# Topological solution of ordinary and partial finite difference equations

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

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

(Recieved 2 April 1980; accepted for publication 11 June 1980)

Using the discrete path formalism, we obtain a topological solution for ordinary, as well as partial, linear inhomogeneous finite difference equations with variable coefficients and arbitrarily specified boundary conditions. The solution is homomorphic to a set of discrete paths constructed from a set of vectors determined by the level differences of the equation.

## I. INTRODUCTION

A certain amount of progress[1,2,3] has been made in studying the solutions of finite difference equations, by making use of the so called combinatorics functions. These are highly generalized factorial expressions[4] that were originally obtained,[1] by a homomorphism, from the restricted partitions[5] of an interval into classes. These classes are determined by the level differences of the equation. Subsequently, the formalism was generalized and made more flexible by replacing the restricted partitions by discrete paths, thus leading to the so called discrete path approach[2] which is a topological, graph theoretic[6] method, requiring a certain amount of combinatorial analysis.[7] Recently, the discrete path formalism was used to derive an explicit topological solution for systems of simultaneous linear equations[3] by establishing a homomorphism between the solution $x_n$, and the set of paths, defined on the corresponding signal flow graph,[8] from all sources to vertex $n$. The homomorphism defines a value function, and the solution is the value of the above set of paths.

The purpose of this paper is to apply the above topological solution to the case of ordinary, as well as partial, finite difference equations. A linear finite difference equation is a system of simultaneous linear equations characterized by a high degree of interdependency among the coefficients. This in turn is reflected by a high degree of symmetry in the corresponding signal glow graph, and permits us to greatly simplify the topological solution. This work constitutes a natural generalization of the solution obtained in Ref. 2, and is one more step in the process of enlarging the class of finite difference equations for which a general solution is known, and consequently enlarging the scope of physical problems that can be solved analytically.

Special cases, of the solution presented here, include solutions of ordinary linear finite difference equations with constant coefficients,[4,9] ordinary linear homogeneous finite difference equations with variable coefficients and initially,[1] finally,[10] or arbitrarily[2] specified boundary conditions, and ordinary linear inhomogeneous finite difference equations with variable coefficients and initially, or finally, specified boundary conditions.[11] The above developments[1,2,10] were motivated by, and needed for, the analytic solution of the Schrödinger equation[10,12] for a quark-antiquark system interacting via a central linear potential.[13] Partial finite difference equations will be needed in solving the Schrödinger equation for a three-quark system.[14-15]

The method presented here handels ordinary and partial finite difference equations in essentially the same way. From the point of view of simultaneous linear equations, the only difference between ordinary and partial finite difference equations is in the number of indices used to label the variables, and consequently the number of dimensions needed to display the vertices of the corresponding signal flow graph. For an ordinary finite difference equation, the signal flow graph can be represented by a linear lattice of vertices.[2] On the other hand, for a partial finite difference equation with two variables, a two dimensional lattice of vertices is needed, and so on. The method also treats homogeneous and inhomogeneous finite difference equations with the same ease; the inhomogeneous term combines with the boundary conditions to form the sources of the signal flow graph. Thus the only difference between the two cases is that in the inhomogeneous case every vertex has a source, while relatively few vertices have sources in the homogeneous case. Finally, the method permits an arbitrary specification of boundary conditions provided they are compatible with the equation.[2] A change in the specification of boundary conditions changes the position of the sources in the homogeneous case, and their relative strength in the inhomogeneous case.

In Sec. II we will present a summary of the needed results on simultaneous linear equations. In Sec. III we will study ordinary finite difference equations, and Sec. IV will deal with partial finite difference equations.

## II. SIMULTANEOUS LINEAR EQUATIONS

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

$$x_j = \sum_{\substack{i\in\Gamma^{-1}(j)\\ i\neq s_j}} w(i,j)x_i + w(s_j,j) \quad j = 1, 2, \ldots N \tag{2.1}$$

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

 0022-2488/80/102475-06$1.00  

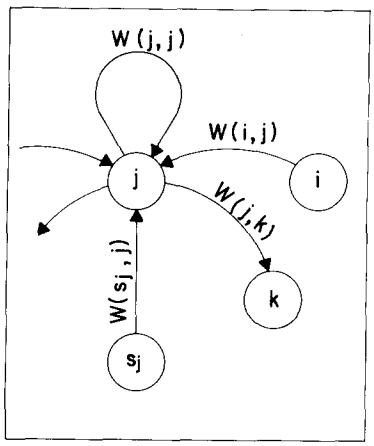



FIG. 1. A typical junction in the signal flow graph corresponding to Eq. (2.1).

and represented[8] by the signal flow graph $G$, as shown in Fig. 1, where $\Gamma^{-1}(j)$ is the set of vertices incident on vertix $j$, $w(i,j)$ is the weight of arc $(i,j)$, and $s_j$ is the source incident on vertex $j$. By definition, a vertex $s_j$ is called a source if there are no vertices incident on it.

As shown in Ref. 3 it is possible to obtain a topological solution of the system of Eq. (2.1) in the form

$$x_j = H[P(j)], \tag{2.2}$$

where $P(j)$ is the set of all paths, on the signal flow graph, from all sources to vertex $j$. $H$ is the homorphism introduced in Ref. 3, and gives the value of a path as the product of the weights of its arcs, and the value of a set of paths as the sum of the values of its elements,

$$H(p) = \prod_{(i,j)\in p} w(i,j), \tag{2.3a}$$

$$H(p) = \sum_{p\in P} H(p), \tag{2.3b}$$

$$H(\emptyset) = 0 \quad \text{and} \quad H(\{e\}) = 1, \tag{2.3c}$$

where $\emptyset$ is the empty set and $\{e\}$ is a set made up of one element of zero length.

Using the above properties of the homomorphism $H$ we can rewrite the solution as

$$x_j = \sum_{s\in S} H[P(s,j)], \tag{2.4}$$

where $S$ is the set of sources and $P(s,j)$ is the set of all paths from source $s$ to vertex $j$. Furthermore since we are considering that there is a maximum of one source per vertex, and that each source is incident on only one vertex, then $P(s_i,j) = \{(s_i,i)\} \otimes P(i,j)$, where $\otimes$ is the concatenation operation,[3] and consequently

$$x_j = \sum_{s_i\in S} w(s_i,i) H[P(i,j)]. \tag{2.5}$$

The fact that the solution is a sum of the contributions of the different sources, is a reflection of the linear character of the equations.

## III. ORDINARY FINITE DIFFERENCE EQUATIONS

### A. The Difference Equation

The ordinary linear inhomogeneous finite difference equation can be written as[2,4,16]

$$x_m = \sum_{k=1}^{N} f_{a_k}(m) x_{m-a_k} + I(m), \quad m \notin J \tag{3.1a}$$

and the boundary conditions given by,

$$x_{j_\alpha} = \lambda_{j_\alpha} \quad j_\alpha \in J, \tag{3.1b}$$

where the coefficients $f_{a_k}(m)$, and the inhomogeneous term $I(m)$ are arbitrary functions of the level $m$, and the numbers $a_k$ are positive or negative intergers belonging to the set A.

$$A = \{a_1, a_2, \ldots, a_k\} \tag{3.2a}$$

and assumed to be ordered according to

$$a_1 < a_2 < \cdots < a_N. \tag{3.2b}$$

The order $h$ of the equation is given by

$$h = a_N - a_1, \tag{3.3}$$

and the boundary conditions specify the values of the solution $x_m$ for $h$ different values of the index $m$. The values of $m$ for which the boundary conditions are specified, will be referred to as boundaries, and given by the elements of the set

$$J = \{j_1, j_2, \ldots, j_h\}. \tag{3.4}$$

Eq. (3.1a) should be written in a way which is compatible with the boundary conditions.[2]

### B. The Signal Flow Graph

Equations (3.1a) and (3.1b) form a set of simultaneous linear inhomogeneous equations defining $x_m$ for all integer values of $m$. To draw the corresponding signal flow graph[8] we make the following observations based on Eqs. (3.1a, b):

(i)Every vertex[17] has one and only one source incident on it. The intensity of the source is given by

$$w(s_m, m) = I(m) \quad m \notin J \tag{3.5a}$$

if the vertex is not a boundary, and by

$$w(s_{j_\alpha}, j_\alpha) = \lambda_\alpha \quad j_\alpha \in J \tag{3.5b}$$

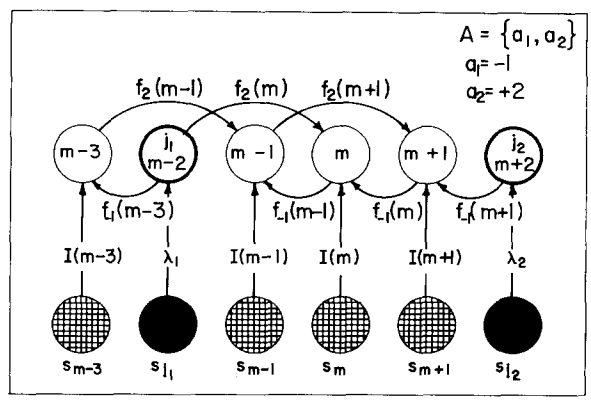



FIG. 2. A typical section of the signal flow graph corresponding to the ordinary finite difference equation $x_m = f_{-1}(m)x_{m-1} + f_{+2}(m)x_{m+2} + I(m)$, which is a special case of Eq. (3.1). Thin circles represent ordinary vertices, thick circles represent boundaries, and dark discs represent sources.

if the vertex is a boundary.

(ii)The set of vertices incident on a given vertex is given by

$$\Gamma^{-1}(m) = \{s_m\} \cup \{m - a_k; a_k \in A\} \quad m \notin J \tag{3.6a}$$

if the vertex is not a boundary, and by

$$\Gamma^{-1}(j_\alpha) = \{s_{j_\alpha}\} \quad j_\alpha \in J \tag{3.6b}$$

if the vertex is a boundary.

(iii)The weight of the arc $(m - a, m)$ eminating from vertex $(m - a)$ incident on vertex $m$, is given by

$$w(m - a, m) = \begin{cases} f_{a_k}(m) & \text{for } a = a_k \in A \\ 0 & \text{for } a \notin A. \end{cases} \tag{3.7}$$

Thus the signal flow graph corresponding to Eqs. (3.1) can be represented by two parallel linear lattices as shown in Fig. 2. The first lattice contains the ordinary vertices and the boundaries, while the second lattice contains the sources. The sources have a strength $\lambda_\alpha$ when incident on a boundary, and $I(m)$ when incident on an ordinary vertex. The vertices are divided into two types; ordinary vertices and boundaries. The boundaries have no vertices, with the exception of a source, incident on them.

### C. The Solution

Applying Eq. (2.5) to the signal flow graph of Fig. 2, and making use of Eqs. (3.5), we obtain the solution of the system of equations (3.1) in the form,

$$x_m = \sum_{j_\alpha \in J} \lambda_\alpha H[P(j_\alpha, m)] + \sum_{\substack{i \notin J \\ i \notin S}} I(i) H[P(i, m)], \tag{3.8}$$

where $J$ is the set of boundaries and $S$ is the set of sources. The value of the set of paths from $l$ to $m$ is given, using Eqs. (2.3a, b) and (3.7), by

$$H[P(l, m)] = \sum_{p \in P(l, m)} \prod_{(\beta, \gamma) \in p} f_{\gamma - \beta}(\gamma) \tag{3.9}$$

for all values of $l$ and $m$, including $l \in J$, provided that $P(l, m) \neq \emptyset$ and $P(l, m) \neq \{e\}$, in which case the value of $P(l, m)$ is given by Eq. (2.3c).

In addition to giving the solution for $x_m$ for all values of $m \notin J$, Eq. (3.8) correctly reproduced the boundary conditions. Let $j_\beta$ be a boundary vertex, then from Eq. (3.8) we have,

$$x_{j_\beta} = \sum_{j_\alpha \in J} \lambda_\alpha H[P(j_\alpha, j_\beta)] + \sum_{\substack{i \notin J \\ i \notin S}} I(i) H[P(i, m)]. \tag{3.10}$$

But since the only arc incident on a boundary is the one eminating from its corresponding source, then there are no paths joining a vertex, or a boundary, to another boundary. That is

$$P(i, j_\beta) = \emptyset \quad \text{for } i \notin J \text{ and } i \notin S \tag{3.11a}$$

and

$$P(j_\alpha, j_\beta) = \begin{cases} \emptyset & \text{for } \alpha \neq \beta \\ \{e\} & \text{for } \alpha = \beta, \end{cases} \tag{3.11b}$$

where $e$ is an arc of zero length joining $j_\beta$ to itself. From Eq. (2.3c) we then have

$$H[P(i, j_\beta)] = 0 \text{ and } H[P(j_\alpha, j_\beta)] = \delta_{\alpha\beta} \tag{3.11c}$$

and Eq. (3.10) reduces to

$$x_{j_\beta} = \sum_{\alpha = 1}^{h} \lambda_\alpha \delta_{\alpha\beta} = \lambda_\beta. \tag{3.12}$$

The value of the set of paths from $l$ to $m$ is what has previously been referred to as a combinatorics function,[2]

$$C(l, m) = H[P(l, m)]. \tag{3.13}$$

Thus in terms of combinatorics functions, Eq. (3.8) can be written as

$$x_m = \sum_{\alpha = 1}^{h} \lambda_\alpha C(j_\alpha, m) + \sum_{\substack{i \notin J \\ i \notin S}} I(i) C(i, m). \tag{3.14}$$

### D. The Boundary Conditions

Equation (3.10) corresponds to the case where the boundary conditions are arbitrarily specified at $h$ different values of $x_m$. On the other hand when the boundary conditions are given as initial conditions

$$x_{j_0 - i} = \lambda_{j_0 - i}, \quad i = 0, 1, 2, \ldots, h - 1 \tag{3.15a}$$

then, as shown in Ref. 2, all the elements of the set A must be positive,[16] and consequently

$$P(i, m) = \emptyset \quad \text{for } i > m.$$

Furthermore in this case $a_N = h$, and hence no segment can by pass the $h$ successive boundaries. That is, there are no paths joining points above and below the boundaries

$$P(i, m) = \emptyset \quad \text{for } i \leqslant j_0 - h \quad \text{and} \quad m > j_0.$$

Hence in this special case Eq. (3.14) reduces to

$$x_m = \sum_{i = 0}^{h - 1} \lambda_{j_0 - i} C(j_0 - i, m) + \sum_{i = j_0 + 1}^{m} I(i) C(i, m) \quad \text{for } m > j_0. \tag{3.15b}$$

Note that, in this case and due to Eq. (2.3c), $C(m, m) = H[P(m, m)] = H(\{e\}) = 1$. Similarly if the boundary conditions are given as final conditions

$$x_{j_0 + i} = \lambda_{j_0 + i}, \quad i = 0, 1, 2, \ldots, h - 1, \tag{3.16a}$$

then the elements of the set A must all be negative, and the boundary conditions decouple vertices on both sides of them leading to the solution

$$x_m = \sum_{i = 0}^{h - 1} \lambda_{j_0 + i} C(j_0 + i, m) + \sum_{i = m}^{j_0 - 1} I(i) C(i, m) \quad \text{for } m < j_0. \tag{3.16b}$$

## IV. PARTIAL FINITE DIFFERENCE EQUATIONS

### A. The Difference Equation

The partial linear inhomogenous finite difference equation of a function of $n$ variables can be written in the form,

$$x(\mathbf{m}) = \sum_{\mathbf{a}_k \in A} f_{\mathbf{a}_k}(\mathbf{m}) x(\mathbf{m} - \mathbf{a}_k) + I(\mathbf{m}) \quad \mathbf{m} \notin J \tag{4.1a}$$

and the boundary conditions specified by

$$x(\mathbf{j}_\alpha) = \lambda(\mathbf{j}_\alpha) \quad \mathbf{j}_\alpha \in J, \tag{4.1b}$$

where

$$\mathbf{m} = (m_1, m_2, ..., m_n), \tag{4.2a}$$

$$\mathbf{a}_k = (a_{k1}, a_{k2}, ..., a_{kn}), \tag{4.2b}$$

$$A = \{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N\}, \tag{4.3a}$$

$$J = \{\mathbf{j}_1, \mathbf{j}_2, ..., ...\}, \tag{4.3b}$$

$$x(\mathbf{m}) \equiv x(m_1, m_2, ..., m_n) \equiv x_{m_1 m_2 ... m_n}, \tag{4.4a}$$

$$I(\mathbf{m}) \equiv I(m_1, m_2, ..., m_n), \tag{4.4b}$$

and

$$f_{\mathbf{a}_k}(\mathbf{m}) \quad f_{a_{k1} a_{k2} ... a_{kn}}(m_1, m_2, ..., m_n). \tag{4.5}$$

The order of the equation is given by the vector $\mathbf{h}$,

$$\mathbf{h} = (h_1, h_2, ..., h_n), \tag{4.6a}$$

where

$$h_i = a_i^{\max} - a_i^{\min} \tag{4.6b}$$

and

$$a_i^{\max} = \max\{a_{1i}, a_{2i}, ..., a_{Ni}\} \tag{4.6c}$$

$$a_i^{\min} = \min\{a_{1i}, a_{2i}, ..., a_{Ni}\}. \tag{4.6d}$$

The compatibility of the boundary conditions with Eq. (4.1a), and their relation to the order $\mathbf{h}$, will be discussed in Sec. 4.

### B. The Signal Flow Graph

Equations (4.1a) and (4.1b) form a set of simultaneous linear inhomogeneous equations defining $x(\mathbf{m})$ for all integer values of $(m_1, m_2, ..., m_n)$. The corresponding signal flow graph consists of two regular $n$-dimensional lattices that are displaced, one with respect to the other, as shown in Fig. 3. The first lattice contains the ordinary vertices and boundaries, while the second lattice contains the sources. For every $n$-dimensional vector $\mathbf{m}$ with integer components there is a vertex on the first lattice and a corresponding source on the second lattice. It is convenient to introduce the following notation for the vertices of the signal flow graph:

$V = \{\mathbf{V}_\alpha, \mathbf{V}_\beta, ...\}$ is the set of all vertices,

$S = \{\mathbf{s}_\alpha, \mathbf{s}_\beta, ...\}$ is the set of all sources,

$J = \{\mathbf{j}_\alpha, \mathbf{j}_\beta, ...\}$ is the set of all boundaries,

$M = \{\mathbf{m}_\alpha, \mathbf{m}_\beta, ...\}$ is the set of ordinary vertices.

