functions suitable for these fields can be physically motivated from the *causality condition*. We call these fields local.[8] We show that for local fields the vacuum expectation values are generalized functions which are boundary values of polynomially bounded analytic functions in the forward tube. For local fields that are not strictly localizable, it is proved that the Wightman functions can grow arbitrarily fast near the light cone. These analytic properties hold independently of the definition of local commutativity we adopt. It appears that, with a suitable definition of local commutativity,[5,8] other results obtained for tempered and strictly localizable fields are also valid. At the same time, for local fields, there exists a limit theorem.[9]

2. CAUSALITY AND TEST FUNCTIONS

Recently the idea that causality may serve as a physical motivation in choosing the test functions has found special attention.[8] In a consistent quantum field theory these test functions have to incorporate the basic requirements (a)–(e). Some arguments for utilizing causality in choosing the test functions come from the theory of nonrenormalizable interactions[5] and from Bogoliubov's causality principle for constructing the scattering matrix in the quantum theory of interacting fields.[8] Here we present a simple argument compatible with the requirements (a)–(e). Let us write the Källén–Lehmann representation for the Fourier transform of the vacuum expectation value of the commutator

$$f(p) = \int e^{ipx}(\Psi_0, [A(\tfrac{1}{2}x), A(-\tfrac{1}{2}x)]\Psi_0)\, dx$$
$$= \int_0^{\infty} dm^2 \rho(m^2)\delta(p^2 - m^2)\epsilon(p_0). \quad (3)$$

Replacing $\delta(p^2 - m^2)\epsilon(p_0)$ by

$$[(p_0 + i\epsilon)^2 - \bar{p}^2 - m^2]^{-1},$$

we obtain the expectation value for the retarded commutator,

$$f^{\mathrm{Ret}}(p) = \lim_{\epsilon\to 0} \int_0^{\infty} dm^2 \rho(m^2)[(p_0 + i\epsilon)^2 - \bar{p}^2 - m^2]^{-1}. \quad (4)$$

But, for nontempered fields, (4) diverges because $\int_0^{\infty} \rho(m^2)\, dm^2$ does so. An acceptable definition of (4) can only be obtained by inserting a convergence factor $g(m^2)$ in (4)[3]:

$$f^{\mathrm{Ret}}(p) = \lim_{\epsilon\to 0} \int_0^{\infty} dm^2 \frac{g(p^2)}{g(m^2)} \rho(m^2) \times [(p_0 + i\epsilon)^2 - \bar{p}^2 - m^2]^{-1}. \quad (5)$$

This definition is acceptable only if inside the forward light cone (in configuration space) the Fourier transform of $f^{\mathrm{Ret}}(p)$, Eq. (5), agrees with the Fourier transform of $f(p)$. Thus, since $\mathcal{F}(f(p)) = \mathcal{F}(f^{\mathrm{Ret}}(p))$ inside the forward cone, it follows that in passing from (3) to (5) the introduction of g can modify f only at the origin of the light cone $x = 0$. Therefore, if $g_1(m^2)$ and $g_2(m^2)$, $g_1 \neq g_2$, have the same growth, the difference

$$f_1^{\mathrm{Ret}}(p) - f_2^{\mathrm{Ret}}(p) = \int_0^\infty dm^2 \rho(m^2) \left(\frac{g_2(m^2)g_1(p^2) - g_1(m^2)g_2(p^2)}{g_1(m^2)g_2(m^2)(p^2 - m^2)} \right) = \omega(p^2) \tag{6}$$

must have a Fourier transform $\omega(-\Box)\delta(x)$ localized at $x = 0$. The function $\omega(p^2)$ has the same order as $g_1(p^2)$ and $g_2(p^2)$. On the other hand, the generalized function $\omega(-\Box)\delta(x)$ can be localized at $x = 0$ if and only if $\omega(p^2)$ has a maximal growth of order one and type zero.[10,8] Consequently, from causality [i.e., from the condition that the Fourier transforms of $f^{\mathrm{Ret}}(p)$ vanish outside the forward cone], it follows that we can choose only convergence factors [i.e., indicatrices $g(p^2)$] with a maximal growth of order one and type zero. We see that the test functions for strictly localizable fields obey the causality principle and that there exist also test function spaces which obey the causality principle but do not incorporate strictly localizable fields (the case of order one and type zero for the indicatrix). A theory is called local[8] if the indicatrix has maximal growth of order one and type zero. Although the change in the growth of the indicatrix from strictly local to local theories is minor, the class of local theories is much larger than the class of strictly local ones.[11] At the same time, when leaving strictly local theories, we see that new problems of principle are springing up: The test functions become analytic, and therefore no test function with compact support exists. We can no longer localize fields in a bounded region of space–time, and the local commutativity has to be formulated using the notion of support of an analytic functional.[5,8] Similar problems appear when we try to formulate nonlocal theories.[8] Therefore, it is of interest to work out the case with order of growth one and type zero which appears like a limiting case between strictly local and nonlocal theories.

Although Jaffe's spaces are quite general and unrestrictive, we do not use them because, for an indicatrix with order of growth one and type zero, the test functions have no definite domain of analyticity. Instead, we use some similar spaces of Gel'fand and Shilov[6] which are called "spaces of type S." These spaces were used also to study entire functions of free fields.[12]

In the following section we give first a brief review of such spaces and then derive some new results which are then used in studying analytic properties of vacuum expectation values for our local fields.

3. SPACES OF TYPE S

Any infinitely differentiable function $\varphi(p)$ (test function in momentum space) belongs to S_α, $\alpha > 0$, if and only if there exist constants C_q and a (depending on φ) such that

$$|\varphi^{(q)}(p)| \leq C_q \exp\left(-a_1 |p_1|^{\alpha^{-1}} - \cdots - a_n |p_n|^{\alpha^{-1}}\right), \tag{7}$$

where n is the number of variables ($p = p_1 \cdots p_n$). The spaces S^α (test functions in configuration space) are defined as Fourier transforms of S_α: $S^\alpha = \mathcal{F}(S_\alpha)$. We have to consider also the spaces $S_{\alpha,a}$ which are made up of all functions $\varphi(p)$ having the property

$$|\varphi^{(q)}(p)| \leq C'_{q\delta} \exp\left[-(a_1 - \delta_1)|p_1|^{\alpha^{-1}} - \cdots - (a_n - \delta_n)|p_n|^{\alpha^{-1}}\right] \tag{8}$$

for any $\delta_i > 0$, $i = 1, 2, \cdots, n$, and also the Fourier transformed spaces $S^{\alpha,a} = \mathcal{F}(S_{\alpha,a})$. We have

$$S_\alpha = \bigcup_a S_{\alpha,a}, \quad S^\alpha = \bigcup_a S^{\alpha,a}.$$

It was shown in Ref. 6 that for $\alpha < 1$ the functions $\psi(x) \in S^{\alpha,a}$ or S^α can be extended (with x replaced by $z = x + iy$) into entire functions $\psi(z)$. For $\alpha = 1$ the functions $\psi(x) \in S^{\alpha,a}$, $a \neq 0$, can be continued as analytic functions $\psi(z)$, into the same complex neighborhood of $\mathbb{R}^n$. For $a \to 0$, the dimension of this neighborhood tends to zero. The functions $\psi(x) \in S^1$ still have an analytic continuation in $\mathbb{C}^n$. For instance (in the case $n = 1$), for each $\psi(x) \in S^1$ there exists a strip around the real axis in which $\psi(z)$ is analytic, but no strip exists in which all functions of S^1 are analytic.

The topology in $S_{\alpha,a}$ can be defined with the help of the norms

$$\|\varphi\|_r = \sup_{q \leq r} M_r(p) |\varphi^{(q)}(p)|,$$

where

$$M_r(p) = \exp\left[(1 - r^{-1})(a_1 |p_1|^{\alpha^{-1}} + \cdots + a_n |p_n|^{\alpha^{-1}})\right].$$

The spaces $S_{\alpha,a}$ are therefore countably normed spaces. The topology in $S_{\alpha,a}$ can be defined through a convergence notion in the following way: A set of elements $\varphi_\nu \in S_{\alpha,a}$ converges to zero if and only if the derivatives $\varphi_\nu^{(q)}(p)$ converge uniformly to zero for each

q and the norms $\|\varphi_\nu\|_r$ are bounded for each r. It is easy to see that $\mathcal{D} \subset S_{\alpha,a} \subset S$, where $\mathcal{D}$ and S are Schwartz' test function spaces. Using the standard methods of distribution theory, we can show that $\mathcal{D}$ is dense in $S_{\alpha,a}$. Because $\mathcal{D}$ is also dense in S, $S_{\alpha,a}$ will be dense in S.

The topology of $S_\alpha = \bigcup_a S_{\alpha,a}$ is the topology of inductive limits. A sequence $\{\varphi_\nu\}$ of S_α converges to zero if and only if all functions $\varphi_\nu(x)$ are in the same space $S_{\alpha,a}$ in which $\{\varphi_\nu(x)\}$ tends to zero. We have that $\mathcal{D} \subset S_\alpha \subset S$ and S_α is dense in S. It follows that we can write

$$S' \subset S'_{\alpha,a} \subset \mathcal{D}', \quad S' \subset S'_\alpha \subset \mathcal{D}', \tag{9}$$

i.e., $S'_{\alpha,a}$, S'_α are spaces of Schwartz distributions which contain the tempered ones. The spaces $(S^{\alpha,a})'$ and $(S^\alpha)'$ are the Fourier transforms of $S'_{\alpha,a}$ and S'_α. Finally, the spaces $S_{\alpha,a}$, S_α, $S^{\alpha,a}$, and S^α are invariant under translations and are nuclear.[6]

From (8) and (7) it follows that $S_{\alpha,a}$ and S_α are similar to Jaffe's spaces in momentum space. In order to have the same asymptotics, we take $\alpha = 1/\mu$, where μ is the order of growth of Jaffe's indicatrix. The theory of strictly localizable fields can be formulated for $\mu < 1$, i.e., for $\alpha > 1$. We have seen that S_1 is the maximal space of test functions in momentum space admitted by the causality condition. We need the following:

Lemma 1: Let $T \in S'_1 \subset \mathcal{D}'$ and $\varphi(p) \in S_1$. The distribution $T\varphi$ is tempered.

Lemma 2: Let $T \in \mathcal{D}'(\mathbb{R}^4)$. Suppose that, for every $\eta \in V^+$ (the forward light cone), we have $e^{-p\eta}T \in S'_1$. Then it follows that T is in S'_1.

We give only the proof of Lemma 2. For Lemma 1 the proof is similar. Let $\psi(p) \in S_1$. It follows that there exists an $a > 0$, $a \in V^+$, such that $\psi(p) \in S_{1,a}$. We take $\eta = \frac{1}{2}a$ and define

$$(T_1, \varphi) = (e^{-p\eta}T, \varphi), \tag{10}$$

where $\varphi(p) = e^{p\eta}\psi(p) \in S_1$. From $\psi_n(p) \to \psi(p)$ in S_1, it follows that

$$\lim_{n\to\infty} (T_1, \psi_n) = 0$$

and therefore T_1 will be a distribution in S'_1, $T_1 \in S'_1 \subset \mathcal{D}'$. For $\psi(p)$ having compact support, we have from (10) that $(T_1, \psi) = (T, \psi)$. Because $\mathcal{D}$ is dense in S_1, we conclude $T = T_1 \in S'_1$.

We remark that spaces which are in closed connection with the spaces of type S for $\alpha \geq 1$ have been considered also by Sebastiao e Silva (ultradistributions)[13] and by Sato (hyperfunctions).[14]

4. VACUUM EXPECTATION VALUES AS BOUNDARY VALUES OF ANALYTIC FUNCTIONS

In what follows we consider functions which depend on n 4-vectors. A variable such as p will stand for a collection on n 4-vectors $p^{(j)}$ with components $p_\nu^{(j)}$. The product pq stands for $\sum_{j=1}^n (p^{(j)}, q^{(j)})$, where $(p^{(j)}, q^{(j)})$ is the Lorentz inner product of the jth 4-vectors. Furthermore, Γ will denote the product of n forward light cones $\Gamma = \otimes_{j=1}^n V^+$. The points in V^+ are 4-vectors such that $p_0^2 - \bar{p}^2 \geq 0$, $p_0 > 0$.

Let $\eta_j, j = 1, 2, \cdots, M$, be M points in the interior of Γ, and let $\eta \in \Gamma$ be a distinct point in the interior of the convex hull of η_j. In other words,

$$\eta = \sum_{j=1}^M \lambda_j \eta_j, \quad \lambda_j > 0, \quad \sum_{j=1}^M \lambda_j = 1. \tag{11}$$

By taking M sufficiently large, we can choose the η_j so that they span $\mathbb{R}^{4n}$. Hence a full neighborhood of η lies in the convex hull of η_j. Let us consider the auxiliar function

$$a(p, \eta, \eta_j) = \exp(-p\eta)\left(\sum_{j=1}^M \exp(-p\eta_j)^{-1}\right). \tag{12}$$

We have the following[3]:

Lemma 3: Under the above conditions there holds

$$|D_p^m a(p, \eta, \eta_j)| \leq C_m \exp(-d\,\|p\|), \tag{13}$$

where

$$\|p\|^2 = \sum_{j=1}^n [(p_{j0})^2 + (\bar{p}_j)^2],$$

$d > 0$ being a geometrical constant depending on η and η_j, and the C_m, $C_0 = 1$, are other constants which do not depend on η and η_j.

In particular, from Lemma 3 it follows that the $a(p, \eta, \eta_j)$ are elements of $S_1(\mathbb{R}^{4n})$. The following theorem plays a central role.

Theorem 1: Suppose that $T(p) \in \mathcal{D}'_p(\mathbb{R}^{4n})$ and

$$\exp(-p\eta)T(p) \in S'_1(\mathbb{R}^{4n}). \tag{14}$$

Then the Laplace transform

$$\mathfrak{L}(T)(\xi - i\eta) = \mathcal{F}_\xi(\exp(-p\eta)T(p)) \tag{15}$$

is a function of $\xi - i\eta$ holomorphic in the tube $\mathcal{T}_n = \mathbb{R}^{4n} - i\Gamma$. For each compact $K \subset \Gamma$, there exists a polynomial $P_K(\xi)$ such that for all $\eta \in K$

$$|\mathfrak{L}(T)(\xi - i\eta)| \leq P_K(\xi). \tag{16}$$

Conversely, every function holomorphic in $\mathcal{T}_n$ which satisfies (16) is the Laplace transform of a $T(p)$ satisfying (14).

Proof: Following an idea of Jaffe,[3] we will show first that condition (14) is equivalent to the following one:

$$\exp(-p\eta)T(p) \in S'(\mathbb{R}^{4n}). \tag{17}$$

Indeed, from (14) we have immediately the relation (17) because $S' \subset S'_1$ [see Eq. (9)]. It remains to prove that from (17) the relation (14) follows. From (12) we have

$$\exp(-p\eta)T(p) = a(p,\eta,\eta_j)\left(\sum_{j=1}^{M}\exp(-p\eta_j)T(p)\right). \tag{18}$$

The sum in the rhs of (18) is in S'_1. Taking into account Lemma 3, we see that $a(p, \eta, \eta_j) \in S_1$. In (18) we have, therefore, to multiply functions from S_1 with generalized functions from S'_1. From Lemma 1 it results that $\exp(-p\eta)T(p)$ is in S' for all $\eta \in \Gamma$. We can apply now standard results[4] to see that the Laplace transform is holomorphic in $\mathcal{T}_n$ and obeys the polynomial bound (16).

We remark that this theorem is also valid for $\alpha > 1$.

Theorem 2: Suppose that a function holomorphic in $\mathcal{T}_n$ satisfies the polynomial bound (16). Then, for $\eta \to 0$, $\eta \in \Gamma$, it converges in the sense of $(S^1)'$ to a generalized function in $(S^1)'$.

Proof: Indeed, from Theorem 1, the holomorphic function is the Laplace transform of a distribution in $\mathcal{D}'(\mathbb{R}^{4n})$ which satisfies (14). Taking into account Lemma 2, we have generally from

$$\exp(-p\eta)T(p) \in S'_1(\mathbb{R}^{4n}), \quad \eta \in \Gamma,$$

that also $T(p) \in S'_1(\mathbb{R}^{4n})$. Consequently, the Laplace transform $\mathfrak{L}(T)(\xi - i\eta)$ converges for $\eta \to 0$ to the Fourier transform $\mathcal{F}_\xi(T(p))$, which is in $(S^1)'$. The convergence is in the sense of generalized functions in $(S^1)'$.

It is interesting to remark that the bound

$$|\mathfrak{L}(T)(\xi - i\eta)| \leq P_K(\xi), \quad \eta \in K,$$

in the case $(S^\alpha)'$, $\alpha = 1$, replaces the bound

$$|\mathfrak{L}(T)(\xi - it\eta)| \leq P_K(\xi)/t^r, \quad r \text{ integer}, \quad 0 < t < 1, \quad \eta \in K,$$

for tempered distributions. For test functions suitable for strictly local theories, essential singularities appear near the real points.[3,9] We conclude that, for fields that are localizable but not strictly localizable, the Laplace transforms (and therefore the Wightman functions) can grow arbitrarily fast near the real points. The same results and concrete examples can be obtained using the Güttinger's series of derivatives of δ functions.[10]

Let us take now a scalar field with momentum space test functions in S_1. With the standard notations (1) and (2) and using the nuclear property and translation invariance of the spaces of type S, we have the following:

Theorem 3: The vacuum expectation values in a local theory are boundary values of analytic functions in $\mathcal{T}_{n-1}$:

$$W(\xi_1, \cdots, \xi_{n-1}) = \lim_{\eta_1 \to 0, \cdots, \eta_{n-1} \to 0} W(\xi_1 - i\eta_1, \cdots, \xi_{n-1} - i\eta_{n-1}). \tag{19}$$

The convergence in (19) is the one in the sense of generalized functions in $(S^1)'$. Furthermore, $W(\xi_1 - i\eta_1, \cdots, \xi_{n-1} - i\eta_{n-1})$ is polynomially bounded in $\mathcal{T}_{n-1}$, but can grow arbitrarily fast near the real points $(\eta_1, \cdots, \eta_{n-1} \to 0)$. The function W has an unique analytic continuation in the extended tube.

The statement of this theorem follows from classical results[1,2] and Theorems 1 and 2.

We conclude that the only difference between tempered and strictly localizable fields, on the one side, and the local fields that are not strictly localizable, on the other side, lies in the fact that the latter can admit arbitrarily fast increasing Wightman functions near the light cone contrary to the former.

In what follows we will discuss other properties of local but nonstrictly localizable fields. Most of them can be proved as usually taking into account that for these fields the Wightman functions are still boundary values of analytic functions. The first consequence is that the Jost points still exist and therefore the weak local commutativity can be formulated. On the other hand, a global edge of the wedge theorem is valid. The usual proof for boundary value in $\mathcal{D}'$ can be followed if we notice three elementary facts: (i) $(S^1)' * S^1 \subset C^\infty$—the space of infinitely differentiable functions; (ii) the possibility of approximating the Dirac function $\delta(x)$ with test functions in S^1 (for example, with Gaussian functions which are both in S_1 and S^1); and (iii) the nuclearity of S^1.[6] It results that the *PCT* theorem remains valid.[1] The local but nonstrictly localizable fields are *PCT*

invariant as a consequence of weak local commutativity.

We remark that for our fields a limit theorem in the sense of Jaffe is valid.[9] The limit theorem enables us to consider entire functions of a free field with an infinite order of growth as local but nonstrictly localizable fields.

Let us give now some details about the scattering theory. We have to look at a scattering theory for fields considered as operator-valued generalized functions over S^1 and which are satisfying the Wightman axioms except the axiom of locality. It was shown recently by Steinmann[15] that a usual scattering theory exist for fields which are operator-valued generalized functions on the test functions of Z, i.e., on test functions whose Fourier transforms have compact support if the Wightman axioms (except local commutativity) and the axiom of asymptotic completeness are satisfied and if instead of local commutativity a certain regularity condition (condition R) in p space is considered. More precisely, one has to demand the existence of generalized retarded products as defined by Steinmann,[16] i.e., operator-valued invariant generalized functions $G_\mu(x_1, \cdots, x_n)$ satisfying the usual algebraic relations and symmetry properties. The Fourier transforms of the matrix elements of $G_\mu(x_1, \cdots, x_n)$ are supposed to be analytic functions with pole-type singularities near the real points in momentum space. There is no condition about the asymptotic behavior for large momenta real or imaginary. We have, for our case, $\mathfrak{D} \subset S_1$ and $Z \subset S^1$, and $\mathfrak{D}$ and Z are dense in S_1 and S^1, respectively. For duals $(S^1)' \subset Z'$ and therefore the usual scattering theory will be valid for our fields if the support properties on the generalized retarded products are replaced with condition R.

5. CONCLUSIONS

We have introduced the test functions of a field theory by utilizing physically motivated causality condition, i.e., the minimum class of test functions compatible with causality. The causality condition implies for Jaffe's indicatrix a maximal growth of order one and type zero. The corresponding fields are called local. Into the category of local fields there fall tempered and strictly localizable fields, but there is also a class of fields that are local but not strictly localizable. For local fields the common analytic properties of the vacuum expectation values are valid. The only difference between local and strictly theories is that in the former, near the light cone, the Wightman functions can grow arbitrarily fast. The convergence of the Wightman functions to boundary values is to be defined in the sense of generalized functions in $(S^1)'$. This convergence is weaker than the corresponding convergence for tempered and strictly localizable fields. In the topology of (S^1) a limit theorem holds.[9] It seems that if local commutativity is replaced with some "technical" conditions, other results of strictly localizable fields remain valid.

ACKNOWLEDGMENTS

The author is indebted to Professor A. Jaffe for sending him his unpublished manuscript, to Professor W. Güttinger for stimulating discussions, and to Professor F. Bopp for the warm hospitality of his Institute. He wishes to express his sincere gratitude to the A. v. Humboldt Foundation for a Fellowship.

* Work supported by the A. v. Humboldt Foundation.

† On leave of absence from the University of Cluj, Romania.

[1] R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964).

[2] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[3] A. M. Jaffe, Phys. Rev. Letters **17**, 661 (1966); Phys. Rev. **158**, 1454 (1967); also "High Energy Behavior of Strictly Localizable Fields. Mathematical Tools," unpublished.

[4] A. S. Wightman, Ann. Inst. Henri Poincaré **1**, 403 (1964).

[5] W. Güttinger, Nuovo Cimento **10**, 1 (1958); Fortschr. Physik **14**, 483 (1966).

[6] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vols. 1, 2.

[7] H. Epstein, V. Glaser, and A. Martin, CERN Preprint, TH 991, 1969.

[8] G. V. Efimov, Kiev, Preprint ITF-68-52, 1968; N. V. Hieu, Ann. Phys. (N.Y.) **33**, 428 (1965).

[9] F. Constantinescu, University of Munich, Preprint, 1969.

[10] F. Constantinescu, Nuovo Cimento Lett. **1**, 892 (1969).

[11] S. S. Khoruzhy, Phys. Letters **26B**, 17 (1967).

[12] A. Rieckers, University of Munich, Preprint, 1968.

[13] J. Sebastiao e Silva, Math. Ann. **174**, 109 (1967).

[14] A. Martineau, Sém. Bourbaki, 13e année, No. 214 (1960/61).

[15] O. Steinmann, Preprint SIN (Zürich), 1970.

[16] O. Steinmann, Commun. Math. Phys. **10**, 245 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Motion in Electric and Magnetic Fields. I. Klein-Gordon Particles

LUI LAM
Department of Physics, Columbia University, New York, New York 10027
and
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

(Received 5 August 1970)

Exact solutions of the Klein–Gordon equation are found for a new solvable configuration of external electromagnetic fields, namely, a static uniform electric field inclined at an arbitrary angle to a static uniform magnetic field of arbitrary strength. Applications are discussed at the end.

I. INTRODUCTION

The relativistic motion of a classical particle in external electric and magnetic fields is quite well understood.[1] However, the corresponding case for quantum particles is yet to be exploited to its full extent. The recent availability of intense electron beams in the 100 keV to MeV range in the laboratory[2] and the proliferation in recent years of meson factories that are capable of generating controllable intense meson beams[3,4] will make practical any test or application derived from our solutions. On the other hand, outside the laboratories, the configuration of parallel electric and magnetic fields[5] (and certainly others too) has appeared in astrophysical problems where relativistic particles are likely to be present.[6] All these add up to an urgent demand for the clear understanding of a relativistic quantum particle in external electromagnetic fields. This is readily accomplished if we have the exact solutions of the equations of motion. Nevertheless, exact solutions in nature are known for their scarcity as well as their elegance and utility. In the case of spin-$\frac{1}{2}$ particles, since Dirac wrote down the relativistic wave equation in 1928, we have witnessed surprisingly few solvable configurations in the literature. Those we know of are the following: a Coulomb potential,[7] a constant magnetic field,[8] a constant electric field,[9] the field of a plane wave,[10] the field of a plane wave with a constant magnetic field in its direction of propagation,[11] and, lastly, four cases in which the electromagnetic potential has a particular functional dependence on the space coordinates.[12] The potential usefulness of an exact solution need not be elaborated here. One only has to remember the case of an electron in a constant magnetic field, which ranges from the Landau levels in solids to Malkin and Man'ko's model of dynamical symmetry.[13]

In this paper, we present in detail for the first time the exact solutions for a spin-0 particle in a new solvable configuration of external fields, namely, a static uniform electric field inclined at an arbitrary angle to a static uniform magnetic field of arbitrary strength. This will also serve to clarify the rather involved mathematics and prepare the way for the more complicated case of spin-$\frac{1}{2}$ particles to be discussed in a separate paper[14] in which the effect of the spin will be emphasized.