Thus

$$V = M \cup J \cup S. \tag{4.7}$$

The sources have a strength of $I(\mathbf{m})$ when incident on an ordinary vertex $\mathbf{m}$, as a strength of $\lambda(\mathbf{j}_\alpha)$ when incident on a boundary vertex $\mathbf{j}_\alpha$. That is

$$w[\mathbf{s}(\mathbf{m}), \mathbf{m}] = I(\mathbf{m}), \quad \mathbf{m} \in M \tag{4.8a}$$

$$w[\mathbf{s}(\mathbf{j}_\alpha), \mathbf{j}_\alpha] = \lambda(\mathbf{j}_\alpha), \quad \mathbf{j}_\alpha \in J. \tag{4.8b}$$

The set of vertices incident on a vertex $\mathbf{m}$ is given by

$$\Gamma^{-1}(\mathbf{m}) = \{\mathbf{s}(\mathbf{m})\} \cup \{\mathbf{m} - \mathbf{a}_k; \mathbf{a}_k \in A\}, \quad \mathbf{m} \in M \tag{4.9a}$$

and that incident on a boundary is given by

$$\Gamma^{-1}(\mathbf{j}_\alpha) = \{\mathbf{s}(\mathbf{j}_\alpha)\}, \quad \mathbf{j}_\alpha \in J. \tag{4.9b}$$

Finally, the weight of an arc incident on an ordinary vertex $\mathbf{m} \in M$ is given by,

$$w(\mathbf{m} - \mathbf{a}, \mathbf{m}) = \begin{Bmatrix} f_{\mathbf{a}_k}(\mathbf{m}) & \text{for } \mathbf{a} = \mathbf{a}_k \in A \\ 0 & \text{for } \mathbf{a} \notin A \end{Bmatrix}. \tag{4.10}$$

### C. The Solution

Applying Eq. (2.5) to the signal flow graph corresponding to Eqs. (4.1a, b) and making use of Eqs. (4.8) we obtain the solution of Eqs. (4.1) in the form

$$x(\mathbf{m}) = \sum_{\mathbf{j}_\alpha \in J} \lambda(\mathbf{j}_\alpha) H[P(\mathbf{j}_\alpha, \mathbf{m})] + \sum_{\mathbf{i} \in M} I(\mathbf{i}) H[P(\mathbf{i}, \mathbf{m})], \tag{4.11}$$

where $P(\mathbf{l}, \mathbf{m})$ is the set of all paths, on the signal flow graph, which join vertex $\mathbf{l}$ to vertex $\mathbf{m}$. Its value is given by

$$H[P(\mathbf{l}, \mathbf{m})] = \sum_{p \in P(\mathbf{l}, \mathbf{m})} \prod_{(\beta, \gamma) \in p} f_{\gamma - \beta}(\gamma) \tag{4.12}$$

for all values of $\mathbf{l}$ and $\mathbf{m}$ on the first lattice, provided that $P(\mathbf{l}, \mathbf{m}) \neq \emptyset$ and $P(\mathbf{l}, \mathbf{m}) \neq \{e\}$, in which case the value of $P(\mathbf{l}, \mathbf{m})$ is given Eq. (2.3c).

As in the case of ordinary finite difference equations, the solution (4.11) correctly reproduces the boundary conditions. Let $\mathbf{j}_\beta$ be a boundary vertex, then

$$x(\mathbf{j}_\beta) = \sum_{\mathbf{j}_\alpha \in J} \lambda(\mathbf{j}_\alpha) H[P(\mathbf{j}_\alpha, \mathbf{j}_\beta)] + \sum_{\mathbf{i} \in M} I(\mathbf{i}) H[P(\mathbf{i}, \mathbf{j}_\beta)]. \tag{4.13}$$

Due to Eq. (4.9b) we have,

$$P(\mathbf{i}, \mathbf{j}_\beta) = \emptyset \quad \text{for } \mathbf{i} \in M \tag{4.14a}$$

and

$$P(\mathbf{j}_\alpha, \mathbf{j}_\beta) = \begin{cases} \emptyset & \text{for } \alpha \neq \beta \\ \{e\} & \text{for } \alpha = \beta \end{cases}. \tag{4.14b}$$

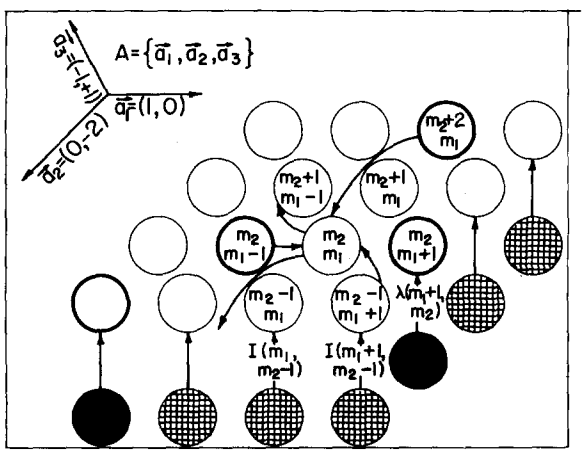



FIG. 3. A typical section of the signal flow graph corresponding to the partial finite difference equation

$$x(m_1, m_2) = f_{1,0}(m_1, m_2) x(m_1 - 1, m_2) + f_{0,-2}(m_1, m_2) \times x(m_1, m_2 + 2) + f_{-1,+1}(m_1, m_2) x(m_1 + 1, m_2 - 1) + I(m_1, m_2).$$

Only arcs into, and out of, vertex $(m_1, m_2)$ have been indicated. Thin circles, thick circles and darkend discs represent, ordinary vertices, boundaries, and sources respectively.

Consequently

$$H[P(\mathbf{i}, \mathbf{j}_\beta)] = 0 \quad \text{and} \quad H[P(\mathbf{j}_\alpha, \mathbf{j}_\beta)] = \delta_{\alpha\beta}. \tag{4.15}$$

Substituting the above values in (4.13) we obtain

$$X(\mathbf{j}_\beta) = \lambda(\mathbf{j}_\beta). \tag{4.16}$$

The combinatorics function over an $n$-dimensional lattice is defined by

$$C(\mathbf{l}, \mathbf{m}) = H[P(\mathbf{l}, \mathbf{m})] \tag{4.17}$$

and Eq. (4.13) can be rewritten as

$$x(\mathbf{m}) = \sum_{\mathbf{j}_\alpha \in J} \lambda(\mathbf{j}_\alpha) C(\mathbf{j}_\alpha, \mathbf{m}) + \sum_{\mathbf{i} \in M} I(\mathbf{i}) C(\mathbf{i}, \mathbf{m}). \tag{4.18}$$

### D. The Boundary Conditions

The boundary conditions, for a partial finite difference equation of a function of $n$ variables, are traditionally stated in the form of $h$ equations

$$x_{m_1 m_2 \ldots m_i \ldots m_n} = \lambda_{im_i}(m_1, m_2, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n) i = 1, 2, \ldots, n$$

and

$$m_i = 0, 1, \ldots, h_i - 1, \tag{4.19}$$

where $\mathbf{h} = (h_1, h_2, \ldots, h_n)$ is the order vector of the equation, and

$$h = \sum_{i=1}^{n} h_i. \tag{4.20}$$

For an ordinary finite difference equation, the value of $n$ is 1 and hence the $\lambda$ are constants.

The above specification has several shortcomings: First it is restrictive, giving the boundary conditions as initial conditions. Second there is overlap in the specification; that is, certain values of $x_{\mathbf{m}}$ are specified by several different functions, and it is important to insure that the different specifications give the same value for $x_{\mathbf{m}}$. Finally, when some levels are missing in the difference equation, the above set of equations overspecifies the boundary conditions leading either to redundancy or inconsistancy.

The first shortcoming can be overcome by replacing the range of $m_i$ in Eq. (4.19) by

$$m_i \in J_i = \{j_{i1}, j_{i2}, \ldots, j_{ih_i}\}. \tag{4.21}$$

This provides a certain liberty in specifying the boundary conditions, but still requires that they be specified along straight lines of vertices parallel to the axes of the signal flow graph lattice. On the other hand the solution (4.11) is valid for an arbitrary specification of boundary conditions. Thus we specify the boundary conditions by the values of $x_{\mathbf{m}}$ over a certain domain $J$ of values of $\mathbf{m}$ as in Eq. (4.1b). This, at the same time, avoids the problem of over specification and incoherence.

We still need to formulate the problem of compatibility[2] of the equation with the boundary conditions. To this end we use $\Gamma^{-n}(\mathbf{m})$, the set of the $n^{\text{th}}$ order antecedents (ancestors) of vertex $\mathbf{m}$. The first-order antecedents of vertex $\mathbf{m}$ are the vertices incident on $\mathbf{m}$, and are given by the elements of the set $\Gamma^{-1}(\mathbf{m})$. The second-order antecedents of $\mathbf{m}$ are the vertices incident on the element of $\Gamma^{-1}(\mathbf{m})$, and so on. Thus

$$\Gamma^{-n}(\mathbf{m}) = \bigcup_{\mathbf{k} \in \Gamma^{-(n-1)}(\mathbf{m})} \Gamma^{-1}(\mathbf{k}). \tag{4.22}$$

Since a source $s$ has no vertices incident on it then

$$\Gamma^{-n}(s) = \emptyset \quad n = 1, 2, \ldots. \tag{4.23}$$

Consider a vertex $\mathbf{m}$ on the signal flow graph, and let us trace backwards the paths incident on it. After $n$ steps backwards we reach the $n^{\text{th}}$ order antecedents of $\mathbf{m}$. The elements of $\Gamma^{-n}(\mathbf{m})$ that are sources will not generate any elements in $\Gamma^{-(n+1)}(\mathbf{m})$, while elements that are not sources will. Compatibility requires that all paths incident on $\mathbf{m}'$, and made up of vectors $\mathbf{a}_k \in A$, should eminate from sources. Hence the condition of compatibility can be stated as,

$$\Gamma^{-\infty}(\mathbf{m}) = \emptyset \quad m \in M. \tag{4.24}$$

## V. CONCLUSION

The solution of partial finite difference equations on $n$ variables has been obtained as the value of a set of paths on an $n$-dimensional lattice. The paths are constructed from a set of vectors determined by the level differences of the equation, and the value of an arc is the corresponding coefficient. The value of a path is the product of the values of its arcs, and the value of a set of paths is the sum of the values of its elements.

The solution is in principle as easy to obtain for homogeneous as for inhomogeneous equations, for ordinary as for partial difference equations, and for initial as for arbitrary boundary conditions. The method is general, intuitive, and simple to apply. The solution and its derivation are fully topological and totally independent of Cramer's rule.

By expanding the solution of a partial differential equation in a multivariable power series, the differential equation can be transformed into a partial difference equation for the expansion coefficients. Thus the solution presented here will provide power series solutions of partial differential equations.


[1] A. F. Antippa and A. J. Phares, J. Math. Physics **18**, 173 (1977).
[2] A. F. Antippa, J. Math. Physics **18**, 2214 (1977).
[3] A. F. Antippa and N. K. Toan, J. Math. Physics, **20**, 2375 (1979).
[4] L. M. Milne-Thomson, *The Calculus of Finite Differences* (MacMillan, London, 1951).
[5] G. E. Andrews, *The Theory of Partitions* (Addison-Wesley, Reading, Mass. 1976).
[6] Frank Harary, *Graph Theory and Theoretical Physics* (Academic, New York, 1967).
[7] C. Berge, *Principles of Combinatorics* (Academic New York, 1971); N. Ya. Vilenkin, *Combinatorics* (Academic, New York, 1971).
[8] W. K. Chen, *Applied Graph Theory* (North-Holland, Amsterdam, 1971).
[9] F. B. Hildebrand, *Methods of Applied Mathematics* (Prentice-Hall, Englewood Cliffs, N. J., 1952).
[10] A. F. Antippa and A. J. Phares, J. Math. Physics **19**, 308 (1978).
[11] A. J. Phares, J. Math. Physics **18**, 1838 (1977).
[12] A. F. Antippa and N. K. Toan, Can. J. Physics **57**, 417 (1979).
[13] B. Harrington, S. Park, and A. Yildiz, Phys, Rev. Letters, **34**, 168 (1975); *ibid.* **34**, 704 (1975).

[14]E. Witten, Nucl. Phys. B **160**, 57 (1979).
[15]J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).
[16]In Ref. 2 paths started at a vertex $m$ and ended at a "terminal point" $j$. In this work we use the opposite convention, in which the direction of a path is from the boundary $j$ to the vertex $m$. This leads to a reversal of the sign of the segments $a_k \in \mathcal{A}$, and is in conformity with the convention used in graph theory.
[17]By vertex we usually mean an ordinary vertex which is not a source. On the other hand, vertices which are sources will always be clearly referred to as such.

# Finite nonabelian subgroups of SU($n$) with analytic expressions for the irreducible representations and the Clebsch–Gordan coefficients [a)]

U. Abresch, A. Bovier, O. Lechtenfeld, M. Lüling, V. Rittenberg, and G. Weymans
*Physikalisches Institut, Bonn University, Bonn, West Germany*

(Received 27 December 1979; accepted for publication 13 June 1980)

We present two sequences of finite nonabelian groups which are semidirect products of three $Z_n$ groups. Although these groups are not simply reducible (in the tensor product of two irreducible representations an irreducible representation is obtained more than once) we give analytic expressions for the irreducible representations and the Clebsch–Gordan coefficients.

## I. INTRODUCTION

As opposed to continuous Lie groups, finite nonabelian groups have found up to now only a few applications in particle physics. As a result the best known groups in the trade are either the point groups [finite subgroups of SO(3)] or the permutation groups.[1] In the last few years, however, finite groups have found two new fields of applications: (a) when combined with local gauge invariance they are used as a symmetry of the flavor sector of unified theories of weak and electromagnetic interactions,[2] (b) in statistical mechanics they are considered in generalized Ising models.[3,4]

One would thus like to have in our hand a larger number of groups and try to use them in physical applications. Unfortunately a classification theory for finite groups does not exist and one has to find supplementary criteria in order to know how to look for them. In this paper we define two sequences of non-Abelian groups of order $n^3$ ($n = 2,3,...$) that we call $\mathscr{D}$ ($n$) and Q($n$) [$\mathscr{D}$ (2) is isomorphic to the dihedral group $D_4$ and Q (2) is isomorphic to the quaternionic group], and which have some remarkable properties. The Q ($n$) groups are subgroups of SU ($n$)[5] and this might make them suitable to be used as an approximation for an SU ($n$) theory on a lattice. On the other hand, all the characteristics of the $\mathscr{D}$ ($n$) and Q ($n$) groups can be written in an analytic way. This includes the irreducible representations (IR) and the Clebsch–Gordon (CG) coefficients as is the case of SU (2). In this way all calculations are simplified enormously and one does not have to deal, as is usually the case, with multiplication tables which define the group and other tables for the CG coefficients.

As a first application[6] spin systems defined on the groups $\mathscr{D}$ ($n$) have been considered and the selfduality conditions have been worked out. This was possible since analytic expressions for the characters of the I.R. of $\mathscr{D}$ ($n$) were available.

From a mathematical point of view we find one result very interesting: When taking tensor products of two irreducible representations one obtains the same representation several times (our groups are not simply reducible). In spite of that, one is able to define in an analytical way an orthonormal basis for the "degenerate" states and in this way to determine the Clebsch–Gordan coefficients. We remind the reader that this problem made it impossible, up to now, to find analytic expressions for the C-G coefficients of SU (3).

As a side point we define the $q$ ($n$)-algebras. Those are related to the Q ($n$) groups in the same way the quaternionic algebra $q$ (2) is related to the quaternionic group Q (2). The $q$ ($n$) algebras are relevant for the formulation of the $Z_n$ Potts models in statistical mechanics.[7]

This paper is organized as follows. In Sec. 2 the nonabelian groups of order $n^3\mathscr{D}$ ($n$) and Q ($n$) are defined. One shows which of the groups $\mathscr{D}$ ($n$) and Q ($n$) are shown to be solvable and to factorize in their Sylow subgroups.

In Sec. 3 we give the analytic expressions for the irreducible representations of the groups $\mathscr{D}$ ($n$) and Q ($n$). The characters of the groups $\mathscr{D}$ ($n$) and Q ($n$) (same $n$) are the same. The Clebsch–Gordan series are given in Sec. 4 and the Clebsch–Gordan coefficients in Sec. 5. Our conclusions and the definition of the $q$ ($n$) algebras are presented in Sec. 6.

## 2. THE $\mathscr{D}$ ($n$) AND Q ($n$) GROUPS

The elements of the group $\mathscr{D}$ ($n$) are labelled by $g^{\gamma}_{\alpha,\beta}$, where $\alpha,\beta,\gamma = 0,1,...,n-1$. Thus the group is of order $n^3$ and the multiplication rule is

$$g^{\gamma}_{\alpha,\beta}g^{\gamma'}_{\alpha',\beta'} = g^{\gamma+\gamma'+\alpha\beta'}_{\alpha+\alpha',\beta+\beta'}, \tag{2.1}$$

where $\alpha,\beta,\gamma \in Z_n$ and all operations in Eq. (2.1) and hereafter are made modulo $n$. From the definition (2.1) one sees that $\mathscr{D}$ ($n$) has an Abelian $Z_n \otimes Z_n$ subgroup given by the elements $g^{\gamma}_{\alpha,0}$. One can easily see that the $\mathscr{D}$ ($n$) groups are *solvable*. (A group $G$ is solvable if the sequence $G^{(i)}$ of the derived groups $G = G^{(0)} \supseteq G^{(1)} \cdots \supseteq G^{(k)}$ ends after a finite number of steps: $G^{(k)} = e$ where $e$ is the unit element. The group $G^{(i+1)}$ is obtained from the commutators of the elements of $G^{(i)}$.) In order to show this we compute the commutator of two elements $g^{\gamma}_{\alpha,\beta}$ and $g^{\gamma'}_{\alpha',\beta'}$, use Eq. (2.1):

$$\begin{aligned}(g^{\gamma}_{\alpha,\beta})^{-1}(g^{\gamma'}_{\alpha',\beta'})^{-1}g^{\gamma}_{\alpha,\beta}g^{\gamma'}_{\alpha',\beta'} &= g^{-\gamma+\alpha\beta}_{-\alpha,-\beta}g^{-\gamma'+\alpha'\beta'}_{-\alpha',-\beta'}g^{\gamma}_{\alpha,\beta}g^{\gamma'}_{\alpha',\beta'} \\ &= g^{\alpha\beta'-\alpha'\beta}_{0,0},\end{aligned} \tag{2.2}$$

and keep in mind that the subgroup $g^{\gamma}_{0,0}$ is abelian.

One can also show that the $\mathscr{D}$ ($n$) groups are nilpotent.[8] (A group $G$ is nilpotent if in the sequence of the derived groups $G = G^{(0)} \supseteq G^{(1)} \supseteq \cdots \supseteq G^{(k)} = e$, $G^{(i-1)}/G^{(i)}$ is in the center of $G/G^{(i)}$ for $i = 1,\cdots,k$.).

[a)] This work came out of a seminar for undergraduate students on finite groups and their applications.

 0022-2488/80/102481-06$1.00  

As an example let us consider the $\mathscr{D}$ (2) group and its faithful two-dimensional representation:

$$g^{\gamma}_{0,0} = (-1)^{\gamma}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (-1)^{\gamma}\mathbf{1};$$
$$g^{\gamma}_{1,0} = (-1)^{\gamma}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = (-1)^{\gamma}\sigma_3 \tag{2.3}$$
$$g^{\gamma}_{0,1} = (-1)^{\gamma}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (-1)^{\gamma}\sigma_1;$$
$$g^{\gamma}_{1,1} = (-1)^{\gamma}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -(-1)^{\gamma}i\sigma_2,$$

where $\gamma = 0,1$ and $\sigma_1$, $\sigma_2$, $\sigma_3$ are the Pauli matrices. From Eq. (2.3) one observes that the group $\mathscr{D}$ (2) coincides with the dihedral group $D_4$ which is a subgroup of SO (3).[1]

We now define the Q ($n$) groups through the multiplication rule

$$g^{\gamma}_{\alpha,\beta}g^{\gamma'}_{\alpha',\beta'} = g^{\gamma+\gamma'+\alpha\beta'+\alpha\alpha'+\beta\beta'}_{\alpha+\alpha',\beta+\beta'}, \tag{2.4}$$

where again $\alpha,\beta,\gamma \in Z_n$ and all operations are made modulo $n$. One can show that the Q ($n$) groups are solvable and nilpotent.

We consider two examples. First the group Q (2) and its faithful 2-dimensional representation:

$$g^{\gamma}_{0,0} = (-1)^{\gamma}\mathbf{1}; \quad g^{\gamma}_{1,0} = (-1)^{\gamma}i\sigma_3, \qquad g^{\gamma}_{0,1} = (-1)^{\gamma}i\sigma_1; \quad g^{\gamma}_{1,1} = (-1)^{\gamma}i\sigma_2, \tag{2.5}$$

where $\gamma = 0,1$ and $\sigma_1$, $\sigma_2$, $\sigma_3$ are the Puali matrices. This group coincides with the quaternionic group which is a subgroup of SU (2) [the matrices (2.5) are unitary and their determinant is one].

The Q (3) group has two faithful three-dimensional representations, we consider one of them

$$g^{\gamma}_{0,0} = \omega^{\gamma}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad g^{\gamma}_{1,0} = \omega^{\gamma}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix} \tag{2.6}$$
$$g^{\gamma}_{0,1} = \omega^{\gamma}\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \ldots,$$

where $\gamma = 0,1,2$ and $\omega = e^{2\pi i/3}$. These matrices are unitary and their determinant is one thus Q (3) is a finite subgroup of SU (3). The second 3-dimensional faithful representation of Q (3) is the complex conjugate representation of (2.6).

For finite groups an essential role is played by the order of the group. We remind the reader that if a group $G$ of order N has a subgroup H of order M then M is a divisor of N. The converse is not true (if M is a divisor of N one does not always have a subgroup of order M). If we write however

$$\mathsf{N} = p_1^{s_1}\cdot p_2^{s_2}\cdot \cdots \cdot p_l^{s_l}, \tag{2.7}$$

where $p_1,p_2,\ldots,p_l$ are prime numbers, then the group $G$ has the subgroups $\mathsf{H}_1,\mathsf{H}_2,\cdots,\mathsf{H}_l$ of order $p_1^{s_1}, p_2^{s_2}, \ldots, p_l^{s_l}$. The subgroups $\mathsf{H}_1,\mathsf{H}_2,\cdots,\mathsf{H}_l$ are called the *Sylow subgroups*.[8]

Nilpotent groups have the important property that they *factorize* into their Sylow subgroups[8] (the same property holds of course for Abelian groups). Thus if we write

$$n = p_1^{s_1}\cdot p_2^{s_2}\cdot \cdots \cdot p_l^{s_l}, \tag{2.7'}$$

we have

$$\mathscr{D}(n) = \mathscr{D}(p_1^{s_1}) \otimes \mathscr{D}(p_2^{s_2}) \otimes \cdots \otimes \mathscr{D}(p_l^{s_l}),$$
$$\mathrm{Q}(n) = \mathrm{Q}(p_1^{s_1}) \otimes \mathrm{Q}(p_2^{s_2}) \otimes \cdots \otimes \mathrm{Q}(p_l^{s_l}). \tag{2.8}$$

As an illustration we derive the factorization property for the $\mathscr{D}$ ($n$) groups. Let us assume that

$$n = k\cdot l, \tag{2.9}$$

where

$$\Phi(k,l) = 1. \tag{2.10}$$

The symbol $\Phi(k,l)$ in (2.10) and hereafter denotes the largest common divisor of $k$ and $l$. Any group element $g^{\gamma}_{\alpha,\beta} \in \mathscr{D}(n)$ can be written as

$$g^{\gamma}_{\alpha,\beta} = g^{l^2\gamma_k + k^2\gamma_l}_{l\alpha_k + k\alpha_l, l\beta_k + k\beta_l}, \tag{2.11}$$

where $\alpha_k,\beta_k,\gamma_k \in Z_k, \alpha_l,\beta_l,\gamma_l \in Z_l$ and the operations in (2.11) are done in $Z_n$ (modulo $n$).

We now can write the group elements $g^{\gamma}_{\alpha,\beta}$ as the product of two *commuting* elements of $\mathscr{D}$ ($n$)

$$g^{\gamma}_{\alpha,\beta} = g^{l^2\gamma_k}_{l\alpha_k,l\beta_k} g^{k^2\gamma_l}_{k\alpha_l,k\beta_l} = g^{k^2\gamma_l}_{k\alpha_l,k\beta_l} g^{l^2\gamma_k}_{l\alpha_k,l\beta_k}. \tag{2.12}$$

Note that the elements $g^{l^2\gamma_k}_{l\alpha_k,l\beta_k}$ form a subgroup of $\mathscr{D}$ ($n$)

$$g^{l^2\gamma_k}_{l\alpha_k,l\beta_k} g^{l^2\gamma'_k}_{l\alpha'_k,l\beta'_k} = g^{l^2(\gamma_k+\gamma'_k+\alpha_k\beta'_k)}_{l(\alpha_k+\alpha'_k),l(\beta_k+\beta'_k)}. \tag{2.13}$$

The operations in Eq. (2.13) are done in $Z_n$ but Eq. (2.13) remains unchanged if we take $l = 1$ and do the operations in $Z_k$ thus

$$\mathscr{D}(n) = \mathscr{D}(k) \otimes \mathscr{D}(l).$$

Because of the factorization property (2.8) the study of the groups $\mathscr{D}$ ($n$) and Q ($n$) can be reduced to the case $n = p^s$, where $p$ is a prime number. One can show that the group $\mathscr{D}(p^s)$ and $\mathrm{Q}(p^s)$ are isomorphic for $p$ odd. In order to find their relation to other known groups, we give here the known definitions of the generalized dihedral and quaternion groups.[8]

*Generalized dihedral group* D($s$) is a group of order $2^s$ obtained from two generators $a$ and $b$ satisfying the relations

$$a^{2^{s-1}} = 1; \quad b^2 = 1; \quad ba = a^{-1}b \quad (s \geqslant 3). \tag{2.14}$$

*Generalized quaternion group* C($s$) is a group of order $2^s$ with the generators $a$ and $b$ satisfying the relations

$$a^{2^{s-1}} = 1; \quad b^2 = a^{2^{s-2}}; \quad ba = a^{-1}b \quad (s \geqslant 3). \tag{2.15}$$

We notice that $\mathrm{D}(3) \equiv \mathrm{D}_4$. One can show only the following isomorphisms

$$\mathscr{D}(2) = \mathrm{D}(3); \quad \mathrm{Q}(2) = \mathrm{C}(3). \tag{2.16}$$

Thus the $\mathscr{D}$ ($n$) and Q ($n$) groups generalize the D(3) and C(3) groups on a different line than (2.14) and (2.15).

Before concluding this section we would like to make a comment about the Lie groups which have the finite groups Q ($n$) and $\mathscr{D}$ ($n$) as subgroups.

From the explicit expression of the irreducible representation which will be given in the next section one can see

that the Q $(n)$ groups have an $n\times n$ irreducible unitary representation with determinant one, they are thus subgroups of SU $(n)$. With the exception of $\mathscr{D}(2)$, which is a subgroup of SO(3), we do not know which simple Lie groups have $\mathscr{D}(2^s)$ as subgroups.

## 3. THE IRREDUCIBLE REPRESENTATIONS OF THE GROUPS $\mathscr{D}(n)$ AND Q$(n)$

We have obtained the expression for the IR of some groups $\mathscr{D}$ $(n)$ and Q $(n)$ by direct calculation and then we have *guessed* the general result. It was then an easy task to check that the IR were unitary and that we have obtained all of them.

The irreducible representations are labelled by the triple $[i,j;k]$, where $k\in Z_n$ and

$$0\leqslant i,\quad j\leqslant \bar{k}-1\,, \tag{3.1}$$

where

$$\bar{k}=\Phi(k,n)\,, \tag{3.2}$$

[$\Phi(k,n)$ is the largest common divisor of $k$ and $n$]. The dimension of the IR $[i,j;k]$ is

$$\mathrm{Dim}([i,j;k])=n/\bar{k}\,. \tag{3.3}$$

The expression of the $[i,j;k]$ irreducible representation of the group $\mathscr{D}$ $(n)$ is:

$$T^{[i,j;k]}_{a,b}(g^{\gamma}_{\alpha,\beta})=\delta_{ka,k(b+\beta)}\omega^{k(\gamma+\alpha b)+i\alpha+j\beta}\,, \tag{3.4}$$

where $a$ and $b$ are the indices of the matrix

$$a,b=0,1,\cdots,(n/\bar{k})-1 \tag{3.5}$$

and $\omega=e^{2\pi i/n}$. In Eq. (3.4) all operations are done modulo $n$. Although we should be interested only in the case $n=p^s$ [see Eq. (2.8)] the expression (3.4) gives automatically the result for any $n$. For example for the group

$$\mathscr{D}(6)=\mathscr{D}(2)\otimes\mathscr{D}(3)\,, \tag{3.6}$$

we can label the irreducible representation either by $[i,j;k]$ $(i,j,k\in Z_6)$ or by the pair $([i_2,j_2;k_2],[i_3,j_3;k_3])$, where $i_2,j_2,k_2\in Z_2$ and $i_3,j_3,k_3\in Z_3$.

It is interesting to compare the expression of the irreducible representations $[i,j;k]$ of $\mathscr{D}$ $(n)$ [see Eq. (3.4)] with the expressions of the one-dimensional irreducible representations $(i,j,k)$ of the *Abelian* group $Z_n\otimes Z_n\otimes Z_n$ where the group elements $h^{\gamma}_{\alpha,\beta}$ verify the multiplication rule

$$h^{\gamma}_{\alpha,\beta}h^{\gamma'}_{\alpha',\beta'}=h^{\gamma+\gamma'}_{\alpha+\alpha',\beta+\beta'}\,, \tag{3.7}$$

and the irreducible representations are

$$T^{(i,j,k)}(h^{\gamma}_{\alpha,\beta})=\omega^{i\alpha+j\beta+k\gamma}\,, \tag{3.8}$$

where $i,j,k,\alpha,\beta,\gamma\in Z_n$. We now compute from Eq. (3.4) the expression of the character of the IR $[i,j;k]$:

$$\chi^{[i,j;k]}(g^{\gamma}_{\alpha,\beta})=\frac{n}{\bar{k}}\delta_{k\alpha,0}\delta_{k\beta,0}\omega^{i\alpha+j\beta+k\gamma}\,. \tag{3.9}$$

We now notice the identity which relates the expressions (3.9) and (3.8)

$$\chi^{[i,j;k]}(g^{\gamma}_{\alpha,\beta})=\frac{1}{n\bar{k}}\sum_{p,q=0}^{n-1}T^{(pk+i,qk+j,k)}(h^{\gamma}_{\alpha,\beta})\,. \tag{3.10}$$

Thus the characters of the IR of the non-Abelian group $\mathscr{D}$ $(n)$ can be expended in terms of the characters of the irreducible representations of the Abelian group $Z_n\otimes Z_n\otimes Z_n$. This identity plays a crucial role when one is looking for duality transformations for spin systems defined on the groups $\mathscr{D}$ $(n)$.[6]

We now give the expression of the IR for the groups Q $(n)$. From the factorization theorem given by the Eq. (2.8) and the isomorphism

$$\mathrm{Q}(p^s)=\mathscr{D}(p^s) \tag{3.11}$$

for $p$ prime, $p\neq 2$ we can limit ourselves to the case $n=2^s$ only. We have

$$T^{[i,j;k]}_{a,b}(g^{\gamma}_{\alpha,\beta})=\delta_{ka,k(b+\beta)}\omega^{k[\gamma+\alpha b-(\alpha^2+\beta^2)/2]+i\alpha+j\beta}\,. \tag{3.12}$$

[In Eq. (3.12) one first computes $k[\gamma+\alpha b-(\alpha^2+\beta^2)/2]$ in $Z$ and then computes the other operations in $Z_n$ (modulo $n$).]

For the case $n=p^s(p\neq 2)$ the representations (3.12) can still be used and they are equivalent to those given by Eq. (3.4). The characters for the IR of the Q $(2^s)$ group obtained from Eq. (3.12) are

$$\begin{aligned}\chi^{[i,j;k]}(g^{\gamma}_{\alpha,\beta})&=\frac{n}{\bar{k}}\delta_{k\alpha,0}\delta_{k\beta,0}\omega^{k[\gamma-(\alpha^2+\beta^2)/2]+i\alpha+j\beta}\\&=\frac{n}{\bar{k}}\delta_{k\alpha,0}\delta_{k\beta,0}\omega^{i\alpha+j\beta+k\gamma}\,.\end{aligned} \tag{3.13}$$

Thus the characters of the $\mathscr{D}(2^s)$ and Q$(2^s)$ groups coincide. This property could already been noticed from the representations [0,0;1] of the groups $\mathscr{D}(2)$ and Q(2) given by Eqs. (2.3) and (2.5).

Let us now consider a few examples. If $n$ is a prime number $p$ the structure of the IR $[i,j;k]$ simplifies very much. For $\mathscr{D}$ $(p)$ there are only two possibilities:

(1) $k=0$. In this case $\bar{k}=\Phi(0,p)=p$, and we have $p^2$ one-dimensional representations $[i,j;0]$ $(i,j\in Z_p)$ with $[i,j;k]^*=[p-i,p-j;0]$.

(2) $k\neq 0$. In this case $\bar{k}=\Phi(k,p)=1$ thus $i=j=0$ and all the representations $[0,0;k]$ $(k=1,2,\cdots,p-1)$ are $p$-dimensional. We also have $[0,0;k]^*=[0,0;p-k]$.

Thus for $\mathscr{D}(2)$ and Q(2) one has 4 one-dimensional representation and one two-dimensional real representation. For Q(3) one has 9 one-dimensional representations and two three-dimensional representations, etc.

## 4. CLEBSCH–GORDAN SERIES

Although the expression of the CG coefficients will be given in the next section, we prefer to present separately the Clebsch–Gordan series. The series are the same for the Q $(n)$ and $\mathscr{D}$ $(n)$ groups. Let us take the tensor product of two irreducible representations $[i_1,j_1;k_1]$ and $[i_2,j_2;k_2]$ of $\mathscr{D}$ $(n)$ and find into which irreducible representations $[i,j;k]$ it decomposes.

A few notations are useful. We denote

$$\begin{aligned}&\bar{k}_1=\Phi(k_1,n);\ \bar{k}_2=\Phi(k_2,n);\ \bar{k}=\Phi(k,n)\\&\kappa=\Phi(\bar{k}_1,\bar{k});\ \delta=\psi(\bar{k}_1,\bar{k}_2,\bar{k})\,,\end{aligned} \tag{4.1}$$

where $\Phi(a,b)$ is the largest common divisor of $a$ and $b$ and $\psi(a,b,c)$ is the least common multiplier of $a$, $b$, and $c$. [Notice that for $k = k_1 + k_2$, $\Phi(\bar{k}_1,\bar{k}) = \Phi(\bar{k}_2,\bar{k})$.]

The Clebsch–Gordan series read

$$[i_1,j_1;k_1][i_2,j_2;k_2]$$

$$= \bigoplus_{u,v,\eta} [i_1 + i_2 - \kappa u, j_1 + j_2 - \kappa v; k]_\eta , \tag{4.2}$$

where

$$\begin{aligned} &k = k_1 + k_2 \\ &0 \leqslant i_1 + i_2 - \kappa u \leqslant \bar{k} - 1 \\ &0 \leqslant j_1 + j_2 - \kappa v \leqslant \bar{k} - 1 , \\ &0 \leqslant \eta \leqslant n/\delta - 1 . \end{aligned} \tag{4.3}$$

The index $\eta$ was introduced in order to distinguish between the various equivalent representations (same $[i,j;k]$) obtained in the tensor product. The multiplicity of the representation $[i,j;k]$ in the product $[i_1,j_1;k_1] \otimes [i_2,j_2;k_2]$ is of course:

$$\mathrm{N}([i,j;k]) = n/\delta . \tag{4.4}$$

Let us give a few examples. First the case Q(2). We have

$$\begin{aligned} &[i_1,j_1;0] \otimes [i_2,j_2;0] = [i_1 + i_2, j_1 + j_2;0] , \\ &\quad (i_1,i_2,j_1,j_2 \in Z_2) ; \\ &[i_1,j_1;0] \otimes [0,0;1] = [0,0;1] , \\ &\quad (i_1,j_1 \in Z_2) , \\ &[0,0;1] \otimes [0,0;1] = \bigotimes_{u,v=0}^{1} [u,v;0] . \end{aligned} \tag{4.5}$$

Keeping in mind that the representations $[i,j;0]$ $(i,j \in Z_2)$ are one-dimensional and that the representation $[0,0;1]$ is the two-dimensional one, the multiplication rule (4.5) reads off easily from an SU(2) point of view with the interpretation that the spinor representation remains two-dimensional for the Q(2) subgroup but that the vector representation splits into three nonequivalent one-dimensional representations.

A similar situation occurs for Q(3) [subgroup of SU(3)]. The 3 and 3* representations of SU(3) remain irreducible under Q(3) (those are the representations $[0,0;1]$ and $[0,0;2]$). The representation 6 decomposes under Q(3) into two representations 3* and the octet representation decomposes into 8 one-dimensional representations (these are the representations $[i,j;0]$, $i,j \in Z_3$ with the exception of $[0,0;0]$, which is the scalar representation). We have, for example,

$$[0,0;1] \otimes [0,0;1] = [0,0;2]_0 \oplus [0,0;2]_1 \oplus [0,0;2]_2 , \tag{4.6}$$

which reads

$$3 \otimes 3 = 3^* \oplus 3^* \oplus 3^* . \tag{4.6'}$$

As seen above the $\mathscr{D}(n)$ and Q $(n)$ groups are not simply reducible. This makes the problem of the determination of the Clebsch–Gordan coefficients a very interesting one since, up to now, (to our knowledge) there is no example in which this problem has been solved in an analytic way.