As is well known, the configuration of an uniform electric field $\mathbf{E}$ inclined to an uniform magnetic field $\mathbf{B}$ can always be Lorentz-transformed to that of parallel fields if they are not originally orthogonal to each other. In fact, if we take $\mathbf{E}$ and $\mathbf{B}$ to lie in the (y, z) plane, the appropriate Lorentz frame in which $\mathbf{E}'$ is parallel to $\mathbf{B}'$ moves with velocity[15] $\boldsymbol{\beta}$ given by[16]

$$(1 + \beta^2)(\mathbf{B} \times \mathbf{E}) + \boldsymbol{\beta}(B^2 + E^2) = 0. \qquad (1)$$

Equation (1) always has a solution with $\beta < 1$, and we note that $\boldsymbol{\beta}$ lies in the x direction and is thus perpendicular to both $\mathbf{E}$ and $\mathbf{B}$. Therefore, if we know the exact solutions for the two cases of $\mathbf{E}$ parallel to $\mathbf{B}$ and $\mathbf{E}$ perpendicular to $\mathbf{B}$, respectively, and if we know how to Lorentz-transform the wavefunctions, we will already have in hand the exact solutions for the general case of $\mathbf{E}$ inclined to $\mathbf{B}$ at an arbitrary angle. The existence of two intrinsic cases is related to the fact that both $\mathbf{E} \cdot \mathbf{B}$ and $E^2 - B^2$ are Lorentz invariants. There is only one drawback to this approach, namely, in general, a stationary state need not be Lorentz transformed into a stationary state.[17] However, for $\mathbf{E}$ and $\mathbf{B}$ both in the z direction the stationary states turn out to have the form (see Sec. III)

$$\varphi(y, z) \exp\,[i(p_x x - \epsilon t)], \qquad (2)$$

which under a Lorentz transformation in the x direction, such as the one described in (1), remains a stationary state in the new frame. In the following, we therefore confine our attention to the parallel and orthogonal configurations only.

For **E** orthogonal to **B** and $E \neq B$, the configuration of fields can further be Lorentz reduced to that of a pure electric field or a pure magnetic field[17] (depending on the relative strength of the fields), and we may obtain the needed solutions by a proper Lorentz transformation on the known exact solutions.[8,9] Nevertheless, for a uniform treatment, they are derived below in conjunction with the Lorentz-invariant case of $\mathbf{E} \cdot \mathbf{B} = 0$, $E = B$.

In the following, exact solutions for a Klein–Gordon particle in parallel and orthogonal fields are presented. The nonrelativistic limit is discussed in Sec. II and compared to the classical results, while the relativistic motion is reserved for Sec. III. Finally, in Sec. IV, the various solutions obtained are analyzed and discussed.

II. NONRELATIVISTIC MOTION

Nonrelativistic motion of a spin-0 particle is described by the Schrödinger equation, which in the presence of external electromagnetic potentials $(\mathbf{A}, A_0)$ is

$$[(\mathbf{p} - e\mathbf{A})^2/2m]\varphi = (p_0 - eA_0)\varphi, \tag{3}$$

where e is the charge of the particle.

A. E ∥ B

For constant, uniform **E** and **B** both in the z direction, we may take

$$A_0 = -zE, \quad A_1 = -yB, \quad A_2 = A_3 = 0, \tag{4}$$

where E and B are constants and are assumed to be nonvanishing for definiteness. Equation (3) then becomes

$$[Q(y) + D(z)]H(y)K(z) = 0, \tag{5}$$

where

$$Q(y) \equiv p_z^2 + (eB)^2(y + p_x/eB)^2, \tag{6}$$

$$D(z) \equiv p_z^2 - 2meE(z + \epsilon/eE), \tag{7}$$

and

$$\varphi = \exp[i(p_x x - \epsilon t)]H(y)K(z), \tag{8}$$

since both p_x and the energy ϵ are constants of motion. Equation (5) is separable and is equivalent to

$$[Q(y) + b]H(y) = 0, \tag{9}$$

$$[D(z) - b]K(z) = 0, \tag{10}$$

where b is a constant. With the change of variable through

$$\alpha \equiv (|eB|)^{\frac{1}{2}}(y + p_x/eB), \tag{11}$$

Eq. (9) reduces to

$$\left(-\frac{d^2}{d\alpha^2} + \alpha^2 + \frac{b}{|eB|}\right)H = 0 \tag{12}$$

and gives the solution

$$H = \exp(-\tfrac{1}{2}\alpha^2)H_n(\alpha), \tag{13}$$

$$b = -|eB|(2n + 1), \quad n = 0, 1, \cdots, \tag{14}$$

where the H_n are the Hermite polynomials. Introducing the new variable

$$\delta \equiv -(2meE)^{\frac{1}{2}}\left(z + \frac{\epsilon}{eE} + \frac{b}{2meE}\right), \tag{15}$$

we see that Eq. (10) becomes

$$\left(-\frac{d^2}{d\delta^2} + \delta\right)K = 0$$

and gives

$$K = \Phi(\delta), \tag{16}$$

where Φ is the Airy function.[18] Collecting (8), (13), and (16), we have

$$\varphi_{p_x,\epsilon,n} = \exp[i(p_x x - \epsilon t)] \exp(-\tfrac{1}{2}\alpha^2)H_n(\alpha)\Phi(\delta) \tag{17}$$

with α and δ given by (11), (15), and (14). Therefore, motion in the (x, y) plane is in quantized orbits while the motion in the z direction is governed by Φ. The quantization of b has no analog in the classical case.

B. E ⊥ B

Taking **B** in the z direction and **E** in the y direction, we may set

$$A_0 = -yE, \quad A_1 = -yB, \quad A_2 = A_3 = 0. \tag{18}$$

Equation (3) becomes

$$\left(\frac{d^2}{d\rho^2} - \rho^2 + k\right)M = 0, \tag{19}$$

where

$$\rho \equiv (|eB|)^{\frac{1}{2}}[y - (mE - Bp_x)/eB^2], \tag{20}$$

$$k \equiv (|eB|)^{-1}[-p_x^2 - p_z^2 + 2m\epsilon + (Bp_x - mE)^2/B^2], \tag{21}$$

and

$$\varphi = \exp[i(p_x x + p_z z - \epsilon t)]M(y), \tag{22}$$

with p_x, p_z, and ϵ constants of motion. Equation (19) is the same type as (12) and thus

$$M = \exp(-\tfrac{1}{2}\rho^2)H_n(\rho) \tag{23}$$

with

$$k = 2n + 1. \tag{24}$$

Equations (21) and (24) give the quantized energy levels

$$\begin{aligned}\epsilon &= \epsilon_{p_x,p_z,n} \\ &= (2m)^{-1}[|eB|(2n + 1) + p_x^2 + p_z^2 \\ &\quad - (mE - Bp_x)^2/B^2].\end{aligned} \tag{25}$$

Therefore, $\varphi = \varphi_{p_x,p_z,n}$, given by (22), (23), and (20).

From the form of the (kinetic) energy in (25), it seems that we may make ϵ as low as we wish by increasing the electric field strength E alone. This is not the case, because, when the E term dominates, $|\epsilon|$ becomes $(m/2)(E/B)^2$ and is comparable to the rest mass m, and at this point the applicability of the Schrödinger equation breaks down. This restriction of $E \ll B$ applies also for the classical motions.[1]

III. RELATIVISTIC MOTION

The Klein–Gordon equation in the presence of external fields is

$$[(\mathbf{p} - e\mathbf{A})^2 + m^2]\varphi = (p_0 - eA_0)^2\varphi. \tag{26}$$

A. E ∥ B[19]

In the gauge of (4), Eq. (26) reduces to

$$[Q(y) + R(z) + m^2]F(y)G(z) = 0, \tag{27}$$

where $Q(y)$ is defined by (6),

$$R(z) \equiv p_z^2 - (\epsilon + eEz)^2 \tag{28}$$

and

$$\varphi = \exp[i(p_x x - \epsilon t)]F(y)G(z). \tag{29}$$

Equation (27) is equivalent to

$$[Q(y) + s]F(y) = 0, \tag{30}$$

$$[R(z) + m^2 - s]G(z) = 0, \tag{31}$$

where s is a constant. (30) is the same as (9); hence

$$F(y) = \exp(-\tfrac{1}{2}\alpha^2)H_n(\alpha), \tag{32}$$

with α defined by (11), and

$$s = -|eB|\,(2n + 1). \tag{33}$$

After the change of variable

$$\gamma \equiv (|eE|)^{\frac{1}{2}}(z + \epsilon/eE), \tag{34}$$

Eq. (31) becomes

$$\left(\frac{d^2}{d\gamma^2} + \gamma^2 - \frac{m^2 - s}{|eE|}\right)G = 0. \tag{35}$$

Expressed in the new variable

$$\theta \equiv -i\gamma^2, \tag{36}$$

with

$$G(z) \equiv \exp(-\theta/2)N(\theta), \tag{37}$$

(36) becomes

$$\left(\theta\frac{d^2}{d\theta^2} + (\tfrac{1}{2} - \theta)\frac{d}{d\theta} - \frac{1 + i(m^2 - s)/|eE|}{4}\right)N = 0. \tag{38}$$

This is the Kummer equation,[20] and there are two linear independent solutions given by

$$N = g_j\left(i\frac{m^2 - s}{4|eE|}, \theta\right), \quad j = 1, 2, \tag{39}$$

where

$$g_1(b, \tau) \equiv F(\tfrac{1}{4}(1 + b), \tfrac{1}{2}, \tau), \tag{40}$$

$$g_2(b, \tau) \equiv \tau^{\frac{1}{2}}F(\tfrac{1}{4}(1 + b) + \tfrac{1}{2}, \tfrac{3}{2}, \tau), \tag{41}$$

and F is the confluent hypergeometric function.[20] Collecting (29), (32), (37), and (39), we have

$$\varphi_{p_x,\epsilon,n} = \exp[i(p_x x - \epsilon t)]\exp(-\tfrac{1}{2}\alpha^2)H_n(\alpha) \times \exp(\tfrac{1}{2}i\gamma^2)g_j\left(i\frac{m^2 - s}{|eE|}, -i\gamma^2\right), \tag{42}$$

with α, γ, and s given by (11), (34), and (33), respectively. As in the nonrelativistic case, the energy levels are not quantized. Comparing (17) and (42), we see that the relativistic effect manifests itself only in the motion in the z direction where the confluent hypergeometric function goes into the Airy function in the nonrelativistic limit. This is expected on physical grounds. However, in making the comparison, one is reminded that ϵ in (42) is the total energy of the particle while the ϵ in (17) is the kinetic energy only.

B. E ⊥ B[21]

In the gauge of (18), writing the wavefunction in the form of

$$\varphi = \exp[i(p_x x - \epsilon t + p_z z)]P(y), \tag{43}$$

we see that (26) reduces to

$$T(y)P(y) = 0, \tag{44}$$

where

$$T(y) \equiv p_y^2 - e^2(E^2 - B^2)y^2 - 2ey(\epsilon E - Bp_x) + p_x^2 + p_z^2 + m^2 - \epsilon^2. \tag{45}$$

Case 1 ($E = B$): In this case $T(y)$ is independent of y if, in addition, $\epsilon = p_x$. We then have free motion for the particle. However, this is unphysical since ϵ, the total energy, is always greater than the momentum unless $m = 0$. Therefore, (45) is linear in y and (44) then has the same form as (10). The solution in this case is easily found to be

$$P(y) = \Phi[u^{\frac{1}{3}}(y + v/u)], \tag{46}$$

where u and v are constants defined by

$$u \equiv 2eB(p_x - \epsilon), \tag{47}$$

$$v \equiv p_x^2 + p_z^2 + m^2 - \epsilon^2. \tag{48}$$

Case 2 ($E \neq B$): With

$$\eta \equiv +(E^2 - B^2)^{\frac{1}{2}}, \tag{49}$$

$$\xi \equiv (|e|\,\eta)^{\frac{1}{2}}[y - (Bp_x - \epsilon E)/(e\eta^2)], \tag{50}$$

(44) becomes

$$\left(\frac{d^2}{d\xi^2} + \xi^2 - a\right)P = 0, \tag{51}$$

where

$$a \equiv [p_x^2 + p_z^2 + m^2 - \epsilon^2 + (Bp_x - \epsilon E)^2/\eta^2]/(|e|\,\eta). \tag{52}$$

Equation (51) is the same as (35) in form, and we have two independent solutions given by

$$P = \exp(\tfrac{1}{2}i\xi^2)g_j(ia, -i\xi^2), \quad j = 1, 2, \tag{53}$$

where g_j, $j = 1, 2$, are defined by (40) and (41). This completes the derivation. The interrelation of the various solutions will be discussed in the next section.