## 5. THE CLEBSCH–GORDAN COEFFICIENTS

We first give the expression of the CG coefficients [they are identical for the $\mathscr{D}(n)$ and Q $(n)$ groups] and afterwards sketch the derivation. With the notation used in Sec. 4 we have

$$\begin{aligned} &C^{[i_1+i_2-\kappa u, j_1+j_2-\kappa v;k]_\eta, a}_{[i_1,j_1;k_1],\, a_1;[i_2,j_2;k_2],\, a_2} \\ &= (\kappa/k)^{1/2} \delta_{k_1+k_2,k} \delta_{k_1 a_1 + k_2 a_2 - k a, -\kappa u} \\ &\quad \times \delta_{\tau(a_1 - a - \xi),0} \omega^{u(l\bar{k}a + l_1 \bar{k}_1 a_1)} \end{aligned} \tag{5.1}$$

here $a_1$, $a_2$ and $a$ denote the states corresponding to the representation $[i_1,j_1;k_1]$, $[i_2,j_2;k_2]$, and $[i,j;k]$, where $(i = i_1 + i_2 - \kappa u, j = j_1 + j_2 - \kappa v)$. We have

$$\begin{aligned} &0 \leqslant a_1 \leqslant (n/\bar{k}_1) - 1; \quad 0 \leqslant a_2 \leqslant (n/\bar{k}_2) - 1; \\ &\quad 0 \leqslant a \leqslant (n/\bar{k}) - 1 , \end{aligned} \tag{5.2}$$

where

$$k = k_1 + k_2 . \tag{5.3}$$

The index $\eta$ distinguishes between equivalent representations. In the right-hand side of the Eq. (5.1) we have

$$\tau = \psi(\bar{k}_1,\bar{k}) , \tag{5.4}$$

where $\psi(a,b)$ is the least common multiplier of $a$ and $b$. The connection between $\xi$ and $\eta$ is a more complicated one:

$$\zeta = \xi_0 + \eta\psi(\bar{k}_1,\bar{k}_2)/\bar{k}_1; \quad 0 \leqslant \xi \leqslant (n/\tau) - 1 . \tag{5.5}$$

The value $\xi_0$ is the smallest solution of the equation

$$k_1\xi + k_2\zeta = i - i_1 - i_2 = \kappa u . \tag{5.6}$$

Here $\zeta \in Z_n$ is arbitrary. We remind the reader that all operations are made modulo $n$. Finally the positive integers $l$ and $l_1$ in Eq. (5.1) are defined by the equation

$$\Phi(\bar{k}_1,\bar{k}) = l_1\bar{k}_1 + l\bar{k} . \tag{5.7}$$

For $n = p^s$ ($p$ is a prime number), $l_1$ and $l$ are 0 or 1 (for $\bar{k}_1 = \bar{k}$, we take by convention $l_1 = l = \frac{1}{2}$.

Before going into the derivation of Eq. (5.1) let us consider again the example of the groups $\mathscr{D}(p)$ where $p$ is a prime number. In this case (see Secs. 3 and 4) there are four cases.

(1) $k_1 \neq 0$, $k_2 \neq 0$, $k = k_1 + k_2 \neq 0$. In this case

$$\bar{k}_1 = \bar{k}_2 = \bar{k} = \kappa = \delta = \tau = \psi(\bar{k}_1,\bar{k}_2) = 1 ,$$

$$i_1 = i_2 = i = j_1 = j_2 = j = \xi_0 = 0 ,$$

$$\xi = \eta$$

and we get

$$C^{[0,0;k]_\eta, a}_{[0,0;k_1],a_1;[0,0;k_2],a_2} = \delta_{k_1+k_2,k} \delta_{\eta,a_1-a} \delta_{k_1a_1+k_2a_2,ka} , \tag{5.8}$$

where

$$0 \leqslant a_1, a_2, a, \eta \leqslant p - 1 .$$

(2) $k_1 \neq 0$, $k_2 \neq 0$, $k = k_1 + k_2 = 0$. We have

$$\bar{k}_1 = \bar{k}_2 = \kappa = l_1 = 1; \quad \delta = \bar{k} = \tau = p; \quad \xi = \eta = l = 0$$

and hence

$$C^{[i,j;0]}_{[0,0;k_1],a_1;[0,0;k_2],a_2} = \frac{1}{\sqrt{p}} \delta_{k_1a_1+k_2a_2,i} \omega^{-ja_1} \tag{5.9}$$

(3) $k_1 \neq 0$, $k_2 = 0$, $k \neq 0$. Now we have

$$\bar{k}_1 = \bar{k} = \kappa = \tau = 1; \quad \bar{k}_2 = \delta = p; \quad \xi = \eta = 0$$

and

$$C^{[0,0;k_1],a}_{[0,0;k_1],a_1;[i_2,j_2,0]} = \delta_{k_1(a_1-a),-i_2} \cdot \tag{5.10}$$

(4) $k_1 = 0$, $k_2 = 0$, $k = 0$. In this case

$$\bar{k}_1 = \bar{k}_2 = \bar{k} = \delta = \kappa = \tau = p; \quad \xi = \eta = 0$$

$$C^{[i,j;0]}_{[i_1,j_1;0],[i_2,j_2;0]} = \delta_{i_1+i_2,i}\delta_{j_1+j_2,j} \cdot \tag{5.11}$$

We now give the derivation of Eq. (5.1) in the case of the $\mathscr{D}$ $(n)$ groups [as mentioned before the results are identical for the Q $(n)$ groups]. We first write the expression of the irreducible representation $[i,j;k]$ [see Eq. (3.4)] in an equivalent form

$$T^{[i,j;k]}(g^{\gamma}_{\alpha,\beta})e_{kb} = \omega^{k(\gamma+\alpha b)+i\alpha+j\beta}e_{k(b+\beta)} \tag{5.12}$$

here $e_0$, $e_{k.1}$,..., $e_{k(n/\bar{k}-1)}$ are the basis vectors.

We are interested in the tensor product of the representation $[i_1,j_1;k_1]$ and $[i_2,j_2;k_2]$ and we start in the standard way[1] by constructing the vectors

$$P_c(e_{k_1b_1}\otimes e_{k_2b_2}) = \frac{1}{n^2\bar{k}}\sum_{\alpha,\beta,\gamma}[T^{[i,j;k]}_{c,0}(g^{\gamma}_{\alpha,\beta})]^* \times T^{[i_1,j_1;k_1]}(g^{\gamma}_{\alpha,\beta})e_{k_1b_1}\otimes T^{[i_2,j_2;k_2]}(g^{\gamma}_{\alpha,\beta})e_{k_2,b_2} \cdot \tag{5.13}$$

Using Eqs. (3.4) and (5.12), we can write Eq. (5.13) in the following form:

$$P_c(e_{k_1b_1}\otimes e_{k_2b_2}) = \frac{1}{n^2\bar{k}}\sum_{\alpha,\beta,\gamma}\delta_{kc,k\beta} \times \omega^{k_1(\gamma+\alpha b_1)+k_2(\gamma+\alpha b_2)-k\gamma+\alpha(i_1+i_2-i)+\beta(j_1+j_2-j)} \times e_{k_1(b_1+\beta)}\otimes e_{k_2(b_2+\beta)} \cdot \tag{5.14}$$

We perform the summations over $\alpha$ and $\gamma$ and obtain

$$P_c(e_{k_1b_1}\otimes e_{k_2b_2}) = (1/\bar{k})\delta_{k_1+k_2,k}\delta_{k_1b_1+k_2b_2,i-i_1-i_2}F, \tag{5.15}$$

where

$$\begin{aligned}F &= \sum_{\beta}\delta_{kc,k\beta}\omega^{(j_1+j_2-j)\beta}e_{k_1(b_1+\beta)}\otimes e_{k_2(b_2+\beta)}\\ &= \sum_{\beta,a_1,a_2}\delta_{kc,k\beta}\omega^{(j_1+j_2-j)\beta}\delta_{k_1(b_1+\beta),k_1a_1} \times\delta_{k_2(b_2+\beta),k_2a_2}e_{k_1a_1}\otimes e_{k_2a_2}\\ &= \sum_{a_1,a_2}\delta_{k_1a_1+k_2a_2-kc,i-i_1-i_2}e_{k_1a_1}\otimes e_{k_2a_2} \times\sum_{\beta}\delta_{kc,k\beta}\delta_{k_1(a_1-b_1),k_1\beta}\omega^{(j_1+j_2-j)\beta}.\end{aligned} \tag{5.16}$$

The last identity makes use of the assumption that the $\delta$'s in Eq. (5.15) don't vanish.

One is now able to do the summation over $\beta$

$$\begin{aligned}&\sum_{\beta}\delta_{kc,k\beta}\delta_{k_1(a_1-b_1),k_1\beta}\omega^{(j_1+j_2-j)\beta}\\ &\quad= \delta_{\tau(a_1-b_1-c),0}\sum_{\beta}\delta_{l\bar{k}c+l_1\bar{k}_1(a_1-b_1),\kappa\beta}\omega^{(j_1+j_2-j)\beta}\\ &\quad= \kappa\delta_{j_1+j_2-j,\epsilon\kappa}\delta_{\tau(a_1-b_1-c),0}\omega^{\epsilon[l\bar{k}c+l_1\bar{k}_1(a_1-b_1)]},\end{aligned} \tag{5.17}$$

where the definitions of $\kappa$, $\tau$, $l$, and $l_1$ are given by the Eqs. (4.1), (5.4), and (5.7). We now use the Eqs. (5.15)–(5.17) and obtain

$$P_c(e_{k_1b_1}\otimes e_{k_2b_2}) = \frac{\kappa}{\bar{k}}\delta_{k_1+k_2,k}\delta_{k_1b_1+k_2b_2,i-i_1-i_2} \times\delta_{j_1+j_2-j,\epsilon\kappa}\sum_{a_1,a_2}\delta_{k_1a_1+k_2a_2-kc,i-i_1-i_2}\delta_{\tau(a_1-b_1-c),0} \times\omega^{\epsilon(l\bar{k}c+l_1\bar{k}_1(a_1-b_1))}e_{k_1a_1}\otimes e_{k_2a_2} \cdot \tag{5.18}$$

If we take $c = 0$ in Eq. (5.18) we get

$$P_0(e_{k_1b_1}\otimes e_{k_2b_2}) = \frac{\kappa}{\bar{k}}\delta_{k_1+k_2,k}\delta_{j_1+j_2-j,\epsilon\kappa} \times\sum_{a_1,a_2}\delta_{k_1(b_1-a_1)+k_2(b_2-a_2),0}\delta_{k_1a_1+k_2a_2,i-i_1-i_2} \cdot \times\delta_{\tau(a_1-b_1),0}\omega^{\epsilon l_1\bar{k}_1(a_1-b_1)} \cdot \times e_{k_1a_1}\otimes e_{k_2a_2} \cdot \tag{5.19}$$

Making various choices for $b_1$ and $b_2$ one may obtain vectors which are proportional to one another. The following vectors however are either orthogonal or zero:

$$R_{\xi,0} = \sqrt{\kappa/\bar{k}}\,\delta_{k_1+k_2,k}\delta_{j_1+j_2-j,\epsilon\kappa} \times\sum_{a_1,a_2}\delta_{k_1a_1+\delta_{k,a_1},i-i_1-i_2}\delta_{\tau(a_1-\xi),0} \cdot \times\omega^{\epsilon l_1\bar{k}_1(a_1-\xi)}e_{k_1a_1}\otimes e_{k_2a_2}, \tag{5.20}$$

where the index $\xi$ takes values within the interval

$$0\leqslant\xi\leqslant(n/\tau)-1. \tag{5.21}$$

There is a simple algorithm which gives the nonzero vectors $R_{\xi,0}$. One writes $\xi$ in the following form

$$\xi = \xi_0 + \eta[\psi(\bar{k}_1,\bar{k}_2)/\bar{k}_1], \tag{5.22}$$

where

$$0\leqslant\eta\leqslant[n/\psi(\bar{k}_1,\bar{k}_2,\bar{k})]-1, \tag{5.23}$$

and $\xi_0$ is the smallest positive integer which is a solution of the equation

$$k_1\xi + k_2\zeta = i - i_1 - i_2. \tag{5.24}$$

$\zeta$ is an arbitrary number in $Z_n$. From the knowledge of the vector $R_{\xi,0}$, we can compute all the other basis vectors of the representation $[i,j;k]_{\eta}$ using Eqs. (5.18) and (5.20). [$\eta$ labels the various equivalent representations and is related to $\xi$ through Eq. (5.22)]:

$$R_{\xi,c} = P_c(R_{\xi,0}); \quad (1\leqslant c\leqslant\bar{k}-1). \tag{5.25}$$

## 6. CONCLUSIONS

We have shown that in the case of the $\mathscr{D}$ $(n)$ and Q $(n)$ groups one is able to find analytic expressions which give in one formula all the irreducible representations of all the groups (any $n$). The formula for the Clebsch–Gordan coefficients is again valid for all the irreducible representations of all the groups. A key point is that although the groups are not simply reducible one is able to find a label for the equivalent irreducible representation obtained in tensor products. This label [$\eta$ in Eqs. (4.2) and (4.3)] is also in $Z_n$. We conjecture that similar results are obtainable for other sequences of solvable non-Abelian groups.

Since the quaternionic group Q(2) is related to the quaternionic algebra $q(2)$, one may ask which are the algebras associated to the Q $(n)$ groups.

Notice from Eq. (2.5) that the quaternions can be labeled in a $(Z_2,Z_2)$ way

$$\sigma_{(0,0)} = \mathbf{1}; \quad \sigma_{(1,0)} = i\sigma_3; \quad \sigma_{(0,1)} = i\sigma_1; \quad \sigma_{(1,1)} = i\sigma_2 \tag{6.1}$$

and that the associative algebra of the $\sigma_{(\alpha,\beta)}$ matrices can be written

$$\sigma_{(\alpha,\beta)}\sigma_{(\alpha',\beta')} = \omega^{\alpha\beta' + \alpha\alpha' + \beta\beta'}\sigma_{(\alpha+\alpha',\beta+\beta')}, \tag{6.2}$$

where $\alpha, \beta, \alpha', \beta' \in Z_2$ and $\omega = e^{2\pi i/2}$. These algebras are simple and their $n$-dimensional representation can immediately be obtained from the [0,0;1] representation of the Q ($n$) groups. For example, the representation for the $\sigma_{(\alpha,\beta)}$ matrices in the Q(3) case are the matrices given by Eq. (2.6) in which we drop the factor $\omega^{\gamma}$. The properties of the $q$ ($n$) algebras will be presented elsewhere. It is important to notice that for the $Z_n$ Potts models used in statistical mechanics the $\sigma_{(1,0)}$ and $\sigma_{(0,1)}$ matrices of $q$ ($n$) play in the transfer matrix the same role as the $\sigma_3$ and $\sigma_1$ matrices for the Ising model.[7]


[1]M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Adison-Wesley, Reading, Mass., 1964).

[2]See, e.g., D. Wyler, Phys. Rev. D **19**, 3369 (1979), and references quoted therein.

[3]J. Bellissard, J. Math Phys. **20**, 1490 (1979); A. B. Zamolodchikov and M. I. Monastyrskii, Sov. Phys. JETP **50**, 167 (1979); J. M. Drouffe, Itzykson and J. B. Zuber, Nucl. Phys. B **147**, 132 (1979).

[4]For a review see R. Savit, Rev. Mod. Phys. **52**, 453 (1980).

[5]For the finite subgroups of SU(3), see W. M. Fairbairn, T. Fulton and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[6]R. Casalbuoni, V. Rittenberg, and S. Yankielowicz, Nucl. Phys. B (to be published).

[7]M. J. Stephen and L. Mittag, J. Math. Phys. **12**, 441 (1971).

[8]M. Hall, *The Theory of Groups* (McMillan, New York, 1959).

# Equivalence of induced representations

P. M. van den Broek
*Institute for Theoretical Physics, University of Nijmegen, Nijmegen, the Netherlands*
R. Dirl
*Institut für Theoretische Physik, T. U. Wien A-1040 Wien, Karlsplatz 13, Austria*

(Received 6 March 1979; accepted for publication 19 October 1979)

Equivalence of induced representations for finite groups is considered in order to determine those equivalence classes of space group representations which are linked by complex conjugation.

## 1. INTRODUCTION

An interesting group theoretical problem is to correlate induced representations which originate from different members of the same orbit. The reason for considering this problem arises there from that, e.g., the complex conjugation (or more general automorphisms) of representations of a given (finite) group are of interest for physical applications.[1,2] In particular complex conjugation of representations is of importance for the problem of determining "coupling coefficients",[3] whose utility in various branches of physics is well known (e.g., "equivalent operators"). So the main objective of this article is to correlate induced representations belonging to different members of an orbit, to specify the general results to space-group representations and to give rules (concerning the star and the projective representations of the corresponding little cogroup) determining those equivalence classes of space-group representations which are linked by complex conjugation. One of the reasons for discussing the last problem is to generalize the results of Ref. 2 to arbitrary space groups, since in Ref. 2 only space groups were considered which contain the inversion as point group operation.

The material is organized as follows: In Sec. 2 the usual induction procedure[4,5,6] is briefly recalled. In the following section we discuss in general the problem of carrying out the induction procedure from different members of the same orbit. There we derive the conditions under which the corresponding representations are identical. In order to be able to apply the results to space-group representaions, we recall briefly their definitions and properties in Sec. 4. Section 5 is devoted to derive representations of the little cogroups whose corresponding space-group representations are obtained by induction from different members of the considered star. These results are then used in Sec. 6 to derive the corresponding equivalence relations for complex conjugation and to verify the results of Ref. 2.

## 2. THE INDUCTION PROCEDURE

Let $H$ be an invariant subgroup of a given finite group $G$. The elements of $H$ are dentoed by $a,b,...$ (identity $e$), the elements of the factor group $K = G/H$ by $\alpha,\beta,...$ (identity $\epsilon$), and for each $\alpha\in K$, a fixed chosen right coset representative is denoted by $r(\alpha)$, where $r(\epsilon) = e$.

Let $\Delta$ be an unitary irreducible representation (unirrep) of $H$. For each $\alpha\in K$ we define $\Delta_\alpha$ by

$$\Delta_\alpha(a) = \Delta(r^{-1}(\alpha)ar(\alpha)), \quad \text{for all } a\in H, \tag{2.1}$$

which admits to define the little cogroup

$$\bar{K} = \{\alpha\in K \,|\, \Delta_\alpha \simeq \Delta\} \subseteq K. \tag{2.2}$$

Hence the corresponding little group of $\Delta$ is given by

$$L = \{ar(\alpha) | a\in H, \alpha\in\bar{K}\}. \tag{2.3}$$

Since for each $\alpha\in\bar{K}$ there exists a unitary matrix $U(\alpha)$ satisfying

$$\Delta(a) = U(\alpha)\Delta(r^{-1}(\alpha)ar(\alpha))U^{+}(\alpha), \quad \text{for all } a\in H, \tag{2.4}$$

the following matrices

$$V(ar(\alpha)) = \Delta(a)U(\alpha), \quad \text{for all } a\in H \text{ and } \alpha\in\bar{K}, \tag{2.5}$$

form a projective unirrep of $L$ which belongs to a factor system $\mu_\Delta$ of $K$, i.e.,

$$V(ar(\alpha))V(br(\beta)) = \mu_\Delta(\alpha,\beta)V(ar(\alpha)br(\beta)). \tag{2.6}$$

The subscript $\Delta$ occurring in the factor system $\mu$ shall indicate that the factor system depends on the given unirrep $\Delta$. In fact only the equivalence class of $\mu$ is determined by $\Delta$. The actual form of $\mu_\Delta$ still depends on the choice of the phases of the matrices $U(\alpha)$.

Due to the known induction procedure the allowable unirreps of $L$ are given by

$$D(ar(\alpha)) = V(ar(\alpha)) \otimes E(\alpha), \quad \text{for all } a\in H \text{ and } \alpha\in\bar{K}, \tag{2.7}$$

where $E$ runs through a complete set of projective unirreps of $\bar{K}$ which belong to the factor system $\mu_\Delta^*$.

Summarizing the induction procedure which gives a complete set of unirreps of $G$, one has a proceed as follows:

(i) Take one $\Delta$ from each orbit of unirreps of $H$, where the orbit of $\Delta$ is defined to be the set $\{\Delta_\alpha \,|\alpha\in K\}$,

(ii) Construct for each such $\Delta$ the allowable unirreps of $L$ by means of (2.5) and (2.7),

(iii) Induce these allowable representations to $G$, where the general definition of the induced representation $(\mathscr{D}\uparrow G)$ of $G$ which is induced from $\mathscr{D}$ is as follows: If a left coset decomposition of $G$ with respect to $L$ is given by $G = \Sigma_i g_i L$, then

$$(\mathscr{D}\uparrow G)_{jt,ks}(g) = \delta(g_j^{-1}gg_k, L)\mathscr{D}_{ts}(g_j^{-1}gg_k), \tag{2.8}$$

where

$$\delta(g, L) = \begin{cases} 0, & \text{if } g\notin L, \\ 1, & \text{if } g\in L. \end{cases} \tag{2.9}$$

 0022-2488/80/102487-04$1.00  

## 3. INDUCTION FROM DIFFERENT MEMBERS OF THE ORBIT

For the following considerations it is assumed that the given unirrep $D$ of $G$ has been obtianed by inducing the allowable unirrep $\mathscr{D}$ of $L$ to $G$, where $\mathscr{D}$ is given through (2.7) and (2.5) for some unirrep $\Delta$ of $H$ and some projective unirrep $E_0$ of $\bar{K}$. For notational convenience we will write $\mathscr{D} = (\Delta, E_0)$.

In this section we will examine how $D$ is obtained, if we do not start from $\Delta$, but from a different member of the orbit of $\Delta$. Suppose we start from $\Delta_\gamma$ where $\gamma \notin \bar{K}$ we derive by means of

$$(\Delta_\gamma)_\alpha(a) = \Delta^\dagger(r^{-1}(\alpha\gamma)r(\alpha)r(\gamma))\Delta(r^{-1}(\alpha\gamma)ar(\alpha\gamma)) \times \Delta(r^{-1}(\alpha\gamma)r(\alpha)r(\gamma)), \tag{3.1}$$

that the equivalence relation

$$(\Delta_\gamma)_\alpha \simeq \Delta_{\alpha\gamma} \tag{3.2}$$

holds, since $r^{-1}(\alpha\gamma)r(\alpha)r(\gamma) \in H$. Consequently

$$\bar{K}_\gamma = \gamma\bar{K}\gamma^{-1}, \tag{3.3}$$

which implies that the corresponding little cogroup $L_\gamma$ is given by

$$L_\gamma = r(\gamma)Lr^{-1}(\gamma). \tag{3.4}$$

Furthermore for each $\beta \in \bar{K}_\gamma$ there exists a unitary matrix $\tilde{U}(\beta)$ such that

$$(\Delta_\gamma)(a) = \tilde{U}(\beta)\Delta_\gamma(r^{-1}(\beta)ar(\beta))\tilde{U}^\dagger(\beta), \tag{3.5}$$

or, with (2.1),

$$\Delta(a) = \tilde{U}(\beta)\Delta(r^{-1}(\gamma)r^{-1}(\beta)r(\gamma)ar^{-1}(\gamma)r(\beta)r(\gamma)) \times \tilde{U}^\dagger(\beta). \tag{3.6}$$

Utilizing (2.4) we rewrite (3.6) as

$$\Delta(a) = \tilde{U}(\beta)U^\dagger(\gamma^{-1}\beta\gamma)\Delta^\dagger(x_\gamma(\beta))\Delta(a) \times \Delta(x_\gamma(\beta))U(\gamma^{-1}\beta\gamma)\tilde{U}^\dagger(\beta), \tag{3.7}$$

where we have introduced the notation

$$x_\gamma(\beta) = r^{-1}(\gamma)r(\beta)r(\gamma)r^{-1}(\gamma^{-1}\beta\gamma), \tag{3.8}$$

which is an element of $H$. Due to Schur's lemma we may take

$$\tilde{U}(\beta) = \Delta(x_\gamma(\beta))U(\gamma^{-1}\beta\gamma), \quad \text{for all } \beta \in \bar{K}_\gamma. \tag{3.9}$$

Consequently in accordance to (2.5) we define by

$$\tilde{V}(br(\beta)) = \Delta_\gamma(b)\tilde{U}(\beta), \quad \text{for all } b \in H \text{ and } \beta \in \bar{K}_\gamma, \tag{3.10}$$

a projective unirrep of $L_\gamma$. Its multiplication law is given by

$$\tilde{V}(br(\beta))\tilde{V}(b'r(\beta')) = \mu_{\Delta_\gamma}(\beta,\beta')\tilde{V}(br(\beta)b'r(\beta')), \tag{3.11}$$

where the factor system $\mu_{\Delta_\gamma}$ of $\bar{K}_\gamma$ is correlated by a simple formula to the original factor system $\mu_\Delta$ of $\bar{K}$

$$\mu_{\Delta_\gamma}(\beta,\beta') = \mu_\Delta(\gamma^{-1}\beta\gamma, \gamma^{-1}\beta'\gamma), \quad \text{for all } \beta,\beta' \in \bar{K}_\gamma. \tag{3.12}$$

In order to verify (3.12) one has to use (3.8), (3.9), and (3.11), where it suffices to set $b = b' = e$. Now it is obvious that the projective unirreps $\tilde{E}$ of $\bar{K}_\gamma$ with factor system $\mu_{\Delta_\gamma}$ can be chosen as

$$\tilde{E}(\beta) = E(\gamma^{-1}\beta\gamma), \quad \text{for all } \beta \in \bar{K}_\gamma, \tag{3.13}$$

where $E$ runs through all projective unirreps of $\bar{K}$, which belong to the factor system $\Delta_\gamma$. Hence the corresponding allowable unirreps of $L_\gamma$ are given by

$$\tilde{\mathscr{D}}(br(\beta)) = \tilde{V}(br(\beta)) \otimes \tilde{E}(\beta), \quad \text{for all } b \in H,\ \beta \in \bar{K}_\gamma, \tag{3.14}$$

and the unirreps of $G$ can be written as

$$(\tilde{\mathscr{D}} \uparrow G)_{jt,ks}(g) = \delta(\tilde{g}_j^{-1}g\tilde{g}_k, L_\gamma)\tilde{\mathscr{D}}_{ts}(\tilde{g}_j^{-1}g\tilde{g}_k), \tag{3.15}$$

where the $\tilde{g}_j$ are left coset representatives of $G$ with respect to $L_\gamma$. The aim of this section is to show that the equivalence relation

$$(\Delta, E_0)\uparrow G \simeq (\Delta_\gamma, \tilde{E}_0)\uparrow G, \tag{3.16}$$

holds. In order to simplify the following considerations we choose the left coset representatives $\tilde{g}_j$ as

$$\tilde{g}_j = g_j r^{-1}(\gamma). \tag{3.17}$$

With this choice (3.16) becomes an identity:

$$[(\Delta, E_0)\uparrow G](g) = [(\Delta_\gamma, \tilde{E}_0)\uparrow G](g) \quad \text{for all } g \in G. \tag{3.18}$$

In order to prove this result we have to show that

$$\delta(\tilde{g}_j^{-1}g\tilde{g}_k, L_\gamma) = \delta(g_j^{-1}gg_k, L), \quad \text{for all } g \in G, \tag{3.19}$$

and

$$(\Delta_\gamma, \tilde{E}_0)(\tilde{g}_j^{-1}g\tilde{g}_k) = (\Delta, E_0)(g_j^{-1}gg_k), \quad \text{for all } g \in g_j L g_k^{-1} = \tilde{g}_j L_\gamma \tilde{g}_k^{-1}. \tag{3.20}$$

The proof of (3.19) is trivial. The proof of (3.20) is less trivial, but still straightforward. If $g_j^{-1}gg_k = ar(\alpha)$ with $a \in H$ and $\alpha \in \bar{K}$ then

$$(\Delta, E_0)(g_j^{-1}gg_k) = V(ar(\alpha)) \otimes E_0(\alpha). \tag{3.21}$$

Now

$$\tilde{g}_j^{-1}g\tilde{g}_k = r(\gamma)ar(\alpha)r^{-1}(\gamma)r^{-1}(\gamma\alpha\gamma^{-1})r(\gamma\alpha\gamma^{-1}). \tag{3.22}$$

Since $r(\gamma)ar(\alpha)r^{-1}(\gamma)r^{-1}(\gamma\alpha\gamma^{-1}) \in H$ we have

$$(\Delta_\gamma, \tilde{E}_0)(\tilde{g}_j^{-1}g\tilde{g}_k) = \tilde{V}(r(\gamma)ar(\alpha)r^{-1}(\gamma)) \otimes \tilde{E}_0(\gamma\alpha\gamma^{-1}). \tag{3.23}$$

But $E_0(\alpha) = \tilde{E}_0(\gamma\alpha\gamma^{-1})$, so (3.20) is proved if we show that

$$V(ar(\alpha)) = \tilde{V}(r(\gamma)ar(\alpha)r^{-1}(\gamma)). \tag{3.24}$$

This equation can be proved in a straightforward way, first using the definitions (2.5) and (3.10), then the definitions (2.1), (3.8), and (3.9). This completes the proof of (3.18) for the case that the leftcoset representatives satisfy (3.17); therefore for arbitrary leftcoset representatiaves (3.16) has been proved.

## 4. UNIRREPS OF SPACE GROUPS

In order to be able to specialize our results to space groups let us briefly recall the basic definitions and notations concerning such groups with their unirreps. Let $G$ be a space group and $H = T$ its subgroup of translations. The elements of $T$ are denoted by $(E\,|\,\mathbf{t})$ (or briefly $\mathbf{t} \in T$), elements of $G/H = K = P$ are denoted by $R$ (representing either an abstract group element of the point group $P$ or the corresponding faithful matrix representation[6]) and coset representatives are denoted by $(R\,|\,\tau(R))$ [where $\tau(R)$ represents nonprimitive lattice translations], so that a general group element takes the form $(R\,|\,\tau(R) + \mathbf{t})$. The multiplication

rule reads as

$$(R_1|\tau(R_1)+\mathbf{t}_1)(R_2|\tau(R_2)+\mathbf{t}_2)$$
$$=(R_1R_2|\tau(R_1R_2)+\mathbf{t}(R_1,R_2)+\mathbf{t}_1+R_1\mathbf{t}_2)\,, \tag{4.1}$$

where

$$\mathbf{t}(R_1,R_2)=\tau(R_1)+R_1\tau(R_2)-\tau(R_1R_2)\,. \tag{4.2}$$

Let us start from the unirrep $\Delta^{\mathbf{q}}$ of $H$ given by

$$\Delta^{\mathbf{q}}(\mathbf{t})=e^{-i\mathbf{q}\cdot\mathbf{t}}\,,\quad \mathbf{q}\in\mathrm{BZ}\,, \tag{4.3}$$

where BZ denotes the Brillouin zone. Because of

$$\Delta^{\mathbf{q}}_R(\mathbf{t})=\Delta^{\mathbf{q}}((R|\tau(R))^{-1}(E|\mathbf{t})(R|\tau(R)))$$
$$=\Delta^{\mathbf{q}}(R^{-1}\mathbf{t})=e^{-iR\mathbf{q}\cdot\mathbf{t}}\,, \tag{4.4}$$

the corresponding little cogroup $\bar{K}=P^{\mathbf{q}}$ is given by

$$P^{\mathbf{q}}=\{R\in P\,|R\mathbf{q}=\mathbf{q}+\mathbf{Q}\{R\mathbf{q}\}\}\,, \tag{4.5}$$

where $\mathbf{Q}\{R\mathbf{q}\}$ is a reciprocal lattice vector. Hence

$$G^{\mathbf{q}}=\{(R|\tau(R)+\mathbf{t})|R\in P^{\mathbf{q}},\,\mathbf{t}\in T\}\,. \tag{4.6}$$

Obviously we can choose

$$U^{\mathbf{q}}(R)=1\,,\quad \text{for all } R\in P^{\mathbf{q}}\,, \tag{4.7}$$

which implies that

$$V^{\mathbf{q}}(R|\tau(R)+\mathbf{t})=\Delta^{\mathbf{q}}(\mathbf{t})\,,\quad \text{for all } R\in P^{\mathbf{q}},\,\mathbf{t}\in T\,, \tag{4.8}$$

and that the corresponding factor system is given by

$$\mu^{\mathbf{q}}(R,R')=e^{i\mathbf{q}\cdot\mathbf{t}(R,R')},\quad \text{for all } R,R'\in P^{\mathbf{q}}\,. \tag{4.9}$$

Hence the allowable unirreps of $G^{\mathbf{q}}$ are given by

$$\mathscr{D}^{\mathbf{q}}(R|\tau(R)+\mathbf{t})=e^{-i\mathbf{q}\cdot\mathbf{t}}B(R)\quad \text{for all } R\in P^{\mathbf{q}},\,\mathbf{t}\in T\,, \tag{4.10}$$

where $B$ runs through all inequivalent projective unirreps of $P^{\mathbf{q}}$ with factor system $\mu^{\mathbf{q}}(R,R')^*=e^{-i\mathbf{q}\cdot\mathbf{t}(R,R')}$.

The corresponding induced unirreps of $G$ can be written as

$$(\mathscr{D}^{\mathbf{q}}\uparrow G)_{jt,ks}(R|\tau(R)+\mathbf{t})$$
$$=\delta(\underline{R}_j^{-1}R\underline{R}_k,P^{\mathbf{q}})\mathscr{D}^{\mathbf{q}}_{ts}((\underline{R}_j|\tau(\underline{R}_j))^{-1}$$
$$\times(R|\tau(R)+\mathbf{t})(\underline{R}_k|\tau(\underline{R}_k)))\,, \tag{4.11}$$

where $(\underline{R}_j|\tau(\underline{R}_j))$ are left coset representatives of $G$ with respect to $G^{\mathbf{q}}$.

## 5. INDUCTION FROM DIFFERENT MEMBERS OF THE STAR

Let $\mathbf{q}'$ be a member of the star of $\mathbf{q}$ where

$$\mathbf{q}'=R_0\mathbf{q}+\mathbf{Q}\,,\quad R_0\in P\,,\quad R_0\notin P^{\mathbf{q}}\,, \tag{5.1}$$

such that $\mathbf{q}'\in$BZ. From the results of Sec. 3 it follows that

$$P^{\mathbf{q}'}=R_0P^{\mathbf{q}}R_0^{-1} \tag{5.2}$$

and

$$G^{\mathbf{q}'}=(R_0|\tau(R_0))G^{\mathbf{q}}(R_0|\tau(R_0))^{-1}\,. \tag{5.3}$$

Equation (3.9) specializes to

$$U^{\mathbf{q}'}(R)=\Delta^{\mathbf{q}}(\mathbf{x}_{R_0}(R))$$
$$=\exp[-i\mathbf{q}\cdot\mathbf{x}_{R_0}(R)],\quad \text{for all } R\in P^{\mathbf{q}'}\,, \tag{5.4}$$

where the vectors $\mathbf{x}_{R_0}(R)$ are given by

$$(E|\mathbf{x}_{R_0}(R))=(R_0|\tau(R_0))^{-1}(R|\tau(R))$$
$$\times(R_0|\tau(R_0))(R_0^{-1}RR_0|\tau(R_0^{-1}RR_0))\,. \tag{5.5}$$

A simple calculation gives

$$\mathbf{x}_{R_0}(R)=R_0^{-1}[-\tau(R_0)+\tau(R)$$
$$+R\tau(R_0)-R_0\tau(R_0^{-1}RR_0)]\,. \tag{5.6}$$

Note that usually one would take $U^{\mathbf{q}'}(R)=1$. Therefore Eq. (5.4) will turn out to be the reason for the occurrence of unimodular factors.

From (3.16) it follows that the allowable unirrep $\mathscr{D}^{\mathbf{q}}$ of $G^{\mathbf{q}}$ given by (4.10) corresponds to the allowable unirrep $\mathscr{D}^{\mathbf{q}'}$ of $G^{\mathbf{q}'}$ which is given by

$$\mathscr{D}^{\mathbf{q}'}(R|\tau(R)+\mathbf{t})$$
$$=e^{-iR_0\mathbf{q}\cdot\mathbf{t}}e^{-i\mathbf{q}\cdot\mathbf{x}_{R_0}(R)}B(R_0^{-1}RR_0)\,,$$
$$\text{for all } R\in P^{\mathbf{q}'}=R_0P^{\mathbf{q}}R_0^{-1}\,. \tag{5.7}$$

Hence if coset representatives are used which are defined by (3.17), it follows from (3.18) that

$$(\mathbf{q},B)\uparrow G=(\mathbf{q}',B')\uparrow G\,, \tag{5.8}$$

where we have introduced obvious notation with

$$B'(R)=\exp[-i\mathbf{q}\cdot\mathbf{x}_{R_0}(R)]B(R_0^{-1}RR_0)\,,$$
$$\text{for all } R\in P^{\mathbf{q}'}\,. \tag{5.9}$$

## 6. COMPLEX CONJUGATION OF SPACE-GROUP REPRESENTATIONS

Now we are in the position to consider in more detail the problem of complex conjugation of space-group unirreps $(\mathbf{q},B)\uparrow G$, where $\mathbf{q}$ belongs to the so-called "representation domain" $\Delta$ BZ of BZ. From (2.8) and (4.10) it follows that

$$[(\mathbf{q},B)\uparrow G]^*=(-\mathbf{q},B^*)\uparrow G\,. \tag{6.1}$$

The vector $-\mathbf{q}$ does not in general belong to $\Delta$ BZ, not even to BZ. Since it is usual to consider only vectors from $\Delta$ BZ we want to rewrite the right-hand side of (6.1) as $(\mathbf{q}',B)\uparrow G$, where $\mathbf{q}'\in\Delta$ BZ. Here $\mathbf{q}'$ is of course of the form

$$\mathbf{q}'=-R_0\mathbf{q}+\mathbf{Q}\,, \tag{6.2}$$

for some $R_0\in P$ and some $\mathbf{Q}$ from the reciprocal lattice. Both $R_0$ and $\mathbf{Q}$ can be easily determined in practical problems. The form of $B'$ can now be determined at once from (5.8) and (5.9):

$$B'(R)=\exp[i\mathbf{q}\cdot\mathbf{x}_{R_0}(R)]B^*(R_0^{-1}RR_0)\,,$$
$$\text{for all } R\in P^{-R_0\mathbf{q}}=R_0P^{-\mathbf{q}}R_0^{-1}\,. \tag{6.3}$$

So our result is

$$[(\mathbf{q},B)\uparrow G]^*=(-R_0\mathbf{q}+\mathbf{Q},B')\uparrow G\,, \tag{6.4}$$

where $B'$ is given by (6.3) and $R_0$ and $\mathbf{Q}$ are easily determined such that $-R_0\mathbf{q}+\mathbf{Q}\in\Delta$ BZ. Note that the equals sign in (6.4) should be replaced by an equivalance sign if the condition (3.17) is not taken into account, which says that the left coset representatives of $G$ with respect to $G^{\mathbf{q}}$ (respectively $G^{-R_0\mathbf{q}}$) should be related as follows: If $(R_j|\tau(R_j))$ is the $j$th left coset representative of $G$ with respect to $G^{\mathbf{q}}$, then $(R_jR_0|\tau(R_jR_0))$ is the $j$th left coset representative of $G$ with respect to $G^{-R_0\mathbf{q}}$. Also the equals sign should be replaced by an equivalence sign if $B'$ is not taken equal to the right-hand side of (6.3) but is only equivalent with it. This may happen for instance if one determines only the characters of $B'$ from

(6.3) and then takes the matrices of $B'$ from some other source.