IV. DISCUSSIONS

Case 2 of Sec. IIIB can be further subdivided into the case of $E > B$ and the case of $E < B$. It is the latter case that connects the relativistic solutions of (43) and (53) to the nonrelativistic solutions of (22) and (23). In fact, when $E < B$, η of (49) becomes purely imaginary. Let us define

$$\eta' \equiv -i\eta \tag{54}$$

$$= +(B^2 - E^2)^{\frac{1}{2}} \tag{55}$$

and

$$\xi' \equiv (|e|\,\eta')^{\frac{1}{2}}[y - (\epsilon E - Bp_x)/(e\eta'^2)]. \tag{56}$$

The variable $-i\xi^2$ of g_j in (53) when written in terms of these new variables becomes ξ'^2, which is real under the condition $E < B$. In general, the confluent hypergeometric function F has the asymptotic expansion[22]

$$F(a, b, \tau) \approx e^{-i\pi a}\frac{\Gamma(b)}{\Gamma(b-a)}\tau^{-a} + \frac{\Gamma(b)}{\Gamma(a)}\tau^{a-b}e^{\tau}, \quad \tau \gg 1. \tag{57}$$

For τ real, in order to get rid of the exponentially increasing term, we must have

$$1/\Gamma(a) = 0. \tag{58}$$

Since Γ functions have poles at nonpositive integers only, (58) implies that

$$a = -n, \quad n = 0, 1, 2, \cdots. \tag{59}$$

In our case, combining (40) and (53), we have

$$1 + ia = -4n, \quad n = 0, 1, 2, \cdots, \tag{60}$$

while (41) and (53) together give

$$(1 + ia) + 2 = -4m, \quad m = 0, 1, 2, \cdots. \tag{61}$$

On the other hand, for τ real and n nonnegative integers, we have[22]

$$H_{2n}(\tau) = [(-1)^n(2n)!/n!]F(-n, \tfrac{1}{2}, \tau^2), \tag{62}$$

$$H_{2n+1}(\tau) = [(-1)^n(2n+1)!/n!]F(-n, \tfrac{3}{2}, \tau^2)2\tau. \tag{63}$$

Therefore, when $E < B$, (53) becomes

$$P \approx \exp(-\xi'^2/2)H_{2n}(\xi') \tag{64}$$

or

$$\exp(-\xi'^2/2)H_{2m+1}(\xi') \tag{65}$$

by means of (40), (41), (53), and (60)–(63). Equations (60), (61), (64), and (65) can be condensed into

$$P \approx \exp(-\xi'^2/2)H_n(\xi'), \quad n = 0, 1, \cdots, \tag{66}$$

and

$$1 + ia = -2n. \tag{67}$$

(67) when combined with (52) and (54) gives

$$(\epsilon - Ep_x/B)^2 = (\eta'/B)^2[(2n+1)\,|e|\,\eta' + p_z^2 + m^2]. \tag{68}$$

Therefore, the energy levels are quantized. It is easy to see that (68) reduces to the correct form when E tends to zero, i.e., the case of a pure magnetic field. Furthermore, we observe that the results of (66) and (68) can indeed be obtained directly by solving the original differential equation (44).

The transition to the nonrelativistic limit is effected by letting $\epsilon \to m$ and $\eta' \to B$ in (56), which is suggested by the discussion at the end of Sec. II. We then have ξ' tending to ρ of (20), and (66) tends to (23). Under the same limits, (68) becomes

$$\epsilon = m + (2m)^{-1}[(2n+1)\,|e|\,B + p_z^2] + Ep_x/B, \tag{69}$$

the same as the nonrelativistic result (25).

That the relativistic motion in the neighborhood of $E = B$ for $\mathbf{E} \perp \mathbf{B}$ is described by different functions depending on the relative strengths of E and B may lead one to suspect that there is some physical discontinuity at the point $E = B$. The answer to this question is probably negative because we expect the solution P to the differential equation,

$$\left(\frac{d^2}{dy^2} + V(y)\right)P(y) = 0, \tag{70}$$

to vary continuously as $V(y)$ changes smoothly from a

quadratic to a linear function in y. This is also suggested by the discussion following (42). We note that this problem also has its counterpart in the classical motions.[1]

As for the nonrelativistic motions in orthogonal fields, we note that the new feature introduced by the presence of the electric field when compared to the pure magnetic field case is that the degeneracy in p_x is now broken and gives an added constant of $Ep_x/B - (m/2)(E/B)^2$ to the energy. Therefore, for the special case of $p_x = 0$, the electric field amounts to only a uniform shift of the energy levels.

The Klein–Gordon equation has been shown to be capable of a single-particle interpretation with the use of the indefinite metric by Feshbach and Villars[23] in the spirit of the work of Pauli and Weisskopf.[24] Recently, Marx[25] has given a probabilistic interpretation to the scattering processes of a Klein–Gordon particle subjected to external electromagnetic fields within the above framework. Along this line, perturbation calculations were suggested and carried out to the lowest order for the pair annihilation process.[26] An interesting question remains on the possibility of comparing the perturbation series with some exact calculations in order to test the convergence of the series and the probabilistic interpretation itself. The exact solutions presented in the preceding section when properly handled may be able to throw light on this subject. This is currently under investigation.

ACKNOWLEDGMENTS

We thank C. K. Au, S. L. Chung, S. H. Hsieh, and T. T. Truong for useful discussions and K. Y. Ho for an expert consultation on the theory of differential equations. J. M. Luttinger and P. M. Platzman have kindly read the manuscript and given helpful comments. The financial support and hospitality from Bell Telephone Laboratories at Murray Hill are gratefully acknowledged.

[1] L. Landau and E. Lifshitz, *Classical Theory of Fields* (Pergamon, New York, 1962).
[2] J. Rander *et al.*, Phys. Rev. Letters **24**, 283 (1970).
[3] L. Lam, Master thesis, University of British Columbia, 1968 (unpublished); D. McMillin, P. Bhargava, L. Lam, and E. W. Vogt, Can. J. Phys. **46**, 1141 (1968).
[4] "A Report on the TRIUMF Project," January 7, 1966, University of British Columbia (unpublished).
[5] F. Occhionero and M. Demianski, Phys. Rev. Letters **23**, 1128 (1969).
[6] D. Datlowe and P. Meyer, Bull. Am. Phys. Soc. **15**, 611 (1970).
[7] C. G. Darwin, Proc. Roy. Soc. (London) **A118**, 654 (1928); W. Gordon, Z. Physik **48**, 11 (1928).
[8] I. I. Rabi, Z. Physik **49**, 7 (1928).
[9] F. Sauter, Z. Physik **69**, 742 (1931).
[10] D. M. Volkov, Z. Physik **94**, 250 (1935).
[11] P. J. Redmond, J. Math. Phys. **6**, 1163 (1965).
[12] G. N. Stanciu, J. Math. Phys. **8**, 2043 (1967).
[13] I. A. Malkin and V. I. Man'ko, Yad. Fiz. **8**, 1264 (1968) [Sov. J. Nucl. Phys. **8**, 731 (1969)].
[14] L. Lam, to be published.
[15] In this paper, we take $\hbar = c = 1$.
[16] We note incidentally that the corresponding formula for β given on p. 73 of Ref. 1 carries a misprint in sign.
[17] L. Lam, Can. J. Phys. **48**, 1935 (1970).
[18] See, e.g., L. Landau and E. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1965).
[19] The existence of exact solutions for this case was first pointed out by L. Lam, Phys. Letters **31A**, 406 (1970).
[20] See, e.g., Wang Chu-Chi and Kuo Tun-Jen, *Theory of Special Functions* (Science Press, Peking,1965).
[21] The existence of exact solutions for this case was first pointed out in Ref. 17.
[22] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[23] H. Feshbach and F. Villas, Rev. Mod. Phys. **30**, 24 (1958).
[24] W. Pauli and V. F. Weisskopf, Helv. Phys. Acta **7**, 709 (1934).
[25] E. Marx, Nuovo Cimento **60A**, 669 (1969).
[26] J. F. Walter and E. Marx, "Pair Annihilation at a Potential Barrier in Time," Drexel Univ. preprint, 1970.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Approximate Topologies*

L. S. SCHULMAN†
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 17 July 1970)

Several examples illustrate the utility of mathematical models which differ in their topological properties from the "real" systems they describe. The examples are the Aharonov–Bohm effect, the periodic lattice, and scattering off an obstacle. By going to the approximate topology, the Hamiltonian loses its essential self-adjointness, but the relation of the approximate to the exact topology leads to a convenient classification of its extensions.

INTRODUCTION

In discussions of the Aharonov–Bohm effect,[1,2] one sometimes invokes the concept of multiple connectedness, but then if this idea is pushed too hard, one backs off and drops some of its consequences while attempting to retain others. Thus, in Ref. 2 (p. 1514) the solenoid and its effects are adiabatically eliminated to prevent use of the license provided by multiple connectedness to employ multivalued wavefunctions.

The purpose of this article is to provide a consistent framework for such procedures. But even more, we show that the license for multivaluedness can and sometimes should be used. This will occur in physical problems where the mathematical model of the coordinate space is topologically different from what we consider to be the "real" space. Nevertheless, by taking the model topology seriously we shall obtain well-defined and physically meaningful results.

In advance of our examples, we state briefly the mathematical principle that is operating. On the approximate coordinate space the Hamiltonian is not essentially self-adjoint[3]; it does possess this property, though, on a covering space. The propagator on the covering space is therefore well defined. On the original space, the propagator is some linear combination of covering space propagators with the coefficients in this linear combination selecting some self-adjoint extension of the original Hamiltonian. How generally this procedure can be applied to a characterization of the extensions of operators which are not essentially self-adjoint would seem to be an interesting mathematical question raised by this article.

As examples, we shall discuss the Aharonov–Bohm effect, the periodic lattice (in a solid, say), and scattering off an opaque body. Also included is a case where approximating the topology would lead to incorrect conclusions.

I. THE AHARONOV–BOHM EFFECT

The Aharonov–Bohm effect can be idealized as follows (Fig. 1): Electrons are emitted from the source and those that ultimately arrive at the detector take either path 0 or path 1 through field-free regions around the solenoid.[4] The wavefunction, when there is no field in the solenoid, is mainly composed of two pieces $\psi_0^{(0)}(x)$ and $\psi_1^{(0)}(x)$, corresponding to the two paths. When the field is on, although $\mathbf{B} \neq 0$ only inside the solenoid, the vector potential $\mathbf{A}(\mathbf{x})$ is nonzero outside. The solution to Schrödinger's equation in the presence of $\mathbf{A}$ is[4]

$$\psi(\mathbf{x}) = \psi_0(\mathbf{x}) + \psi_1(\mathbf{x}), \tag{1}$$

$$\psi_j(\mathbf{x}) = \psi_j^{(0)}(\mathbf{x}) \exp\left\{ie\int_{\text{path } j}^{\mathbf{x}} \mathbf{A}\cdot d\mathbf{s}\right\}, \quad j = 0, 1. \tag{2}$$

Because the integral of $\mathbf{A}$ along different paths is different, interference patterns are created which depend on $\mathbf{A}$ and, in particular, on the integral of $\mathbf{A}$ around the closed circuit which is just the total magnetic flux.

We now analyze this in terms of a multiply connected space.[5] The mathematical object we wish to compute is the Green's function for some given time evaluated at the source and the detector. For multiply connected spaces, this can be obtained[6] most conveniently by going to the universal covering space, performing the usual path integral in this simply connected space to each of the pre-images (under the covering projection) of the detector, and summing the contributions. If we drop the z direction (along the axis of the solenoid), the space of interest M is the plane minus a disk. The covering space of M, denoted M^*, is essentially the same as the Riemann surface for the logarithm. The Green's function is therefore[6]

$$G(x_2, t; x_1, 0) = \sum_{n=-\infty}^{\infty} e^{in\alpha} G_n(x_2, t_2; x_1, 0), \tag{3}$$

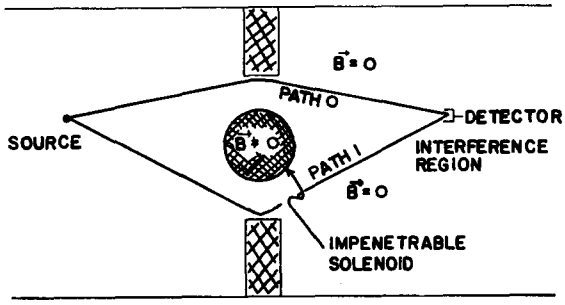

FIG. 1. Idealized Aharonov–Bohm experiment.

where G_n is the sum over paths (of $e^{iS/\hbar}$; see, e.g., Ref. 6) which loop around the solenoid n times (negative n for clockwise loops). In Fig. 2 are shown some of the paths in M, and in Fig. 3 the corresponding paths in M^*. From Fig. 3 it is evident that a better notation for $G_n(x_2, t_2; x_1, 0)$ would be $G^c(x_2^{(n)}, t_2; x_1, 0)$, where $x_2^{(n)}$ is the nth pre-image of x_2 under the covering projection and G^c is a single Green's function defined on M^*. Of course, there are many choices for x_1 also, but one has been picked for definiteness.

The phases $n\alpha$ are selected so that rotation of one of the end points through 2π changes G by an x-independent phase factor, which is unobservable. Since M is multiply connected, we cannot use the requirement of single valuedness to dispose of α.

Forget for the moment about magnetic fields, and consider the physical requirements for a Green's function describing a particle which is free except for an "impenetrable" cylinder. Since the impenetrability is an idealization, we cannot tolerate multivalued wavefunctions (the space is, in principle, simply connected), and it is necessary to set $\alpha = 0$.

With a magnetic field in the solenoid, but not outside, there is a nonzero vector potential in M. Now unless one specifies a definite path and does not move the end point around too much (like all around the solenoid), the trick used in Eq. (2) (i.e., different gauge transformations for different semiclassical paths) cannot work. However, by remaining in M^*, this trick will always work since it is merely a gauge transformation. That is, since $\mathbf{B} = 0$ in M^*, $\mathbf{A} = \nabla\Omega$ for some function Ω. Now $\Omega(x)$ can be taken to be the

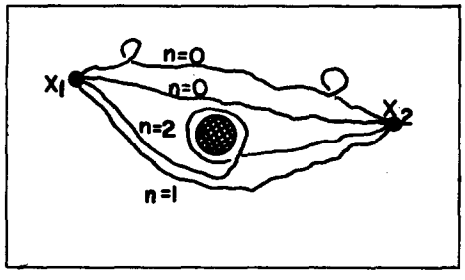

FIG. 2. Paths from source to detector.

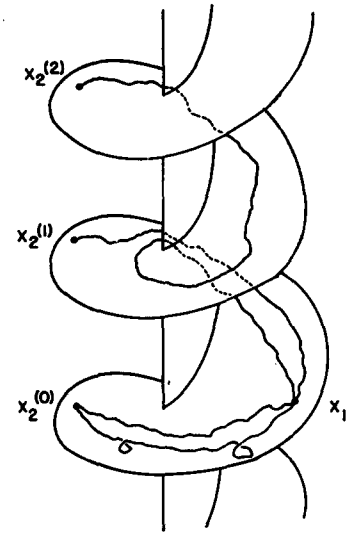

FIG. 3. Paths in the covering space.

integral $\int \mathbf{A} \cdot d\mathbf{s}$ from some fiducial point to x, and is unique because M^* is simply connected. If G_0^c is the Green's function in M^* for $\mathbf{A} = 0$, the Green's function for $\mathbf{A} = \nabla\Omega \neq 0$ is

$$G_{\mathbf{A}}^c = G_0^c(\xi_2, t; \xi_1, 0) \exp\{ie[\Omega(\xi_2) - \Omega(\xi_1)]\}, \quad \xi_1, \xi_2 \in M \tag{4}$$

$[-\Omega(\xi_1)$ is added for symmetry].

The presence of $\exp(ie\Omega)$ does not affect the periodicity arguments offered previously, and, by taking once more $\alpha = 0$ to be a physically imposed requirement, the proper Green's function is obtained.

The formulation of the foregoing procedure from the standpoint of the space M alone (as opposed to looking inside the solenoid) is also possible and, in fact, is the whole point of the present paper. On M a Hamiltonian (with or without $\mathbf{A}$) is given, but because M is multiply connected, the Hamiltonian by itself does not serve to define the dynamics. That the representation of the Hamiltonian as a differential or multiplicative operator may be inadequate has long been emphasized by mathematicians, and indeed, with particular reference to physics, the essential self-adjointness of a Hamiltonian is an important question. In a 1966 lecture,[7] Wightman discussed the physical significance of this property. For example, for a free particle in a box the Hamiltonian is not specified as an operator until the boundary conditions (e.g., vanishing at edges) are given. These boundary conditions are thus dynamical data and the physics of the situation is determined by and determines them. Our space M is quite similar to the box and the quantity which is not determined by the Hamiltonian (in its representation as a differential and multiplicative operator) is the parameter α. Our argument

concerning α involved looking more closely at the hole in M; that is, additional physical information was marshalled to complete the specification of the problem.

It is possible to recast the problem so that the freedom offered by the phase α does not remain unexploited. This formulation may also have some bearing on the role of potentials in quantum mechanics (which was the main object of the Aharanov–Bohm investigation) since the potentials are eliminated.

The function Ω introduced above, satisfies

$$\Omega(x^{(n+1)}) = \Omega(x^{(n)}) + F, \tag{5}$$

where

$$F = \int \mathbf{A} \cdot d\mathbf{s} \tag{6}$$

is independent of x and n. On M^*, the polar angle in M likewise becomes a single-valued function and satisfies (with obvious notation)

$$\varphi(x^{(n+1)}) = \varphi(x^{(n)}) + 2\pi. \tag{7}$$

φ can be used to define a function $\omega \equiv \Omega - F\varphi/2\pi$ which is periodic in M^* and hence well defined in M.

In general, gauge transformations on M^* affect the physics on M. The function ω, however, can be transformed away because it is single valued on M (and surrounds no flux lines anywhere). Writing

$$\Omega = (\Omega - F\varphi/2\pi) + F\varphi/2\pi = \omega + F\varphi/2\pi, \tag{8}$$

we see that there is no point in considering vector potentials other than $\nabla F\varphi/2\pi$. Letting

$$\Phi(x^{(n)}) \equiv \varphi(x^{(n)}) - 2n\pi, \tag{9}$$

we can write the Green's function on M,

$$G(x_2, t; x_1, 0) = \exp\{ieF/2\pi[\Phi(x_2) - \Phi(x_1)]\} \times \sum_{n=-\infty}^{\infty} \exp(ieFn) G_n(x_2, t; x_1, 0). \tag{10}$$

Evidentally, for interference experiments of the kind discussed by Aharonov and Bohm, the phase factor outside the sum in Eq. (10) plays no role. This suggests that it can be eliminated altogether.

The Green's function of Eq. (10) is single valued and propagates single-valued functions on M forward in time:

$$\psi(x_2, t) = \int_M dx_1 G(x_2, t; x_1, 0)\psi(x_1, 0). \tag{11}$$

All quantities in this equation can be redefined so that they are all multivalued in the same way and the equation continues to hold. Let

$$\psi_{\text{new}}(x, t) = \exp[(-ieF/2\pi)\Phi(x)]\psi_{\text{new}}(x, t),$$
$$G_{\text{new}} = \exp\{-ieF/2\pi[\Phi(x_2) - \Phi(x_1)]\} G_{\text{old}}. \tag{12}$$

The new quantities have the property

$$f(x^+) = \exp(-ieF) f(x^-), \tag{13}$$

where f denotes ψ_{new} or G_{new} and x^+ and x^- are the same point in physical space, but on opposite sides of the cut which runs through M from the solenoid to ∞. (This could be better stated in terms of limits, but the meaning is clear.) The result of these redefinitions is a multivalued wavefunction on a multiply connected space being propagated by a multivalued Green's function. The multivaluedness depends on a single parameter F. There need be no vector potential in M, and the physical effects of the solenoid are all included in this parameter F.

Thus, because the space is multiply connected, the Hamiltonian is not essentially self-adjoint and admits a one-parameter family of extensions. Extending the Hamiltonian means putting in new physical information. The information in this case involves a region outside M, namely the solenoid, and the particular quantity of interest is $F = \int \mathbf{B} \cdot d\boldsymbol{\sigma}$, an integral over the cross section of the solenoid.

Our purpose in the foregoing discussion was a demonstration of the idea of approximate topology and of its relation to lack of essential self-adjointness, but we cannot leave unnoticed the fact that the physical consequences of the Aharonov–Bohm effect have been obtained without the use of an electromagnetic potential. We exclude the electron from the solenoid, create an ambiguity, and look to physical quantities associated with the solenoid (total flux) to resolve this ambiguity. Aharonov and Bohm also exclude the electron from the solenoid but retain a vector potential which "remembers" what is going on inside. It seems fair to say that these are just two different ways to do quantum mechanics.

The problem of two (or more) solenoids leads, in our formalism, to an amusing situation. The fundamental (homotopy) group of two nonintersecting solenoids is the same as that of the figure eight and is not commutative. But each homotopy class has associated with it a certain phase factor in the Green's function, and it follows from the path integral representation that this factor is a 1-dimensional unitary representation of the fundamental group. The solution of the physical problem therefore involves a commutative representation of a noncommutative group.

II. ELECTRON IN A LATTICE

This example is an application of the idea of approximate topology to solid-state physics,[8,9] and gives some additional perspective to some of the current work in this area. By considering cells of a

lattice to be identical, calculations can be confined to a single cell. However, since this cell is now considered "all" cells, opposite sides of the cell are identified with one another and the space of interest (say, M) is topologically a 3-dimensional torus. The fundamental group of M is Z^3, the additive group of triples of integers, and the universal covering space is R^3 which in a way brings us back to the original solid. The unitary representations of Z^3 [i.e., the phase factors appearing in the Green's function, as in Eq. (3)] are labeled by triples of real numbers, conveniently gathered into a 3-vector $\mathbf{k}$. By noticing that, as in Eq. (13) above, $\mathbf{k}$ appears in the boundary conditions on the wavefunctions,

$$\psi(\xi + \mathbf{R}) = \exp(i\mathbf{k} \cdot \mathbf{R})\psi(\xi) \qquad (14)$$

(where ξ and $\xi + \mathbf{R}$ are on the faces of the cell, separated by the lattice vector $\mathbf{R}$), the physical meaning of $\mathbf{k}$ is immediately identified: It is the momentum label for the Bloch functions.

Once again, there is a Hamiltonian defined on a multiply connected space M, and it is not essentially self-adjoint. The various extensions are labeled by parameters with physical meaning which influence the dynamics of the system.

There are several ways in which perfect periodicity of the lattice is only an approximate description of nature. The solid is at some finite temperature, there are lattice defects, it is of finite size, one-electron theory itself is a rather narrow framework, etc. How do these physical features dictate the nature of the extended Hamiltonians and the way in which they are used?

An electron in a real solid cannot exist as a perfect Bloch wave, for otherwise its wavefunction would extend through all space. This leads to a conclusion which may be greeted with some surprise. We do not extend the Hamiltonian once and for all, but rather superpositions of the eigenfunctions of different extensions of the Hamiltonian combine to form the physical state. Furthermore, if we are given some initial wavefunction in the cell (M), there is no unique combination (superposition) of Bloch functions to which it corresponds. (*Proof:* Consider a function vanishing on the surfaces of the cell.) Thus, if one wishes to propagate a state in M, it is not enough to give some initial $\psi(\mathbf{x})$, but rather for each $\mathbf{k}$ he must specify $\psi_{\mathbf{k}}(\mathbf{x})$ and propagate with its particular extension of the Hamiltonian (or with the Green's function $G_{\mathbf{k}}$ of Ref. 8). If, instead of the notation $\psi_{\mathbf{k}}(\mathbf{x})$, we were to write $C(\mathbf{k}, \mathbf{q})$, it would become apparent that we have made the connection with Zak's kq representation[10] indicated in Ref. 8.

The handling of some of the defects within the perfect periodicity model takes the form of a dynamics in $\mathbf{k}$. Besides the need for superpositions, there will be transitions from one value of $\mathbf{k}$ to another. Some aspects of this were treated in Ref. 8 in connection with external electromagnetic fields; for example, a constant uniform electric field causes $\mathbf{k} \to \mathbf{k} + e\mathbf{E}t$.

Uniform electric and magnetic fields may be more easily adaptable to the perfect lattice model than other kinds of defects, since for them it is only the potentials which are different in different cells, while the fields are the same.

III. SCATTERING FROM AN OBSTACLE

A final example of a model of a physical system with a topology different from the "real" system may be of interest, although it is different in spirit from those already discussed. Buslaev[11] has given a path integral derivation of the short-time or long-wavelength limit of the Green's function for diffraction by a smooth obstacle. Rather than consider separately Dirichlet or Neumann boundary conditions at the surface of the obstacle, he employs the following construction: Call the accessible subset of R^3, d. The boundary of d is called l and is assumed smooth. The complement of d is the obstacle, assumed to be convex. Let D be a two-sheeted domain, consisting of two copies of d joined together along l. D covers d, and Green's functions in d are computed from those in D by adding those corresponding to the given points in d. Depending on the relative sign in this combination, Dirichlet or Neumann boundary conditions are obtained. A more elaborate version of this method is also indicated on p. 74 of Ref. 11. We distinguish this technique from the previous examples, however, because d is simply connected. Nevertheless, the model serves Buslaev's purpose very well, and the physical boundary conditions once again are expressed in the way the Green's function on d is built from that on D.

IV. THE TOP

Finally, there are situations where an approximate topology cannot be used. This occurs when a physical system is, in principle, multiply connected, and a description of it in terms of a simply connected space will eliminate physical states known to be present. For example, a model of a particle with intrinsic spin is the top, whose coordinates can be taken to be the group manifold of the group $SO(3)$.[6,12] This space is multiply connected, and the model is appropriate for particles of half-integral spin. However, if one considers this top to be the limit of n tightly bound spinless particles, the coordinate space is R^{3n}, which is

simply connected. The n particle model of the top can therefore never exhibit half-integral angular momentum.