Let us consider the special case

$$I \in P \simeq G/T, \tag{6.5}$$

i.e., the inversion $I$ belongs to the point group $P$. This special case was considered in detail in Ref. 2. For this case we have to distinguish two cases. Especially if for a given $\mathbf{q}$ we have $I \notin P^{\mathbf{q}}$ or $I \in P^{\mathbf{q}}$.

*Case* $I \notin P^{\mathbf{q}}$: Because of (6.2) it is possible to choose $R_0 = I$ what leads us to

$$B'(R) = e^{i\mathbf{q}\cdot\mathbf{x}_I(R)} B^*(R), \quad \text{for all } R \in P^{\mathbf{q}}, \tag{6.6}$$

where

$$\mathbf{x}_I(R) = \tau(I) - 2\tau(R) - R\tau(I). \tag{6.7}$$

Changing the factor system of the projective unirreps according to Eq. (1.9) of Ref. 7, i.e.,

$$B(R) = e^{-i\mathbf{q}\cdot\tau(R)} F(R), \tag{6.8}$$

$$B'(R) = e^{-i\mathbf{q}\cdot\tau(R)} F'(R), \tag{6.9}$$

we obtain

$$F'(R) = e^{i(\mathbf{q} - R^{-1}\mathbf{q})\cdot\tau(I)} F(R)^*, \tag{6.10}$$

which coincides with Eq. (3.8) of Ref. 2, if one uses that $R\mathbf{q} = \mathbf{q} + \mathbf{Q}\{R\mathbf{q}\}$, if $R \in P^{\mathbf{q}}$. Obviously if $\mathbf{q}$ does not belong to the boundary of $\Delta$ BZ, we have $\mathbf{Q}\{R\mathbf{q}\} = \mathbf{0}$, which implies that

$$F(R)^* = F'(R), \quad \text{for all } R \in P^{\mathbf{q}}. \tag{6.11}$$

Note that we do not have $B'(R) = B^*(R)$ in this case; so the disappearance of the unimodular factor is due to the choice of the factor system.

*Case* $I \in P^{\mathbf{q}}$: Since we have $-\mathbf{q} = \mathbf{q} + \mathbf{Q}$ in this case we can take $R_0$ equal to $E$ in (6.2). Equation (6.4) then becomes

$$[(\mathbf{q}, B)\uparrow G]^* = (\mathbf{q}, B')\uparrow G, \tag{6.12}$$

and Eq. (6.3) becomes

$$B'(R) = B^*(R), \quad \text{for all } R \in P^{\mathbf{q}}. \tag{6.13}$$

If changing the factor system in the same way as in (6.8) and (6.9) we arrive to the equivalence relation

$$F'(R) \simeq r^{-i\mathbf{Q}\cdot\tau(R)} F^*(R). \tag{6.14}$$

Equation (6.14) is equal to Eq. (3.16) of Ref. 2. Opposite to case $I \notin P^{\mathbf{q}}$, here the change of the factor system is the reason for the occurrence of the unimodular factor in (6.14).

*Concluding remarks*: The aim of this paper was to compare equivalent induced representations and to derive conditions under what circumstances such representations are identical. The results obtained have been applied to space-group representations. There we succeeded in deriving a simple relation which determines those equivalence classes which are linked by complex conjugation.


[1] J.F. Cornwell, Phys. Status Solidi **37**, 225 (1970).
[2] R. Dirl, J. Math. Phys. **20**, 1566 (1979).
[3] R. Dirl, J. Math. Phys. **20**, 1562 (1979).
[4] G.W. Mackey, Acta Mathematica **99**, 265 (1958).
[5] R. Dirl, J. Math. Phys. **18**, 2065 (1977).
[6] C.J. Bradley and A.P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, England, 1972).
[7] R. Dirl, J. Math. Phys. **20**, 664 (1979).

# Spinor fields invariant under space–time transformations[a)]

J. Beckers, J. Harnad,[b)] and P. Jasselette

*Physique Théorique et Mathématique, Université de Liège, [c)] Belgium*

(Received 12 June 1979; accepted for publication 15 October 1979)

Spinor fields *invariant* under the subgroups of the Poincaré group *or* under the maximal subgroups of the conformal group of space–time are analyzed. It is shown that only certain Poincaré subgroups, all of dimension less than or equal to six, can leave two component spinor fields invariant, with rather severe restrictions on the fields. Tables listing all such invariant fields for subgroups of dimension greater than or equal to four are given. Construction of Dirac spinors and connections between invariant spinors and tensors are discussed: In particular it is shown that from any two-component spinor invariant under a Poincaré subgroup a real skew-symmetric tensor invariant under the same group may be constructed.

## 1. INTRODUCTION

The subgroup structure of the Euclidean, Poincaré, and Minkowskian conformal groups have received much recent consideration.[1–6] In particular, Patera, Winternitz, and Zassenhaus[2,3] and others[4–6] have determined all the maximal subalgebras of the conformal Lie algebra c(3,1) of space–time and identified the complete subalgebra structures of each of these. For the Poincaré group P(3,1), the Patera–Winternitz–Zassenhaus[2] classification is equivalent to one obtained by Bacry–Combe–Sorba.[4]

These analyses have been applied to the study of invariant fields appearing in physics in the following ways: (i) electromagnetic fields invariant under subgroups of the Poincaré group were systematically studied[7–9]; (ii) classifications of symmetry breaking interactions in the Schrödinger equation were obtained[10–12]; (iii) electromagnetic 1 and 2 forms, symmetric (0,2) tensors, and scalar densities invariant under all the maximal subgroups of C(3,1) were determined[13–15]; (iv) solutions to the Yang–Mills equations invariant under the compact subgroups of the conformal group were obtained.[16,17] For further applications, see Beckers, Harnad, Perroud, and Winternitz[13] hereafter referred as BHPW, and also Refs. 17–19.

In this paper, we shall be concerned with the problem of determining the fundamental spinor fields invariant under subgroups of P(3,1) or under the maximal subgroups of C(3,1). This work is a continuation of BHPW, the latter dealing with tensor fields and densities invariant under subgroups of the conformal group of space–time.

The contents of this paper are summarized as follows. Section 2 presents conventions, notations, and the two methods of calculation—finite and infinitesimal—for determining spinor fields invariant under subgroups of P(3,1). Section 3 deals with the explicit determination of $(\frac{1}{2},0)$-spinor fields invariant under Poincaré subgroups of dimension $\geqslant 4$ and illustrates the methods with two specific examples. The results are summarized in Tables I–III. In Sec. 4, the invariance conditions for spinors under conformal transformations are given and it is shown that no nontrivial spinor field is invariant under the maximal subgroups of C(3,1). Section 5 discusses the construction of invariant four component (Dirac) spinors and the connections between invariant 2-component spinors and real skew-symmetric tensor fields of rank 2.

## 2. NOTATIONS AND INVARIANCE CONDITIONS

Let $M$ denote Minkowski space with metric $g_M$ identified with the diagonal matrix

$$g_M \equiv \mathrm{diag}(1,-1,-1,-1). \tag{2.1}$$

According to the notations of BHPW, the Poincaré transformations are generated by the following vector fields:

$$M_{\mu\nu} = -x_\mu \partial_\nu + x_\nu \partial_\mu , \tag{2.2a}$$

$$P_\mu = -\partial_\mu , \tag{2.2b}$$

where the $M_{\mu\nu}$'s generate homogeneous Lorentz transformations and the $P_\mu$'s space–time translations. The corresponding Lie algebra [the Poincaré algebra p(3,1)] is

$$[M_{\mu\nu},M_{\sigma\tau}] = g_{\mu\sigma}M_{\nu\tau} + g_{\nu\tau}M_{\mu\sigma} - g_{\mu\tau}M_{\nu\sigma} - g_{\nu\sigma}M_{\mu\tau} , \tag{2.3}$$

$$[M_{\mu\nu},P_\sigma] = g_{\mu\sigma}P_\nu - g_{\nu\sigma}P_\mu .$$

We denote an arbitrary vector field induced by a one-parameter subgroup of P(3,1) as

$$X = \tfrac{1}{2}\omega^{\mu\nu}M_{\mu\nu} + \alpha^\mu P_\mu , \quad (\omega^{\mu\nu} = -\omega^{\nu\mu}), \tag{2.4}$$

corresponding to the *infinitesimal* transformation

$$x \to x' : x'^\mu = x^\mu - \omega^\mu{}_\nu x^\nu + \alpha^\mu . \tag{2.5}$$

The $\omega^{\mu\nu}$ are real constants more conveniently defined by

$$\phi^i = \omega^{0i}, \qquad \theta^i = \tfrac{1}{2}\epsilon^{ijk}\omega_{jk}, \quad (i,j,k = 1,2,3), \tag{2.6}$$

in correspondence with the frequently used (Lorentz) basis for the $M_{\mu\nu}$'s, i.e.,

$$K_i = M_{0i} = -x^0\partial_i - x^i\partial_0 ,$$

$$L_i = -\tfrac{1}{2}\epsilon_{ijk}M^{jk} = -\epsilon_{ijk}\,x^j\partial_k . \tag{2.7}$$

In terms of Eqs. (2.6) and (2.7), the vector field $X$ can

[a)]Supported in part by the National Research Council of Canada and "Les Accords Culturels Belgo-Québecois, 1978".

[b)]Permanent address: Centre de Recherches Mathématiques, Université de Montréal, Montreal H3C 3J7, Canada.

[c)]Postal address: Institut de Physique au Sart Tilman, Bâtiment B.5, B-4000 Liege 1, Belgium.

 0022-2488/80/102491-09$1.00  

also be written:

$$X = \boldsymbol{\phi}\cdot\mathbf{K} - \boldsymbol{\theta}\cdot\mathbf{L} + \alpha\cdot P. \tag{2.8}$$

For later use, let us identify the three-dimensional Abelian subgroup generated by the algebra

$$\{P_0 + P_3, A_1, A_2\}, \tag{2.9}$$

where

$$A_1 \equiv L_2 - K_1, \quad A_2 \equiv L_1 + K_2\ . \tag{2.10}$$

We wish to study invariant *spinor* fields and therefore must regard the group generated by the homogeneous Lorentz subalgebra $\{M_{\mu\nu}\}$ as the universal covering group of the homogeneous Lorentz group, i.e., the group SL(2,$\mathbb{C}$) of complex unimodular 2 by 2 matrices. The spinors transform according to irreducible representations of SL(2,$\mathbb{C}$) and their properties are well known.[20-22] To each finite Lorentz transformation $\Lambda$ corresponds a matrix $U \in$ SL(2,$\mathbb{C}$) defined up to a sign $[\Lambda \leftrightarrow (U, -U)]$ which acts on fundamental $(\frac{1}{2},0)$ spinors $v \in \mathbb{C}^2$ as

$$U : v \longmapsto Uv. \tag{2.11}$$

The correspondence is determined by associating with each point $x = (x^0, x^1, x^2, x^3)$ the Hermitian $2\times 2$ matrix:

$$\tilde{x} \equiv x^\mu \sigma_\mu = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix} \in \mathbb{H}(2) \tag{2.12}$$

(with $\sigma_i$ the Pauli matrices and $\sigma_0 = 1$) upon which the $\Lambda$ action is

$$\tilde{x} \longrightarrow U\tilde{x}U^{+} \equiv \widetilde{\Lambda x} \tag{2.13a}$$

and translations are given by

$$\tilde{x} \longrightarrow \tilde{x} + \tilde{a} \tag{2.13b}$$

where

$$\tilde{a} \equiv a^\mu \sigma_\mu \in \mathbb{H}(2).$$

Under an *infinitesimal* ($\Lambda \simeq 1 - \omega, \alpha \ll 1$) Poincaré transformation (2.5), we get

$$\tilde{x}' \sim U\tilde{x}U^{+} + \tilde{\alpha}, \tag{2.14}$$

where

$$U \equiv \mathbb{1} + \tfrac{1}{2}\boldsymbol{\Omega}\cdot\boldsymbol{\sigma}, \quad \tilde{\alpha} \approx \tilde{a}, \tag{2.15}$$

and

$$\boldsymbol{\Omega} = \boldsymbol{\phi} + i\boldsymbol{\theta}. \tag{2.16}$$

With these conventions, the $(\frac{1}{2},0)$ representation of sl(2,$\mathbb{C}$) is given by

$$\rho(L_i) = -(i/2)\sigma_i, \quad \rho(K_i) = \tfrac{1}{2}\sigma_i\ . \tag{2.17}$$

In particular, we have

$$\rho(A_1) = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \quad \rho(A_2) = \begin{pmatrix} 0 & -i \\ 0 & 0 \end{pmatrix}. \tag{2.18}$$

The spinor fields corresponding to this $(\frac{1}{2},0)$ representation are two component quantities denoted hereafter as $\Psi = (\Psi^1, \Psi^2)$.

The invariant spinors can be obtained through two alternative methods (see BHPW for the analogous methods applied to tensors and densities):

(1) a *global* method based on the determination of (regular and singular) orbits, their corresponding isotropy groups, and on the action of the group matrices on the spinors (a regular orbit is one belonging to an open stratum; the others, of lower dimension, are called singular: see BHPW, Sec. 3);

(2) an *infinitesimal* method, involving the solution of the Lie differential equations expressing invariance.

Let us summarize the procedure for each.

### A. Global method

Given a (pseudo-) Riemannian space $M$ on which a Lie group $G$ of isometries acts, the determination of all possible spinor fields $\in D^{(1/2,0)}$ which are invariant under $G$ proceeds through the following steps:

(i) Decompose $M$ into $G$ orbits. (Generally, there are a finite number of *generic*, open strata, whose union is a dense submanifold of $M$, and it is sufficient to consider only these.)

(ii) Choose a convenient point $p_0$ on each orbit such that their union forms, if possible, a smooth cross section $\Sigma$ to the orbits in the generic strata, and identify the isotropy group $G_0(p_0) \subset G$ at each $p_0$. Such a cross section might only exist locally.

(iii) Find a spinor at $p_0$ invariant under $G(p_0)$ by solving the linear algebraic equations expressing this invariance.

(iv) Apply the group transformations in order to obtain the invariant spinors at any point on the orbit of each $p_0$.

(v) Allow the free parameters determined in (iii) to vary smoothly along the cross section $\Sigma$.

We refer the reader to BHPW for a more complete discussion (in the case of tensors and densities) and for specific examples involving subgroups of the *conformal* group.

For tensors, no restriction need be made on $G$, except that it be a smooth transformation group. For spinors, however, we are limited to isometries in order that the group action may be lifted to the bundle of spinor frames. This may be extended to include conformal transformations by choosing a scaling weight for the spinor representation. However, for more general transformation groups, the differential action cannot be defined on finite-component spinor representations. In the following we shall only be concerned with Minkowski space and subgroups of P(3,1) and C(3,1).

### B. Infinitesimal method

The problem reduces here to solving the equations which express the vanishing of the Lie derivative of the spinor

$$L_{X_i}\Psi = 0, \tag{2.19}$$

where $\{X_i\}$ are the vector fields induced by the one-parameter subgroups of $G$. For finite P(3,1) transformations, the invariance condition for $(\frac{1}{2},0)$ spinors

$$\Psi(gx) = \rho(g)\Psi(x), \quad \forall\, g \in G, \tag{2.20}$$

becomes

$$\Psi(U\tilde{x}\,U^{+} + \tilde{a}) = U\Psi(\tilde{x}). \tag{2.21}$$

Retaining only first order terms [Eqs. (2.14)–(2.16)] gives the infinitesimal invariance condition

TABLE I. Six, five, and four-dimensional NON-SPIPS.

| PWZ notation [2] | Generators |
|---|---|
| *Six dimensions* | |
| $P_{1,2};P_{3,2};P_{4,2};$ | $\{\mathbf{L},\mathbf{K}\};\{\mathbf{L},\mathbf{P}\};\{L_3,K_1,K_2,P_0,P_1,P_2\};$ |
| $P_{6,2};P_{8,1};P_{9,1}$ | $\{L_3,A_1,A_2,P_0+P_3,P_1,P_2\};\{K_3,A_2,P_\mu\};\{L_3,K_3,P_\mu\}$ |
| *Five dimensions* | |
| $P_{2,3};P_{8,3};$ | $\{L_3,K_3,A_1,A_2,P_0+P_3\};\{K_3,A_2,P_0,P_2,P_3\};$ |
| $P_{9,2};P_{11,1};$ | $\{L_3,K_3,P_0+P_3,P_1,P_2\};\{L_3\cos f-K_3\sin f,P_\mu\};$ |
| $P_{12,1};P_{13,1}$ | $\{L_3,P_\mu\};\{K_3,P_\mu\}$ |
| *Four dimensions* | |
| $P_{2,4};P_{3,3};P_{4,3};$ | $\{L_3,K_3,A_1,A_2\};\{\mathbf{L},P_0\};\{L_3,K_1,K_2,P_3\};$ |
| $P_{6,3};P_{9,3};P_{9,4};$ | $\{L_3,A_1,A_2,P_0+P_3\};\{L_3,K_3,P_1,P_2\};\{L_3,K_3,P_0,P_3\};$ |
| $P_{12,2};P_{12,3};P_{12,4};$ | $\{L_3,P_0+P_3,P_1,P_2\};\{L_3,\mathbf{P}\};\{L_3,P_0,P_1,P_2\};$ |
| $P_{13,3};\widetilde{P}_{6,7};\widetilde{P}_{6,8}$ | $\{K_3,P_0,P_1,P_3\};\{L_3,A_1+P_2,A_2+P_1,P_0+P_3\};\{L_3,A_1-P_2,A_2-P_1,P_0+P_3\}$ |

$$\mathscr{D}\Psi\equiv\left(\alpha^\mu\partial_\mu+(\mathbf{x}\cdot\boldsymbol{\phi})\frac{\partial}{\partial t}+(t\boldsymbol{\phi}+\mathbf{x}\times\boldsymbol{\theta})\cdot\frac{\partial}{\partial\mathbf{x}}\right)\Psi$$
$$=\tfrac{1}{2}(\boldsymbol{\Omega}\cdot\boldsymbol{\sigma})\Psi. \tag{2.22}$$

Equation (2.22) corresponds to the following set of conditions on the two spinor components:

$$\mathscr{D}\Psi^1=\tfrac{1}{2}[\Omega^3\Psi^1+(\Omega^1-i\Omega^2)\Psi^2] \tag{2.23a}$$

and

$$\mathscr{D}\Psi^2=\tfrac{1}{2}[(\Omega^1+i\Omega^2)\Psi^1-\Omega^3\Psi^2]. \tag{2.23b}$$

Let us remark that the analogous conditions for invariance of 2 forms $F$ are written (Combe–Sorba[8])

$$\mathscr{D}\mathbf{E}+\boldsymbol{\theta}\times\mathbf{E}+\boldsymbol{\phi}\times\mathbf{B}=0 \tag{2.24a}$$

and

$$\mathscr{D}\mathbf{B}+\boldsymbol{\theta}\times\mathbf{B}-\boldsymbol{\phi}\times\mathbf{E}=0, \tag{2.24b}$$

where $E^i=F^{0i}$ and $B^i=\frac{1}{2}\epsilon^{ijk}F_{jk}$ are, respectively, the "electric" and "magnetic" components. The Bacry–Combe–Richard[7] conditions correspond to $\mathscr{D}\equiv0$ in Eqs. (2.24).