ACKNOWLEDGMENTS

I would like to thank Professor A. Lenard and Professor J. S. Dowker for having stimulated my interest in some of these questions.

* Supported in part by the National Science Foundation and the U.S. Army Research Office (Durham).

† Address during academic year 1970–1971: Physics Department, Technion-Israel Institute of Technology, Haifa, Israel.

[1] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2] Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511 (1961).

[3] F. Reisz and B. Sz.-Nagy, *Functional Analysis* (Unger, New York, 1955). Section 119 of this book gives examples of direct relevance to the present work. "Essential" self-adjointness means a given object has a unique self-adjoint extension, although in a particular context the full description may not be specified.

[4] The discussion here presumes a semiclassical approximation.

[5] Roughly speaking, a space M is simply connected if all curves (between given end points; let the two end points coincide for this discussion) can be continuously deformed into one another. If this cannot be done, M is multiply connected, and the curves fall into different classes, called homotopy classes, within which curves are deformable into one another. Two classes give a third class by attaching the heads of curves in one class to the tails of those in the other. There is a group structure under this operation and the group is called the fundamental group of M. For each space M there is a "universal covering space" M^* which locally looks very much like M, but is simply connected. The map $p: M^* \to M$ which is locally 1–1 is called the covering projection and for each point of M has a pre-image for each element of the fundamental group. Thus $SU(2)$ is the universal covering space for $SO(3)$. The covering projection is the usual homomorphism from $SU(2)$ to $SO(3)$. The pre-images of the identity in $SO(3)$ are $\pm I$, where I is the 2×2 identity matrix.

[6] L. S. Schulman, Phys. Rev. **176**, 1558 (1968).

[7] A. S. Wightman, in *Theoretical Physics* (Proceedings of the Fifth Eastern Theoretical Physics Conference at Brown University, 1966), D. Feldman, Ed. (Benjamin, New York, 1967).

[8] L. S. Schulman, Phys. Rev. **188**, 1139 (1969).

[9] T. Lukes, Phys. Letters **29A**, 69 (1969).

[10] J. Zak, Phys. Rev. Letters **19**, 1385 (1967); Phys. Rev. **168**, 686 (1968); **177**, 1151 (1969).

[11] V. S. Buslaev, "Continuum Integrals and the Asymptotic Behavior of the Solutions of Parabolic Equations as $t \to 0$. Applications to Diffraction," in *Topics in Mathematical Physics*, Sh. Birman, Ed. (Consultants Bureau, New York, 1967), p. 67.

[12] F. Bopp and R. Haag, Z. Naturforsch. **5a**, 644 (1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Conserved Particle Structures in General Relativity

J. G. WILLIAMS*
Theoretical Physics Institute, University of Alberta, Edmonton, Canada

(Received 18 May 1970)

When general relativity is examined from the point of view of algebraic topology, it is found that the theory exhibits conserved particlelike structures. The number of such particles associated with a given metric is determined by the homotopy class of the metric. This paper is concerned with showing that if the theory is quantized, then the quantum mechanical states corresponding to 1-particle systems in the unquantized theory may have properties similar to fermions. In particular, it is proved that it is possible for wavefunctionals to exist which are double-valued under 2π rotation.

1. INTRODUCTION

A number of studies[1–5] have been made of field theories which possess conserved particlelike structures called "kinks." Such theories are said to "admit kinks." The fields are classical and so single-valued under the action of the rotation group. The aim of these studies is to examine the possibility that when theories of this type are quantized, the quantum mechanical states corresponding to 1-kink classical field configurations are, in fact, fermion states.

An example of a theory that admits kinks is the nonlinear theory of mesons and baryons suggested by Skyrme.[1] This involves mappings φ from 3-dimensional space R^3 into the 3-sphere S^3:

$$\varphi: R^3 \to S^3.$$

S^3 can be parametrized by four real variables $(\phi_1, \phi_2, \phi_3, \phi_4)$ subject to the restriction

$$\sum_{i=1}^{4} \phi_i^2 = 1.$$

To prevent the escape of interesting structures at infinity, we only consider mappings φ which satisfy the boundary condition

$$\varphi(\mathbf{x}) \to (0, 0, 0, 1) \quad \text{as} \quad |\mathbf{x}| \to \infty,$$

where $\mathbf{x}$ is any point in R^3 and $(0, 0, 0, 1)$ is a fixed point in S^3. Among the ϕ_i there are three independent fields that may be used to represent the three π-meson fields. The number of kinks present equals the degree of the mapping φ. This can take on any integer value, positive or negative, and it is hoped

simply connected. The n particle model of the top can therefore never exhibit half-integral angular momentum.

ACKNOWLEDGMENTS

I would like to thank Professor A. Lenard and Professor J. S. Dowker for having stimulated my interest in some of these questions.

* Supported in part by the National Science Foundation and the U.S. Army Research Office (Durham).

† Address during academic year 1970–1971: Physics Department, Technion-Israel Institute of Technology, Haifa, Israel.

[1] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2] Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511 (1961).

[3] F. Reisz and B. Sz.-Nagy, *Functional Analysis* (Unger, New York, 1955). Section 119 of this book gives examples of direct relevance to the present work. "Essential" self-adjointness means a given object has a unique self-adjoint extension, although in a particular context the full description may not be specified.

[4] The discussion here presumes a semiclassical approximation.

[5] Roughly speaking, a space M is simply connected if all curves (between given end points; let the two end points coincide for this discussion) can be continuously deformed into one another. If this cannot be done, M is multiply connected, and the curves fall into different classes, called homotopy classes, within which curves are deformable into one another. Two classes give a third class by attaching the heads of curves in one class to the tails of those in the other. There is a group structure under this operation and the group is called the fundamental group of M. For each space M there is a "universal covering space" M^* which locally looks very much like M, but is simply connected. The map $p: M^* \to M$ which is locally 1–1 is called the covering projection and for each point of M has a pre-image for each element of the fundamental group. Thus $SU(2)$ is the universal covering space for $SO(3)$. The covering projection is the usual homomorphism from $SU(2)$ to $SO(3)$. The pre-images of the identity in $SO(3)$ are $\pm I$, where I is the 2×2 identity matrix.

[6] L. S. Schulman, Phys. Rev. **176**, 1558 (1968).

[7] A. S. Wightman, in *Theoretical Physics* (Proceedings of the Fifth Eastern Theoretical Physics Conference at Brown University, 1966), D. Feldman, Ed. (Benjamin, New York, 1967).

[8] L. S. Schulman, Phys. Rev. **188**, 1139 (1969).

[9] T. Lukes, Phys. Letters **29A**, 69 (1969).

[10] J. Zak, Phys. Rev. Letters **19**, 1385 (1967); Phys. Rev. **168**, 686 (1968); **177**, 1151 (1969).

[11] V. S. Buslaev, "Continuum Integrals and the Asymptotic Behavior of the Solutions of Parabolic Equations as $t \to 0$. Applications to Diffraction," in *Topics in Mathematical Physics*, Sh. Birman, Ed. (Consultants Bureau, New York, 1967), p. 67.

[12] F. Bopp and R. Haag, Z. Naturforsch. **5a**, 644 (1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Conserved Particle Structures in General Relativity

J. G. WILLIAMS*
Theoretical Physics Institute, University of Alberta, Edmonton, Canada

(Received 18 May 1970)

When general relativity is examined from the point of view of algebraic topology, it is found that the theory exhibits conserved particlelike structures. The number of such particles associated with a given metric is determined by the homotopy class of the metric. This paper is concerned with showing that if the theory is quantized, then the quantum mechanical states corresponding to 1-particle systems in the unquantized theory may have properties similar to fermions. In particular, it is proved that it is possible for wavefunctionals to exist which are double-valued under 2π rotation.

1. INTRODUCTION

A number of studies[1–5] have been made of field theories which possess conserved particlelike structures called "kinks." Such theories are said to "admit kinks." The fields are classical and so single-valued under the action of the rotation group. The aim of these studies is to examine the possibility that when theories of this type are quantized, the quantum mechanical states corresponding to 1-kink classical field configurations are, in fact, fermion states.

An example of a theory that admits kinks is the nonlinear theory of mesons and baryons suggested by Skyrme.[1] This involves mappings φ from 3-dimensional space R^3 into the 3-sphere S^3:

$$\varphi: R^3 \to S^3.$$

S^3 can be parametrized by four real variables $(\phi_1, \phi_2, \phi_3, \phi_4)$ subject to the restriction

$$\sum_{i=1}^{4} \phi_i^2 = 1.$$

To prevent the escape of interesting structures at infinity, we only consider mappings φ which satisfy the boundary condition

$$\varphi(\mathbf{x}) \to (0, 0, 0, 1) \quad \text{as} \quad |\mathbf{x}| \to \infty,$$

where $\mathbf{x}$ is any point in R^3 and $(0, 0, 0, 1)$ is a fixed point in S^3. Among the ϕ_i there are three independent fields that may be used to represent the three π-meson fields. The number of kinks present equals the degree of the mapping φ. This can take on any integer value, positive or negative, and it is hoped

that a kink may be interpreted as a baryon. (This theory has a simple 1-dimensional analog involving mappings from the real line R^1 into the circle S^1. A kink may then be pictured intuitively as a 2π twist in an infinitely long strip.[5,6])

A more complicated example is that of general relativity. Here the fields are represented by the metric tensor. At a particular instant of time, we may regard this as a mapping g from R^3 into the set of 4×4 real symmetric matrices of signature 1, which we denote by $S_{4,1}$:

$$g: R^3 \to S_{4,1}.$$

It is again convenient to impose boundary conditions. We shall only consider mappings that satisfy

$$g(\mathbf{x}) \to I_{4,1} \quad \text{as} \quad |\mathbf{x}| \to \infty,$$

where $I_{4,1}$ is the matrix diag $(1, 1, 1, -1)$.

Since $S_{4,1}$ is a manifold of higher dimension than R^3, the usual concept of degree does not apply to the mappings g. However, the idea of the degree of a mapping can be generalized by introducing the more powerful techniques of algebraic topology, and a mapping can be specified by the homotopy class to which it belongs. Mappings belong to the same homotopy class if and only if they are continuously deformable into each other (homotopic). The homotopy classes of many manifolds (including $S_{4,1}$, as we shall see) can be labeled by a set of integers.[7] Since time is a continuous parameter, these labels can not change as time evolves, and so they may be interpreted as conserved particle numbers.

Given a theory that admits kinks, we may investigate the 1-kink mappings to determine whether or not they have fermionlike properties. A characteristic property of a many-fermion state is the double-valuedness of the wavefunctional under exchange. For the type of theories considered in this paper, it has been shown[3,4] that double-valuedness under exchange implies the double-valuedness of the one-kink wavefunctional under 2π rotation. If a theory is given which admits kinks and if it is also possible to define, on the space of 1-kink fields, functionals which are double-valued under 2π rotation, then we shall say that the theory "admits spin."

In a previous paper,[8] the author has proved that the nonlinear theory of Skyrme admits spin. On the basis of this, it is the purpose of the present paper to show that general relativity admits spin.

2. THEORY OF KINKS

Before beginning the investigation of the topological structure of $S_{4,1}$, it will be convenient to review the general theory of kinks due to Finkelstein.[2-5,8] Let θ be a mapping from 3-dimensional Euclidean space R^3 into a pathwise connected manifold Y,

$$\theta: R^3 \to Y,$$

subject to the boundary condition

$$\theta(\mathbf{x}) \to y_0 \quad \text{as} \quad |\mathbf{x}| \to \infty.$$

$\mathbf{x}$ is any point of R^3, and y_0 is a fixed point of Y. We shall let Q denote the set of all such mappings θ. The set of pathwise connected components Q_a, Q_b, $Q_c, \cdots$ of Q are called homotopy classes and form an Abelian group called the third homotopy group $\pi_3(Y)$. If $\pi_3(Y) \neq 0$, we shall say that the theory admits kinks. The 1-kink mappings belong to the generators of $\pi_3(Y)$.

Let Q_1 denote a generator of $\pi_3(Y)$. We shall investigate the conditions under which functionals may be defined on Q_1 which are double-valued under 2π rotation. (The concept of "double-valuedness" can be formulated precisely by introducing the idea of universal covering space.[2,9]) If it is possible to define a double-valued functional on Q_1, then the space of closed paths in Q_1 must contain two different classes of paths such that all paths belonging to one class are deformable to a point (trivial paths), whereas all paths belonging to the other class are not deformable to a point (nontrivial paths). A path formed by describing a nontrivial path twice must be trivial. The quantity which contains the information about the path structure of Q_1 is the first homotopy group $\pi_1(Q_1)$. This is the set of different classes of closed paths in Q_1 which pass through some fixed point $\theta_1 \in Q_1$. A necessary and sufficient condition for the existence of double-valued functionals on Q_1 is that the group $\pi_1(Q_1)$ have an element of order 2. The existence of double-valued functionals alone does not imply the existence of spin properties. One must show that the double-valuedness arises through following a path, starting and ending at some $\theta_1 \in Q_1$, which corresponds to rotating the system through 2π. When this happens, we shall follow the terminology of Finkelstein and say that the theory "admits spin." One need only prove the double-valuedness for a single member θ_1 of Q_1 and a single 2π rotation. It then follows[3,4] that the double-valuedness holds for all members of Q_1 and for all 2π rotations.