Since we shall be interested in Dirac spinors (see Sec. 5) we also need information on spinors of the other (nonequivalent) fundamental representation $D^{(0,1/2)}$ of SL(2,$\mathbb{C}$) acting on dotted spinors $\{v^{\dot a}\}$ by the complex conjugate transformation $v^{\dot a}\to\overline{U}^{\dot a\dot b}v^{\dot b}$. The corresponding spinor fields will be denoted $\phi(x)$. In correspondence with Eq. (2.21), we have for $(0,\frac{1}{2})$ spinors

$$\phi(U\tilde xU^{+}+\tilde a)=\overline{U}\phi(\tilde x) \tag{2.25}$$

and, in infinitesimal form, we get

$$\mathscr{D}\phi=\tfrac{1}{2}(\boldsymbol{\Omega}^*\cdot\boldsymbol{\sigma}^*)\phi \tag{2.26}$$

($\boldsymbol{\Omega}^*$ and $\boldsymbol{\sigma}^*$ are the complex conjugates of $\boldsymbol{\Omega}$ and $\boldsymbol{\sigma}$). In general, a $(p/2,q/2)$ spinor will transform under any $U\in$SL(2,$\mathbb{C}$) according to

$$f_U:\Psi^{a_1\cdots a_p\dot b_1\cdots\dot b_q}$$
$$\to U^{a_1c_1}\cdots U^{a_pc_p}\overline{U}^{\dot b_1\dot d_1}\cdots\overline{U}^{\dot b_q\dot d_q}\Psi^{c_1\cdots c_p\dot d_1\cdots\dot d_q}$$

so that the invariance conditions generalizing Eqs. (2.21) and (2.25) become

$$\Psi^{a_1\cdots a_p\dot b_1\cdots\dot b_q}(U\tilde xU^{+}+\tilde a)$$
$$=U^{a_1c_1}\cdots U^{a_pc_p}\overline{U}^{\dot b_1\dot d_1}\cdots\overline{U}^{\dot b_q\dot d_q}\Psi^{c_1\cdots c_p\dot d_1\cdots\dot d_q}(\tilde x) \tag{2.27}$$

TABLE II. Six- and five-dimensional SPIPS and their invariant $(\frac{1}{2},0)$-spinor fields.

| PWZ notation | Generators | Fields [a] |
|---|---|---|
| *Six dimensions* | | |
| $P_{10,1}$ | $\{A_1,A_2,P_\mu\}$ | $(A,0)$ |
| $P_{7,2}$ | $\{K_3,A_1,A_2,P_0+P_3,P_1,P_2\}$ | $(\xi^{-1}A,0)$ |
| $\left.\begin{matrix}\widetilde{P}_{6,5}\\ \widetilde{P}_{6,6}\end{matrix}\right\}$ | $\{L_3\pm(P_0-P_3),A_1,A_2,P_0+P_3,P_1,P_2\}$ | $(e^{\mp(i/4)(t-z)},0)$ |
| $P_{5,2}$ | $\{L_f,A_1,A_2,P_0+P_3,P_1,P_2\}$ | $(\xi^{-1-i\cot f}A,0)$ |
| *Five dimensions* | | |
| $P_{14,1}$ | $\{A_2,P_\mu\}$ | |
| $\widetilde{P}_{10,6}$ | $\{A_1,A_2+\frac{1}{2}(P_0-P_3),P_0+P_3,P_1,P_2\}$ | $(A,0)$ |
| $P_{7,3}$ | $\{K_3,A_1,A_2,P_0+P_3,P_2\}$ | |
| $P_{8,2}$ | $\{K_3,A_2,P_0+P_3,P_1,P_2\}$ | $(\xi^{-1}A,0)$ |
| $\widetilde{P}_{7,6}$ | $\{K_3,+(1/a)P_1,A_1,A_2,P_0+P_3,P_2\},a>0$ | |
| $P_{10,2}$ | $\{A_1,A_2,P_0+P_3,P_1,P_2\}$ | $[S(t-z),0]$ |
| $\widetilde{P}_{8,10}$ | $\{K_3+(1/a)P_1,A_2,P_0,P_2,P_3\},a>0$ | $(e^{ax/2}A,0)$ |

[a] $\xi=[(t-z)/2]^{1/2}$; $A$, $B$ = arbitrary constants.

TABLE III. Four-Dimensional SPIPS and their invariant $(\frac{1}{2},0)$-spinor fields.

| PWZ notation | Generators | Fields[a] |
|---|---|---|
| $P_{10,3}$ | $\{A_1,A_2,P_0+P_3,P_2\}$ | |
| $P_{14,2}$ | $\{A_2,P_0+P_3,P_1,P_2\}$ | $[S(\xi),0]$ |
| $\widetilde{P}_{10,9};\widetilde{P}_{10,10}$ | $\{A_1\pm P_2,A_2,P_0+P_3,P_1\}$ | |
| $P_{14,3}$ | $\{A_2,P_0,P_2,P_3\}$ | $[S(x),0]$ |
| $P_{8,5}$ | $\{K_3,A_2,P_0+P_3,P_2\}$ | $[\xi^{-1}S(x),0]$ |
| $\widetilde{P}_{10,8}$ | $\{A_1,A_2-\frac{1}{2}(P_0-P_3),P_0+P_3,P_1\}$ | $\left(S\left[y+\frac{(t-z)^2}{2}\right],0\right)$ |
| $\widetilde{P}_{8,12}$ | $\left\{K_3+\frac{1}{2a}P_1,A_2,P_0+P_3,P_2\right\}$, $a>0$ | $\{e^{ax}S[e^{2ax}(t-z)],0\}$ equivalently $\{\xi^{-1}S[2ax+\ln(t-z)],0\}$ |
| $P_{15,1}$ | $\{P_\mu\}$ | $(A,B)$ |
| $P_{13,2}$ | $\{K_3,P_0+P_3,P_1,P_2\}$ | $(\xi^{-1}A,\xi B)$ |
| $P_{11,2}$ | $\{L_f,P_0+P_3,P_1,P_2\}$ | $(\xi^{-1+i\cot f}A,\xi^{1-i\cot f}B)$ |
| $\widetilde{P}_{12,11};\widetilde{P}_{12,12}$ | $\{L_3\pm\frac{1}{4}(P_0-P_3),P_0+P_3,P_1,P_2\}$ | $(e^{\mp i(t-z)}A,e^{\pm i(t-z)}B)$ |
| $\widetilde{P}_{12,13}$ | $\left\{L_3+\frac{1}{2a}P_0,\mathbf{P}\right\}$, $a>0$ | $(e^{-iat}A,e^{iat}B)$ |
| $\widetilde{P}_{12,14}$ | $\left\{L_3+\frac{1}{2b}P_3,P_0,P_1,P_2\right\}$, $b\neq0$ | $(e^{-ibz}A,e^{ibz}B)$ |
| $\widetilde{P}_{13,10}$ | $\left\{K_3-\frac{1}{2a}P_2,P_0,P_1,P_3\right\}$, $a>0$ | $(e^{-ay}A,e^{ay}B)$ |
| $\widetilde{P}_{14,10}$ | $\{A_2-\frac{1}{2}(P_0-P_3),P_0+P_3,P_1,P_2$ | $[A+i(t-z)B,B]$ |
| $\widetilde{P}_{14,11};\widetilde{P}_{14,12}$ | $\{A_2\pm P_1,P_0,P_2,P_3\}$ | $(A\pm ixB,B)$ |
| $P_{8,4}$ | $\{K_3,A_2,P_0+P_3,P_1\}$ | $\left(\xi^{-1}A-\frac{i}{2}y\xi^{-1}B,\xi B\right)$ |
| $P_{7,4}$ | $\{K_3,A_1,A_2,P_0+P_3\}$ | $[\xi^{-1}A+\frac{1}{2}(x-iy)\xi^{-1}B,\xi B]$ |
| $P_{8,6}$ | $\left\{K_3,A_2,P_0+P_3,P_2-\frac{1}{b}P_1\right\}$, $b\neq0$ | $\left[\xi^{-1}A-\frac{i}{2}(bx+y)\xi^{-1}B,\xi B\right]$ |
| $\widetilde{P}_{10,7}$ | $\left\{A_1+\frac{1}{b}P_2,A_2-\frac{1}{2}(P_0-P_3),P_0+P_3,P_1\right\}$, $b\neq0$ | $\left(A+\left\{i(t-z)-b\left[y+\frac{(t-z)^2}{2}\right]\right\}B,B\right)$ |
| $\widetilde{P}_{8,11}$ | $\{K_3+aP_2,A_2,P_0+P_3,P_3\}$, $a>0$ | $\left[\xi^{-1}A-\frac{i}{2}(y+a\ln\xi^2)\xi^{-1}B,\xi B\right]$ |
| $\widetilde{P}_{7,7}$ | $\{K_3+aP_1,A_1,A_2,P_0+P_3\}$, $a>0$ | $[\xi^{-1}A+\frac{1}{2}(x+a\ln\xi^2-iy)\xi^{-1}B,\xi B]$ |
| $\widetilde{P}_{8,13}$ | $\left\{K_3+aP_1,A_2,P_0+P_3,P_2-\frac{1}{b}P_1\right\}$, $a>0$, $b\neq0$ | $\left[\xi^{-1}A-\frac{i}{2}(bx+y+ab\ln\xi^2)\xi^{-1}B,\xi B\right]$ |
| $P_{5,3}$ | $\{L_f,A_1,A_2,P_0+P_3\}$ | $[\xi^{-1-i\cot f}A+\frac{1}{2}(x-iy)\xi^{-1+i\cot f}B,\xi^{1+i\cot f}B]$ |

[a] $\xi=[(t-z)/2]^{1/2}$; $A$, $B$ = arbitrary constants.

in finite form, and

$$\mathscr{D}\Psi^{a_1\cdots a_p\dot{b}_1\cdots\dot{b}_q}=\sum_k\left[\tfrac{1}{2}(\Omega\cdot\sigma)^{a_kc_k}\Psi^{a_1\cdots c_k\cdots a_p\dot{b}_1\cdots\dot{b}_q}+\tfrac{1}{2}(\Omega^*\cdot\sigma^*)^{\dot{b}_k\dot{d}_k}\Psi^{a_1\cdots a_p\dot{b}_1\cdots\dot{d}_k\cdots\dot{b}_q}\right] \tag{2.28}$$

in infinitesimal form.

## 3. $(\frac{1}{2},0)$-SPINOR FIELDS INVARIANT UNDER POINCARÉ SUBGROUPS

According to the subgroup classifications[2,4] with respect to the Poincaré group, there are, up to a conjugation, 11 subgroups of dimension $n=6$, 13 with $n=5$, 39 with $n=4$. In Tables I–III, these are all listed in the notation of Ref. 2 and a representative of each conjugacy class of subalgebras is identified by its basis elements. We mention neither the subgroups of dimension $n>6$ for reasons which will become evident below, nor those of dimension $n\leqslant3$ for brevity.

Before the determination of the nontrivial $(\frac{1}{2},0)$-spinor fields invariant under the subgroups of P(3,1), let us distinguish between two kinds of subgroups of P(3,1):

(a) Those leading to nontrivial invariant $(\frac{1}{2},0)$ spinors will be called "SPIPS" (Spinorial Poincaré Subgroup): a SPIPS $G$ is a subgroup of P(3,1) which leaves invariant at least one $(\frac{1}{2},0)$-spinor field $\Psi$ ($\Psi\neq0$) whose components are (real or complex) functions defined and differentiable on $M$ (or at least on the *generic* strata);

(b) the remaining subgroups of P(3,1) will be called "NON-SPIPS".

Applying the methods described in Sec. 2 leads to the following results: (1) The uniform and constant spinor field $\Psi=(\Psi_1,\Psi_2)=(A,B)$ has the group generated by

$$K\equiv\{A_1,A_2,P_\mu\} \tag{3.1}$$

(up to a conjugation) as stabilizer. This Poincaré subgroup of dimensions 6 will be called the "kinematical" group[7] of $\Psi$. (2) Every SPIPS is of dimension less than or equal to 6.

(3) There are five SPIPS of dimension 6 and 7 of dimension 5 and 27 of dimension 4. These are specified in Tables II ($n = 6,5$) and III ($n = 4$). The explicit invariant spinor fields are also given. The NON-SPIPS of dimension 4, 5, and 6 are listed in Table I.

We now prove points (1) and (2) and illustrate one example of a SPIPS determined by the global method and one by the infinitesimal method.

Statement (1) follows directly from the global or the infinitesimal analyses. A basis can always be chosen in Minkowski space, so that the constant $\Psi_2$ vanishes. Then, from the representations (2.17) and (2.18) of the generators, it is obvious that $K$, and nothing else, stabilizes $\Psi$. By the infinitesimal method, we immediately get from Eq. (2.22)

$$(\boldsymbol{\Omega}\cdot\boldsymbol{\sigma})\Psi = (\boldsymbol{\phi} + i\boldsymbol{\theta})\cdot\boldsymbol{\sigma}\Psi = 0.$$

If $\Psi_2$ vanishes, this gives

$$\Omega^3 = 0, \quad \Omega^1 = -i\Omega^2,$$

or

$$\phi^3 = \theta^3 = 0 \quad \text{and} \quad \phi^1 = \theta^2, \quad \theta^1 = -\phi^2.$$

So the vector field $X$ defined in Eq. (2.8) reduces in this case to

$$X = \phi^1 A_1 + \phi^2 A_2 + \alpha^\mu P_\mu , \tag{3.2}$$

which is the general element in $K$.

Statement (2) follows from an argument similar to that used by Combe–Sorba[8] for the case of skew tensors. We may assume that the origin in $M$ lies on a regular orbit since, if it did not, we could take another representative of the same conjugacy class of subgroups for which it does. At the origin, the isotropy group $G(0)$ is contained in the homogeneous Lorentz subgroup, and by the above result must be conjugate to a subgroup generated by $\{A_1, A_2\}$. Since $\dim G(0) \leqslant 2$, and the orbit, which is diffeomorphic to $G/G(0)$, has dimension $\leqslant 4$, we must have $\dim G \leqslant 6$.

Let us now consider the example of the algebra

$$\widetilde{P}_{7,7} \equiv \{K_3 + aP_1, A_1, A_2, P_0 + P_3\} \quad (a > 0) \tag{3.3}$$

from Table III. In order to apply the *global* method, we must identify the group to which it corresponds. We take this to be the subgroup $\widetilde{G}_{7,7} \subset \mathrm{SL}(2,\mathbb{C}) \times \mathbb{H}(2)$ of the Poincaré spinor group consisting of elements $\{U,\tilde{a}\}$ of the form

$$\widetilde{G}_{7,7} = \left\{ U(\alpha,\beta,\gamma) = \begin{pmatrix} \exp\left(\frac{\alpha}{2}\right) & -\gamma - i\beta \\ 0 & \exp\left(-\frac{\alpha}{2}\right) \end{pmatrix}, \right.$$

$$\left. \tilde{a}(\alpha,\lambda) = \begin{pmatrix} 2\lambda & a\alpha \\ a\alpha & 0 \end{pmatrix} \right\}, \quad \alpha,\beta,\gamma,\lambda \in R. \tag{3.4}$$

*i. Generic strata (regular orbits)*

Applying $\{U(\alpha,\beta,\gamma),\tilde{a}(\alpha,\lambda)\}$ to the point $p_0 \equiv (1,0,0,-1)$ maps it to the point with coordinates

$$t + z = 2\lambda + 2\beta^2 + 2\gamma^2, \quad t - z = 2\exp(-\alpha)$$

$$x = a\alpha - 2\gamma \exp\left(-\frac{\alpha}{2}\right),$$

$$y = 2\beta \exp\left(-\frac{\alpha}{2}\right).$$

The orbit is therefore the open submanifold $t - z > 0$. Similarly, the orbit of $(-1,0,0,1)$ is the submanifold $t - z < 0$. The isotropy group for both these points is the identity element; therefore, their orbits belong to the same stratum. Their union is an open dense submanifold. Therefore, any other orbit is of smaller dimension and belongs to a different stratum. Thus, the union of these two orbits forms the single, generic stratum.

*ii. Singular strata*

Applying $\{U(\alpha,\beta,\gamma),\tilde{a}(\alpha,\lambda)\}$ to $p_0(y_0) \equiv (0,0,y_0,0)$ gives $p = (t,x,y,z)$ with

$$t = z = \lambda - \beta y_0 \exp\left(\frac{\alpha}{2}\right),$$

$$x = a\alpha,$$

$$y = y_0.$$

This defines a family of orbits, parametrized by $y_0$, with $t - z = 0, y = y_0$. The isotropy groups $G_0(y_0)$ are given by

$$G_0(y_0) = \left\{ \begin{pmatrix} 1 & -\gamma - i\beta \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2\beta y_0 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$

For $y_0 \neq 0$ these are all conjugate in $\widetilde{G}_{7,7}$, and therefore the union of such orbits forms a stratum. For $y_0 = 0$, however, the isotropy group belongs to a different conjugacy class (in $\widetilde{G}_{7,7}$), and therefore the orbit of the origin is itself a singular stratum.

*iii. Invariant fields*

Since the isotropy group at $(1,0,0,-1)$ is the identity, there is no isotropy constraint and we may take the field to have any value

$$\Psi(1,0,0,-1) = \begin{pmatrix} A \\ B \end{pmatrix}.$$

Applying the group transformation $\{U(\alpha,\beta,\gamma),\tilde{a}(\alpha,\lambda)\}$, the invariance condition (2.21) gives

$$\Psi(t,x,y,z) = \begin{pmatrix} \xi^{-1}A + \frac{1}{2}[x + a\ln\xi^2 - iy]\xi^{-1}B \\ \xi B \end{pmatrix}, \tag{3.5}$$

where $\xi = [(t-z)/2]^{1/2}$. This expression is valid on the orbit $t - z > 0$. Similarly, starting from $(-1,0,0,1)$, we define

$$\Psi(-1,0,0,1) = \begin{pmatrix} A' \\ B' \end{pmatrix}$$

and, applying the group transformation, obtain

$$\Psi(t,x,y,z) = \begin{pmatrix} \eta^{-1}A' + (x + \ln\eta^2 - iy)\eta^{-1}B' \\ \eta B' \end{pmatrix}, \tag{3.5'}$$

where $\eta = [(z-t)/2]^{1/2}$. This is valid for the orbit $t - z < 0$. The values of the arbitrary constants $(A,B)$ and $(A',B')$ are *a priori* unrelated. However, we may relate them by adding some reasonable further assumption, such as analyticity, which implies that the expression (3.5) holds

throughout both regions $x - z < 0$ and $x - z > 0$, with a suitable Riemann sheet structure chosen for $[(t-z)/2]^{1/2}$ and $\ln[(t-z)/2]$. In Table III we have only listed the expression (3.5). For this, as for every other case listed in which a singularity occurs dividing the space into disjoint regions, it is to be understood that the arbitrary constants $(A,B)$ and functions $S(u)$ may be chosen differently in the various regions.