$\pi_1(Q_1)$ is evaluated by using the isomorphism[10]

$$\pi_1(Q_a) \simeq \pi_1(Q_b).$$

Let θ_0 denote the constant mapping which maps the whole of R^3 into the fixed point y_0 of Y. Mappings which are homotopic to θ_0 are called trivial. The

homotopy class of such mappings will be called Q_0. Thus

$$\pi_1(Q_1) \simeq \pi_1(Q_0).$$

$\pi_1(Q_0)$ can be obtained from

$$\pi_1(Q_0) \simeq \pi_4(Y).$$

This latter isomorphism results from the fact that the space of closed paths in Q_0 (beginning and ending at θ_0) is homeomorphic to the space of mappings from R^4 into Y (with appropriate boundary conditions).[11]

Another useful fact is that if Y is a topological group, then the different homotopy classes Q_a are homeomorphic to each other. The reasoning is as follows. For any given $\theta, \sigma \in Q$ we may define a group operation * in Q by

$$(\theta * \sigma)(\mathbf{x}) = \theta(\mathbf{x}) \circ \sigma(\mathbf{x}),$$

where $\circ$ is the group operation in Y. Thus, Q is a topological group, and so its pathwise connected components, namely the Q_a, must be homeomorphic to each other.[12] Furthermore, it can be shown[13] that if $\theta \in Q_a$ and $\sigma \in Q_b$, then $\theta * \sigma \in Q_a + Q_b$, where $+$ is the group operation in $\pi_3(Y)$.

We are now in a position to discuss some examples. Consider the theory for which $Y = S^3$. It is well known that

$$\pi_3(S^3) = Z,$$

where Z denotes the group of integers. Thus the theory admits kinks. The homotopy classes can be labeled by a single integer $i, \cdots, Q_{-2}, Q_{-1}, Q_0, Q_1, Q_2 \cdots$. An example of a 1-kink mapping is the stereographic projection defined by

$$\begin{aligned} \phi_i &= 2ax_i/(r^2 + a^2), \quad i = 1, 2, 3, \\ \phi_4 &= (r^2 - a^2)/(r^2 + a^2), \end{aligned} \tag{1}$$

where $(\phi_1, \phi_2, \phi_3, \phi_4) \in S^3$, $\mathbf{x} \in R^3$, and r denotes $|\mathbf{x}|$. It is also well known that

$$\pi_4(S^3) = Z_2,$$

where Z_2 denotes the group of integers modulo 2. The question now arises as to whether or not this theory admits spin. This question has now been answered in the affirmative.[8]

With regard to general relativity, it was noted by Finkelstein and Misner[5] that

$$\pi_3(S_{4,1}) = Z \tag{2}$$

and

$$\pi_4(S_{4,1}) = Z_2. \tag{3}$$

In Sec. 4 of this paper it will be shown that general relativity admits spin. However, it will first be necessary to review some of the topological properties of $S_{4,1}$ and to understand why Eqs. (2) and (3) are true.

3. $S_{4,1}$ AS A BUNDLE SPACE

Any real nonsingular matrix A can be written in one and only one way as a product,

$$A = QS,$$

where Q is a real orthogonal matrix and S is a symmetric positive-definite matrix. Let O_4 denote the group of 4×4 orthogonal matrices. It can be proved[14] that $A \in S_{4,1}$ if and only if $Q \in O_4 \cap S_{4,1}$ and $QS = SQ$. Furthermore, if O_n is the group of $n \times n$ orthogonal matrices and $S_{n,k}$ is the set of $n \times n$ real symmetric matrices of signature k, then it can be shown[14] that $O_n \cap S_{n,k}$ is homeomorphic to the set of k planes through the origin in R^n. This set is called the Grassmann manifold $M_{n,k}$. $M_{4,1}$ is the set of straight lines through the origin in R^4 and is commonly called real projective space P^3. It is easy to show[15] that P^3 is homeomorphic to the 3-dimensional rotation group SO_3. Hence $Q \in SO_3$.

We define the mapping

$$p: S_{4,1} \to SO_3$$

by

$$p(QS) = Q.$$

$S_{4,1}$ is a bundle space with SO_3 as base and p as projection. This is a special case of the more general result[14] that $S_{n,k}$ is a bundle space with $M_{n,k}$ as base. The fiber F, for the $S_{4,1}$ case, is defined to be $p^{-1}(I_{4,1})$. This is the set of matrices of the form

$$F = \begin{pmatrix} B & 0 \\ 0 & b \end{pmatrix},$$

where B is a 3×3 symmetric positive-definite matrix and b is a negative real number. A space which is homeomorphic to a Euclidean space is called a cell. The space of definite matrices of a fixed order is a cell. It follows that F is the product of two cells and so is a cell. Hence, from the point of view of algebraic topology, F is trivial and

$$\pi_n(F) = 0, \quad \text{all } n.$$

Because of this, one expects all of the interesting topological properties of $S_{4,1}$ to be displayed by SO_3.

If X and Y are any two spaces, a mapping

$$f: X \to Y,$$

with an appropriate boundary condition, can be used to define a homomorphism

$$f_*: \pi_n(X) \to \pi_n(Y).$$

Given a mapping $\theta: R^n \to X$, then we see that the composition $f\theta$ is a mapping from R^n into Y. We denote the homotopy class of θ by $[\theta]$ and that of $f\theta$ by $[f\theta]$; f_* is defined by

$$f_*[\theta] = [f\theta].$$

In this way a homomorphism p_* can be defined between $\pi_n(S_{4,1})$ and $\pi_n(SO_3)$. In fact, the homotopy groups $\pi_n(S_{4,1})$, $\pi_n(SO_3)$, and $\pi_n(F)$ are related by an exact sequence[16,17]

$$\cdots \longrightarrow \pi_n(F) \xrightarrow{i_*} \pi_n(S_{4,1}) \xrightarrow{p_*} \pi_n(SO_3) \xrightarrow{d_*} \pi_{n-1}(F) \longrightarrow \cdots,$$

where i_*, p_*, and d_* denote homomorphisms. The triviality of the fiber implies

$$0 \xrightarrow{i_*} \pi_n(S_{4,1}) \xrightarrow{p_*} \pi_n(SO_3) \xrightarrow{d_*} 0,$$

and, because the sequence is exact, it follows[17] that p_* is an isomorphism. Hence

$$\pi_n(S_{4,1}) \simeq \pi_n(SO_3).$$

The homotopy groups of SO_3 are well known. Apart from the $n = 1$ case, they are identical to the homotopy groups of the universal covering group of SO_3. This is S^3 (or SU_2); $\pi_3(SO_3)$ is Z, and $\pi_4(SO_3)$ is Z_2. This proves Eqs. (2) and (3). Equation (2) indicates that kinks exist in general relativity and that they are labeled by a single integer. Equation (3) means that it is reasonable to make further investigations to ascertain whether or not the theory admits spin.

4. PROOF OF THE ADMISSION OF SPIN FOR $S_{4,1}$

Let $Q_1(Y)$ denote the set of 1-kink mappings corresponding to either of the cases $Y = S^3$, SO_3, and $S_{4,1}$. Let path $[Q_1(Y)]$ be the set of closed paths in $Q_1(Y)$ beginning and ending at some particular point in $Q_1(Y)$ and path $[Q_0(Y)]$ be the set of closed paths in $Q_0(Y)$ beginning and ending at the constant mapping which takes the whole of R^3 into a fixed point y_0 of Y. Let $M(R^4, Y)$ denote the space of mappings

$$\psi: R^4 \to Y,$$

with the boundary condition

$$\psi(u_1, u_2, u_3, u_4) \to y_0 \quad \text{as any} \quad |u_i| \to \infty,$$

where (u_1, u_2, u_3, u_4) is any point of R^4. As mentioned before, $M(R^4, Y)$ is homeomorphic to path $[Q_0(Y)]$. Thus we may construct homeomorphisms D and H:

$$D: \text{path } [Q_0(S^3)] \to M(R^4, S^3),$$
$$H: \text{path } [Q_0(SO_3)] \to M(R^4, SO_3).$$

Because S^3 and SO_3 are topological groups, $Q_1(S^3)$ is homeomorphic to $Q_0(S^3)$, and $Q_1(SO_3)$ is homeomorphic to $Q_0(SO_3)$. Thus, it is possible to define homeomorphisms d and h:

$$d: \text{path } [Q_1(S^3)] \to \text{path } [Q_0(S^3)],$$
$$h: \text{path } [Q_1(SO_3)] \to \text{path } [Q_0(SO_3)].$$

Let c denote the usual covering mapping

$$c: S^3 \to SO_3.$$

A theorem quoted by Hu (Ref. 16, p. 89) states that every covering space is a bundle space with discrete fiber. Thus S^3 is a bundle space with SO_3 as the base and c as the projection. The fact that the fiber is discrete means that all its homotopy groups are zero. This situation is similar to the case of the fibering of $S_{4,1}$ over SO_3, and we may construct an exact sequence and show that the mapping

$$c_*: \pi_n(S^3) \to \pi_n(SO_3)$$

is an isomorphism.

Suppose $\varphi_1 \in Q_1(S^3)$. Because c_* is an isomorphism, this implies that the composition $c\varphi_1 \in Q_1(SO_3)$.[18] For example, the composition of the stereographic projection of Eq. (1) and the covering mapping c belongs to $Q_1(SO_3)$. Let χ be a closed path in $Q_1(S^3)$ beginning and ending at φ_1. χ is a member of path $[Q_1(S^3)]$. $c\chi$ is a closed path in $Q_1(SO_3)$. In this sense, c can be regarded as a mapping between two path spaces:

$$c: \text{path } [Q_1(S^3)] \to \text{path } [Q_1(SO_3)].$$

On the other hand, if $F \in M(R^4, S^3)$, then $cF \in M(R^4, SO_3)$, and so c can also be regarded as a mapping between two mapping spaces:

$$c: M(R^4, S^3) \to M(R^4, SO_3).$$

Because $\pi_4(S^3) = \pi_4(SO_3) = Z_2$, there are only two classes of mappings belonging to $M(R^4, S^3)$ and only two classes of mappings belonging to $M(R^4, SO_3)$. We may draw the following diagram.

$$\begin{array}{ccccc} \text{path } [Q_1(S^3)] & \xrightarrow{d} & \text{path } [Q_0(S^3)] & \xrightarrow{D} & M(R^4, S^3) \\ \downarrow{\scriptstyle c} & & & & \downarrow{\scriptstyle c} \\ \text{path } [Q_1(SO_3)] & \xrightarrow{h} & \text{path } [Q_0(SO_3)] & \xrightarrow{H} & M(R^4, SO_3) \end{array}$$

Let $\chi \in \text{path } [Q_1(S^3)]$ be a 2π rotation path. An example of such a path is obtained by making the

replacement

$$x_i \to \sum_{j=1}^{3} R_{ij}(s)x_j$$

in Eq. (1), where

$$\|R_{ij}(s)\| = \begin{pmatrix} \cos 2\pi s & \sin 2\pi s & 0 \\ -\sin 2\pi s & \cos 2\pi s & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and where s is the path parameter varying between 0 and 1. Clearly, $c\chi \in$ path $[Q_1(SO_3)]$ is also a 2π rotation path. Because the S^3 theory admits spin, χ is a nontrivial path. We shall show that $c\chi$ is also nontrivial.

The fact that χ is nontrivial and Dd is a homeomorphism implies that $D\,d\chi$ is a nontrivial member of $M(R^4, S^3)$. Because c_* is an isomorphism, $cD\,d\chi$ is a nontrivial member of $M(R^4, SO_3)$. Since $(Hh)^{-1}$ is a homeomorphism, $(Hh)^{-1}cD\ d\chi$ is a nontrivial path belonging to path $[Q_1(SO_3)]$. Because $(Hh)^{-1}$ and Dd are homeomorphisms, the mappings c and $(Hh)^{-1}cDd$ are homotopic, and so $c\chi$ and $(Hh)^{-1}cD\,d\chi$ are deformable into each other. The nontriviality of $(Hh)^{-1}cD\,d\chi$ implies that $c\chi$ is nontrivial. Since $c\chi$ is a 2π rotation path, it follows that the SO_3 theory admits spin.

Let us now consider $S_{4,1}$ and the subspace $O_4 \cap S_{4,1}$ of $S_{4,1}$. This subspace is homeomorphic to SO_3. Let

$$g_1 : R^3 \to S_{4,1}$$

be the composition of the stereographic projection of Eq. (1) and the covering mapping c which maps S^3 onto $O_4 \cap S_{4,1}$ (or equivalently onto SO_3). Consider the composition

$$pg_1 : R^3 \to SO_3.$$

As noted previously, $pg_1 \in Q_1(SO_3)$. Because p_* is an isomorphism, $g_1 \in Q_1(S_{4,1})$. Thus an example of a 1-kink mapping in general relativity is the composition of the stereographic projection with the usual two-fold covering of SO_3 by S^3.

Since $S_{4,1}$ is a bundle space with SO_3 as base, we may draw a diagram formally similar to the S^3, SO_3 case.

$$\begin{array}{ccccc} \text{path } [Q_1(S_{4,1})] & \xrightarrow{k} & \text{path } [Q_0(S_{4,1})] & \xrightarrow{K} & M(R^4, S_{4,1}) \\ \downarrow p & & & & \downarrow p \\ \text{path } [Q_1(SO_3)] & \xrightarrow{k'} & \text{path } [Q_0(SO_3)] & \xrightarrow{H} & M(R^4, SO_3) \end{array}$$

However, since $S_{4,1}$ is not a topological group, we cannot assume that $Q_1(S_{4,1})$ is homeomorphic to $Q_0(S_{4,1})$. We shall follow Rubinstein[4] and define the mapping k as follows. Let χ be a path belonging to path $[Q_1(S_{4,1})]$ so that, for a particular s, $\chi(s) \in Q_1(S_{4,1})$. Let g_{-1} be a member of $Q_{-1}(S_{4,1})$. We may define a path $\eta \in$ path $[Q_0(S_{4,1})]$ by

$$\eta(s) = \chi(s) \cup g_{-1}. \tag{4}$$

The symbol $\cup$ represents the usual homotopy theory operation of joining two mappings. The homotopy classes $[\eta(s)]$, $[\chi(s)]$, and $[g_{-1}]$ are related by

$$[\eta(s)] = [\chi(s)] + [g_{-1}],$$

where $+$ is the group operation for $\pi_3(S_{4,1})$. Clearly, $[\eta(s)] = Q_0(S_{4,1})$, and so $\eta \in$ path $[Q_0(S_{4,1})]$. It has been shown by Rubinstein[4] that if χ is a nontrivial (trivial) path, then η is also a nontrivial (trivial) path. Equation (4) defines a mapping k from path $[Q_1(S_{4,1})]$ into path $[Q_0(S_{4,1})]$. In a similar manner, we may define a mapping k':

$$k' : \text{path } [Q_1(SO_3)] \to \text{path } [Q_0(SO_3)].$$

k and k' are not homeomorphisms.

Consider the two mappings g_1 and pg_1. (These are essentially the same mapping.) By applying a 2π rotation, g_1 and pg_1 give rise to two paths χ_1 and $p\chi_1$, $\chi_1 \in$ path $[Q_1(S_{4,1})]$ and $p\chi_1 \in$ path $[Q_1(SO_3)]$. $p\chi_1$ is nontrivial because the SO_3 theory admits spin. We shall prove that χ_1 is also nontrivial. Consider the two mappings $pKk\chi_1$ and $Hk'p\chi_1$, both belonging to $M(R^4, SO_3)$. Both k and k' map nontrivial paths into nontrivial paths. Because $p\chi_1$ is nontrivial, it follows that $Hk'p\chi_1$ is nontrivial. However, the two mappings $pKk\chi_1$ and $Hk'p\chi_1$ are, in fact, the same mapping. Therefore, $pKk\chi_1$ is nontrivial. The triviality of $pKk\chi_1$ depends upon that of χ_1. Hence, the path χ_1 is nontrivial. Since χ_1 is a 2π rotation path in $Q_1(S_{4,1})$, it follows that general relativity admits spin.

5. A 1-KINK MAPPING

It is of interest to construct explicitly an example of a 1-kink mapping in general relativity. Recalling that the composition of the stereographic projection and the covering mapping c is a 1-kink mapping, let us first construct a mapping from S^3 into $O_4 \cap S_{4,1}$ which is topologically equivalent to c. Let G be a member of $O_4 \cap S_{4,1}$. Therefore, G can be written in the form

$$G = P \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -1 \end{pmatrix} P^{-1},$$

where $P \in O_4$. A simple way of defining a mapping

from S^3 into O_4 is to write P in the form

$$P = \begin{pmatrix} -\phi_4 & \phi_3 & \phi_2 & \phi_1 \\ -\phi_3 & -\phi_4 & \phi_1 & -\phi_2 \\ -\phi_2 & -\phi_1 & -\phi_4 & \phi_3 \\ -\phi_1 & \phi_2 & -\phi_3 & -\phi_4 \end{pmatrix}.$$

Substituting this into the equation for G gives

$$G = \begin{pmatrix} 1-2\phi_1^2 & 2\phi_1\phi_2 & -2\phi_1\phi_3 & 2\phi_1\phi_4 \\ 2\phi_2\phi_1 & 1-2\phi_2^2 & 2\phi_2\phi_3 & -2\phi_2\phi_4 \\ -2\phi_3\phi_1 & 2\phi_3\phi_2 & 1-2\phi_3^2 & 2\phi_3\phi_4 \\ 2\phi_4\phi_1 & -2\phi_4\phi_2 & 2\phi_4\phi_3 & 1-2\phi_4^2 \end{pmatrix}.$$

This defines a mapping from S^3 into $O_4 \cap S_{4,1}$. Call it c'. Any member $(\phi_1, \phi_2, \phi_3, \phi_4) \in S^3$ determines G uniquely. Given any $G \in O_4 \cap S_{4,1}$, the diagonal elements determine any ϕ_i to within a sign. The off-diagonal elements of G serve to correlate the signs of the ϕ_i so that $(\phi_1, \phi_2, \phi_3, \phi_4)$ is determined to within an over-all sign. Clearly, c' is topologically equivalent to the usual twofold covering mapping c from S^3 into SO_3. The composition of c' with the stereographic projection of Eq. (1) then gives a mapping from R^3 into $S_{4,1}$, which is a member of $Q_1(S_{4,1})$ and so a 1-kink mapping.

6. SUMMARY

The aim of this paper has been to explain how general relativity can admit kinks and how the 1-kink states can have fermionlike properties. The knowledge that the S^3 theory admits spin and that S^3 is a bundle space with SO_3 as base was used to show that the SO_3 theory admits spin. The fact that $S_{4,1}$ is a bundle space with SO_3 as base then led to the conclusion that general relativity admits spin. An example of a 1-kink mapping in general relativity was constructed. Whether or not such objects play any role in nature is an open question.

ACKNOWLEDGMENTS

I am grateful to Dr. C. S. Hoo for a number of helpful discussions. I should also like to thank the University of Alberta for the award of a postdoctoral fellowship and Professor Y. Takahashi for the hospitality of the Theoretical Physics Institute.

* Address after 1 September 1970: Department of Mathematical, Physics, University of Birmingham, England.

[1] T. H. R. Skyrme, Proc. Roy. Soc. (London) **A247**, 260 (1958); **A260**, 127 (1961); Nucl. Phys. **31**, 556 (1962).
[2] D. Finkelstein, J. Math. Phys. **7**, 1218 (1966).
[3] D. Finkelstein and J. Rubinstein, J. Math. Phys. **9**, 1762 (1968).
[4] J. Rubinstein, Ph.D. thesis, Yeshiva University, 1967.
[5] D. Finkelstein and C. W. Misner, Ann. Phys. **6**, 230 (1959).
[6] T. H. R. Skyrme, Proc. Roy. Soc. (London) **A262**, 237 (1961); J. K. Perring and T. H. R. Skyrme, Nucl. Phys. **31**, 550 (1962); U. Enz, Phys. Rev. **131**, 1392 (1963); A. C. Scott, Am. J. Phys. **37**, 52 (1969); J. Rubinstein, J. Math. Phys. **11**, 258 (1970).
[7] This is true for such common manifolds as S^1, SO_3, and $S_{4,1}$ and can be shown to be true for any simply connected manifold. See S.-T. Hu, *Homotopy Theory* (Academic, New York, 1959), p. 306, Corollary 8.3.
[8] J. G. Williams, J. Math. Phys. **11**, 2611 (1970).
[9] J. G. Hocking and G. S. Young, *Topology* (Addison-Wesley, Reading, Mass., 1961).
[10] G. W. Whitehead, Ann. Math. **47**, 460 (1946).
[11] G. W. Whitehead, *Homotopy Theory* (M.I.T. Press, Cambridge, Mass., 1966), Lemma 1.4.2.
[12] D. Montgomery and L. Zippin, *Topological Transformation Groups* (Interscience, New York, 1955), p. 39.
[13] P. J. Hilton, *An Introduction to Homotopy Theory* (Cambridge U.P., Cambridge, 1966), p. 60.
[14] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, N.J., 1951), Chap. 40.
[15] D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966), p. 93.
[16] S.-T. Hu, *Homotopy Theory* (Academic, New York, 1959).
[17] R. C. Hwa and V. L. Teplitz, *Homology and Feynman Integrals* (Benjamin, New York, 1966), pp. 37–40.
[18] Strictly speaking, the fact that c_* is an isomorphism implies that either $c\varphi_1 \in Q_1(SO_3)$ or $c\varphi_1 \in Q_{-1}(SO_3)$, where $Q_{-1}(SO_3)$ is the set of 1-antikink mappings and is the inverse of $Q_1(SO_3)$ in the group $\pi_3(SO_3)$. Since the kink or antikink nature of $Q_1(SO_3)$ or $Q_{-1}(SO_3)$ can be assigned in an arbitrary manner, we shall ignore this complication.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 2 FEBRUARY 1971

Application of the Semidirect Product of Groups*

ROBERT GEROCH†
Department of Physics, Syracuse University, Syracuse, New York
AND
E. T. NEWMAN
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 10 August 1970)

A simple example of the application of the semidirect product of groups is presented.

The asymptotic symmetry group in general relativity, the BMS group, was introduced by Sachs.[1] We show that this group, as well as the Poincaré group and the invariance group of electrodynamics, are examples of the semidirect product applied to certain group representations.[2]

Let G be a group and V an Abelian group. Suppose we are given in addition an action of G on V. That is to say, for every $g \in G$ we have an automorphism A_g on V such that $A_{g_1} \circ A_{g_2} = A_{g_1 g_2}$ and $(A_g)^{-1} = A_{g^{-1}}$. The semidirect product $V \circledS G$ of V and G is then defined as follows. The underlying set of $V \circledS G$ is the product of sets $V \times G$. The group operation is defined by[3]

$$(g_1, v_1)(g_2, v_2) = (g_1 g_2, v_1 + A_{g_1}(v_2)).$$

In particular, if V is the vector space of a representation of G, then G acts on V as required. With each representation of G we thus associate a new group. We can consider V as an Abelian normal subgroup of $V \circledS G$: G is the factor group.

We discuss three applications of these remarks. Let $G = \mathfrak{L}$, the (homogeneous) Lorentz group.

Suppose first that we choose for V the vector representation W of $\mathfrak{L}$, i.e., the $D^{(\frac{1}{2},\frac{1}{2})}$ representation. Then $\mathfrak{L} \circledS W$ is the inhomogeneous Lorentz group $\mathfrak{P}$.

Next, consider $\mathfrak{L}$ as the group of conformal mappings on the Riemann sphere S. Let X be the vector space of smooth, real-valued functions on S. Define the action of $\mathfrak{L}$ on X as follows: If $L \in \mathfrak{L}$, $f \in X$, $Lf(p) = Kf[L(p)]$, where $p \in S$ and K is the conformal factor associated with the action of L on S. (This is the infinite-dimensional representation, sometimes called the closest relative representation, associated with $D^{(\frac{1}{2},\frac{1}{2})}$.[4]) Then $\mathfrak{L} \circledS X$ is the BMS group.[1] (Similarly, one obtains "generalizations" of the BMS group.)

Finally, let $G = \mathfrak{P}$, the Poincaré group, considered as acting on Minkowski space M. Let Y be the vector space of smooth, real-valued functions on Minkowski space. If $P \in \mathfrak{P}$ and $f \in Y$, let $Pf(q) = f(q)$, $q \in M$. Then Y is a representation of $\mathfrak{P}$, and $\mathfrak{P} \circledS Y$ is the invariance group of electrodynamics. The invariance group of linearized general relativity is obtained in a similar way (choosing for Y the vector fields on Minkowski space).

ACKNOWLEDGMENTS

We wish to thank R. Penrose and J. Goldberg for useful discussions.

* This research was supported in part by the Aerospace Research Laboratory OAR, U.S. Air Force, Contract F33615-70-C-1081.

† Present address: Department of Physics, University of Texas, Austin, Texas. Supported in part by AFOSR and ARL.

[1] R. Sachs, Phys. Rev. **128**, 2851 (1962).

[2] Essentially the same conclusions have recently been reached by P. McCarthy (private communication).

[3] F. Gürsey, in *Les Houches Lectures, 1963: Relativity, Groups and Topology*, edited by C. M. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1964).

[4] A. Held, E. T. Newman, and R. Posadas, J. Math. Phys. **11**, 3145 (1970); I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. V.