For the singular orbits, the isotropy constraints at the points $(0,0,y_0,0)$ imply that the field $\Psi_S$ takes the form

$$\Psi_S(0,0,y_0,0) = \begin{pmatrix} S(y_0) \\ 0 \end{pmatrix}.$$

Applying the group transformation $\{U(\alpha,\beta,\gamma),\tilde{a}(\alpha,\lambda)\}$, the invariance condition (2.21) gives

$$\Psi_S(t,x,y,z) = \begin{pmatrix} \exp\left(\frac{x}{2a}\right) S(y) \\ 0 \end{pmatrix} \tag{3.6}$$

for $t - z = 0$. A smooth cross section $(0,0,y,0)$ to the singular orbits exists, and we may pick $S(y)$ to be a smooth function. However, there is no way to extend these fields off the singular submanifold $t - z = 0$ without breaking the invariance. It is precisely on this submanifold that the singularities of $\Psi$ occur.

We now apply the *infinitesimal* method to the subgroup of dimension 4 generated by

$$\widetilde{P}_{8,12} \equiv \{K_3 + aP_1, A_2, P_0 + P_3, P_2\} \quad (a > 0). \tag{3.7}$$

Invariance under $P_2$ $[\alpha = (0,0,1,0), \phi = \theta = 0]$ and $P_0 + P_3$ $[\alpha = (1,0,0,1), \phi = \theta = 0]$ implies through Eq. (2.23) that the spinor components are of the form

$$\Psi^i = \Psi^i(t-z,x). \tag{3.8}$$

Invariance under $A_2 \equiv L_1 + K_2$ $[\alpha = 0, \phi = (0,1,0), \theta = (-1,0,0)]$ gives

$$\left[y\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial z}\right) + (t-z)\frac{\partial}{\partial y}\right]\Psi^1 = -i\Psi^2,$$
$$\left[y\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial z}\right) + (t-z)\frac{\partial}{\partial y}\right]\Psi^2 = 0, \tag{3.9}$$

which, substituting in Eq. (3.8) gives $\Psi^2 = 0$.

Finally, invariance under $K_3 + aP_1$ $[\alpha = (0,1,0,0), \phi = (0,0,1), \theta = 0]$ leads to

$$\left(a\frac{\partial}{\partial x} + z\frac{\partial}{\partial t} + t\frac{\partial}{\partial z}\right)\Psi^1 = \tfrac{1}{2}\Psi^1.$$

Introducing the new variable $\rho = t - z$ and using Eq. (3.8) gives

$$a\frac{\partial}{\partial x}\Psi^1(x,\rho) = \frac{\partial\Psi^1(x,\rho)}{\partial(\ln\rho)} + \tfrac{1}{2}\Psi^1(x,\rho).$$

Changing variables to $u = (x/a) + \ln\rho$ and $v = (x/a) - \ln\rho$ reduces the equation to the form

$$\frac{\partial\Psi^1(u,v)}{\partial v} = \tfrac{1}{4}\Psi^1(u,v).$$

The final solution is then

$$\Psi^1(t-z,x) = e^{x/4a}(t-z)^{-1/4}F[(x/a) + \ln(t-z)]$$

or equivalently

$$\Psi^1(t-z,x) = (t-z)^{-1/2}G[(x/a) + \ln(t-z)].$$

The invariant spinor field thus takes the general form

$$\Psi = \begin{pmatrix} (t-z)^{-1/2}G[(x/a) + \ln(t-z)] \\ 0 \end{pmatrix} = \begin{pmatrix} (t-z)^{-1/2}G[(t-z)e^{x/a}] \\ 0 \end{pmatrix}. \tag{3.10}$$

This result can evidently also be obtained through the global method. Since the solutions to these differential equations must be interpreted in a local sense, the comments above regarding singularities and disjoint regions apply equally.

All the results concerning invariant fields are collected in Table II and III, where we have introduced the following notations: $\xi = [(t-z)/2]^{1/2}$, $A$, $B$, are arbitrary constants; $S(u)$ is an arbitrary function of the specified variables $u$; and $L_f = L_3 \cos f - K_3 \sin f$ when $0 < f < \pi, f \neq \pi/2$.

Many of the groups appearing in our tables are subgroups of others, and thus obviously have associated invariant fields (e.g., $P_{10,1} \supset P_{14,1}, \widetilde{P}_{10,6}$; $P_{7,2} \supset P_{7,3}, P_{8,2}, \widetilde{P}_{7,6}$, etc.). Our motivation in including these nevertheless is that usually the invariant fields associated to the subgroups are of more general form.

## 4. INVARIANCE OF SPINOR FIELDS UNDER CONFORMAL TRANSFORMATIONS

As already mentionned, BHPW have determined the different tensors invariant under the maximal subgroups of the conformal group.[23-25] Here we address ourselves to the corresponding problem with respect to $(\frac{1}{2},0)$-spinor fields and we limit ourselves to the *nine* maximal subgroups[3] up to conjugacy under *conformal* transformations. Conjugacy classes under *Poincaré* transformations are obtained by applying the transition elements listed in Table I of BHPW to the given representative.

If conformal transformations are considered, the basis (2.2) has to be completed by

$$D = x^\mu\partial_\mu, \quad C_\mu = x^2\partial_\mu - 2x_\mu x^\nu\partial_\nu \tag{4.1}$$

associated with dilations $D$ and special conformal transformations $C$. The algebra (2.3) is supplemented by the following nonzero commutators:

$$[D,P_\mu] = -P_\mu, \quad [D,C_\mu] = C_\mu,$$
$$[M_{\mu\nu},C_\sigma] = g_{\mu\sigma}C_\nu - g_{\nu\sigma}C_\mu, \tag{4.2}$$
$$[P_\mu,C_\nu] = 2g_{\mu\nu}D + 2M_{\mu\nu}.$$

The vector field $X \in c(3,1)$ associated with the *infinitesimal* transformations

$$x^\mu \to x'^\mu = x^\mu - \omega^\mu{}_\nu x^\nu + \alpha^\mu - \lambda x^\mu - c^\mu x^2 + 2(c\cdot x)x^\mu \tag{4.3}$$

are now defined in correspondence with Eq. (2.4) or (2.8), by

$$X = \phi\cdot\mathbf{K} - \theta\cdot\mathbf{L} + \alpha\cdot P + \lambda\cdot D + c\cdot C. \tag{4.4}$$

Following Mack and Salam,[24] we shall regard spinors as induced representations of the conformal group, such that the isotropy group at the origin, generated by $\{M_{\mu\nu},D,C_\mu\}$, acts upon $\Psi(0)$ through the $(\frac{1}{2},0)$ representation of $SL(2,\mathbb{C})$, has canonical scaling dimension $d = \frac{3}{2}$, and is trivial when restricted to the special conformal transformations. Under a

finite conformal transformation $g: x \mapsto x'_g \equiv f_g(x)$, the Jacobian matrix has the following form:

$$ {}^g J^\mu{}_\nu(x) \equiv \frac{\partial x'^\mu_g}{\partial x^\nu} = {}^g \Lambda^\mu{}_\nu(x) \exp[\lambda_g(x)], $$

where ${}^g \Lambda^\mu{}_\nu(x) \in SO(3,1)$, $\lambda_g(x) \in \mathbb{R}$.

The corresponding transformation of $\Psi$ is

$$ g: \Psi(x) \mapsto \Psi'(x) \equiv \exp[\tfrac{3}{2}\lambda_g(x)] U_g(x) \Psi[f_g^{-1}(x)], $$

where $U_g(x) \in SL(2,C)$ is determined, up to a sign, as the element corresponding to ${}^g \Lambda^\mu{}_\nu(x)$ under the homomorphism $SL(2,C) \to SO(3,1)$. The invariance condition thus becomes

$$ \Psi(x'_g) = \exp[\tfrac{3}{2}\lambda_g(x)] U_g(x) \Psi(x). \tag{4.5} $$

For finite transformations, an arbitrariness of sign must be retained [or the transformation group reidentified as a subgroup of SU(2,2)], but for infinitesimal ones, the sign may be fixed by requiring continuity in a neighborhood of the identity transformation. The resulting invariance condition, corresponding to the infinitesimal transformation (4.3), is given by

$$ (\tilde{\mathscr{D}} + \tfrac{3}{2}\tilde{\lambda})\Psi = \tfrac{1}{2}(\tilde{\boldsymbol{\Omega}}\cdot\boldsymbol{\sigma})\Psi, \tag{4.6} $$

where

$$ \tilde{\mathscr{D}} \equiv \mathscr{D} - \lambda x^\nu \partial_\nu + 2(c\cdot x)(x\cdot\nabla) - x^2(c\cdot\nabla) \tag{4.7a} $$

$$ = \alpha^\mu \partial_\mu + (\mathbf{x}\cdot\boldsymbol{\phi})\frac{\partial}{\partial t} + (t\,\boldsymbol{\phi} + \mathbf{x}\times\boldsymbol{\theta})\cdot\frac{\partial}{\partial \mathbf{x}} + 2(c^0 t - \mathbf{c}\cdot\mathbf{x})\left(t\frac{\partial}{\partial t} + \mathbf{x}\cdot\frac{\partial}{\partial \mathbf{x}}\right) - \lambda\left(t\frac{\partial}{\partial t} + \mathbf{x}\cdot\frac{\partial}{\partial \mathbf{x}}\right) - x^2\left(c^0\frac{\partial}{\partial t} + \mathbf{c}\cdot\frac{\partial}{\partial \mathbf{x}}\right), \tag{4.7b} $$

$$ \tilde{\lambda} \equiv \lambda + \lambda(x) = \lambda - 2c^\nu x_\nu, \tag{4.8} $$

$$ \tilde{\boldsymbol{\Omega}} \equiv \tilde{\boldsymbol{\phi}} + i\tilde{\boldsymbol{\theta}}, \tag{4.9a} $$

with

$$ \tilde{\boldsymbol{\phi}} = \boldsymbol{\phi} + \boldsymbol{\phi}(x) = \boldsymbol{\phi} + 2c^0\mathbf{x} - 2t\mathbf{c}, \tag{4.9b} $$

$$ \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \boldsymbol{\theta}(x) = \boldsymbol{\theta} + 4\mathbf{c}\times\mathbf{x}, \tag{4.9c} $$

$$ \tilde{\boldsymbol{\Omega}}\cdot\boldsymbol{\sigma} = \begin{pmatrix} \tilde{\phi}^3 + i\tilde{\theta}^3 & \tilde{\phi}^1 + \tilde{\theta}^2 + i(\tilde{\theta}^1 - \tilde{\phi}^2) \\ \tilde{\phi}^1 - \tilde{\theta}^2 + i(\tilde{\theta}^1 + \tilde{\phi}^2) & -(\tilde{\phi}^3 + i\tilde{\theta}^3) \end{pmatrix}. \tag{4.10} $$

These expressions are the spinor analogs of those introduced in BHPW for the case of tensor fields (see Sec. 12 of Ref. 13).

Equations (4.5) or (4.6) may now be applied to the determination of spinor fields invariant under subgroups of the conformal group C(3,1). Here, we shall only discuss its maximal subgroups, for which we have the following result: There are no nonvanishing spinor fields invariant under any of the maximal subgroups of the conformal group.

There are nine conjugacy classes of maximal subgroups of C(3,1) which, in the notation of Ref. 13, are SIM(3,1), OPT(3,1), O(3,2), O(4,1), O(2) ⊗ O(2,2), O(3) ⊗ O(2,1), O(2,1) ⊗ O(2,1), O(2) ⊗ O(4), and S[U(2,1) ⊗ U(1)]. Each class contains a representative for which the origin is on the generic stratum and all the isotropy groups for these representatives contain the rotations generated by $L_3$. However, no nonvanishing spinor field is invariant under such rotations, as can be seen immediately from the isotropy conditions at the origin. Actually, for the groups SIM(3,1), OPT(3,1), O(3,2), O(4,1), O(2) ⊗ O(2,2), and O(3) ⊗ O(2,1), the Poincaré part is already contained in our list of NON-SPIPS of dimension ⩾4 (Table I), as may be seen by consulting the list of generators given in the table of BHPW.

## 5. COMMENTS

### A. Four components invariant spinors

It is clear from Eq. (2.25) that if a $(\frac{1}{2},0)$ spinor $\Psi$ is invariant under a Poincaré subgroup, its complex conjugate $\bar{\Psi}$ will also be an invariant field, regarded as transforming under the conjugate $(0,\frac{1}{2})$ representation, and conversely. Therefore, all invariant $(0,\frac{1}{2})$ fields are complex conjugates of invariant $(\frac{1}{2},0)$ fields. This result is equally clear from the infinitesimal invariance condition (2.26) for $(0,\frac{1}{2})$ fields, which is obtained from the corresponding one (2.22) for $(\frac{1}{2},0)$ fields by replacing

$$ \rho(\phi^i K_i - \theta^i L_i) = \tfrac{1}{2}\phi^i\sigma_i + \tfrac{i}{2}\theta^i\sigma_i = \tfrac{1}{2}(\boldsymbol{\Omega}\cdot\boldsymbol{\sigma}) \tag{5.1} $$

by its complex conjugate

$$ \bar{\rho}(\phi^i K_i - \theta^i L_i) = \tfrac{1}{2}\phi^i\sigma_i^* - \tfrac{i}{2}\theta^i\sigma_i^* = \tfrac{1}{2}(\boldsymbol{\Omega}^*\cdot\boldsymbol{\sigma}^*). \tag{5.2} $$

Four component Dirac spinors transform according to the direct sum representation[26]

$$ D^{(1/2,0)} \oplus D^{(0,1/2)} $$

and therefore are determined by a pair $(\Psi^a, \Phi^{\dot{b}})$ of 2-component spinors. The condition for invariance of the Dirac spinor is that both $\Psi$ and $\Phi$ be invariant. In the basis corresponding to the direct sum decomposition

$$ R(X) = \rho(X) \oplus \bar{\rho}(X), \tag{5.3} $$

we have

$$ R(\boldsymbol{\phi}\cdot\mathbf{K} - \boldsymbol{\theta}\cdot\mathbf{L}) = \tfrac{1}{2}\phi^i\hat{\sigma}_i + \tfrac{1}{2}\theta^i\hat{\tau}_i, \tag{5.4} $$

with

$$ \hat{\sigma}_i = \begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i^* \end{pmatrix}, \quad \hat{\tau}_i = \begin{pmatrix} \sigma_i & 0 \\ 0 & -\sigma_i^* \end{pmatrix}. \tag{5.5} $$

In order to express an invariant Dirac field $\binom{\Psi}{\Phi}$ in the standard basis, we must apply the matrix

$$ M = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & \sigma_2 \\ 1 & -\sigma_2 \end{pmatrix}. \tag{5.6} $$

which relates our representation (5.5) to the usual one as given for example by Bjorken and Drell[27]

$$ \sigma_i^{BD} = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \quad \tau_i^{BD} = \begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix}. \tag{5.7} $$

It follows from the results of Sec. 4 that there exists no nontrivial four components spinor invariant under any of the maximal subgroups of the conformal group C(3,1). In particular, there are no nontrivial "Dirac" spinors invariant under O(3,2), O(4,1), or O(2) ⊗ O(4), although invariant scalar densities and gauge fields under these groups are known to exist and have interesting properties.[28,29]

### B. Spinors and tensors

Let us end this section by discussing the connections between invariant spinors and tensors. This point is of particular interest owing to the fact that skew-symmetric tensors of rank 2 (2 forms) have already been studied in connection with the Poincaré subgroups[8,9] and the maximal subgroups of the conformal group.[13]

It is well known[21,22] that tensors may be constructed from spinors or re-expressed in a spinor basis through the connecting quantities $\sigma_\mu^{a\dot{b}}$ introduced by Infeld and Van der Waarden.[30] Such quantities transform as covariant 4 vectors (indices $\mu = 0,1,2,3$) with respect to Lorentz transformations and as second rank spinors (indices $a,\dot{b} = 1,2$) with respect to spinor transformations. The net result of the two laws of transformation is that the $\sigma_\mu^{a\dot{b}}$ are invariant. Thus, by contraction with $\sigma_\mu^{a\dot{a}}$ or $\sigma_{a\dot{a}}^\mu$, to every $\mu$-index tensor is associated an $a\dot{a}$-spinor, and conversely. In particular, we have

$$F_{\mu\nu} \leftrightarrow F_{a\dot{a}b\dot{b}}\ , \tag{5.8}$$

i.e.,

$$F_{\mu\nu} = \sigma_\mu^{a\dot{a}}\, \sigma_\nu^{b\dot{b}}\, F_{a\dot{a}b\dot{b}} \tag{5.8a}$$

*and*

$$F_{a\dot{a}b\dot{b}} = \sigma_{a\dot{a}}^\mu\, \sigma_{b\dot{b}}^\nu\, F_{\mu\nu}\ . \tag{5.8b}$$

Moreover,[31] if $F$ is a *real*, skew-symmetric 2-index tensor, for example an "electromagnetic" field, we have

$$F_{a\dot{a}b\dot{b}} = \epsilon_{ab}\bar{\phi}_{\dot{a}\dot{b}} + \epsilon_{\dot{a}\dot{b}}\phi_{ab}\ , \tag{5.9a}$$

where

$$\phi_{ab} \equiv \tfrac{1}{2}\epsilon^{\dot{a}\dot{b}}F_{ab\dot{a}\dot{b}} \tag{5.9b}$$

is a *symmetric* 2 spinor, $\bar{\phi}_{\dot{a}\dot{b}}$ its complex conjugate, and

$$\epsilon^{ab} = \epsilon_{\dot{a}\dot{b}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -\epsilon^{\dot{a}\dot{b}} = -\epsilon^{ab}. \tag{5.10}$$

(Raising and lowering of spinor indices is by contraction with $\epsilon_{\dot{a}\dot{b}}$ or $\epsilon^{\dot{a}\dot{b}}$.)

It is clear from Eqs. (5.8) and (5.9) that if we have a spinor $\phi_{ab}$, invariant under a Poincaré subgroup, the corresponding tensor is also invariant owing to the invariance of the $\sigma_\mu^{a\dot{b}}$, $\epsilon_{ab}$, and $\epsilon_{\dot{a}\dot{b}}$ and conversely. Starting from an invariant $(\frac{1}{2},0)$ spinor $\Psi$ we may thus form an invariant skew tensor by defining

$$F_{a\dot{a}b\dot{b}} = \epsilon_{ab}\bar{\Psi}_{\dot{a}}\bar{\Psi}_{\dot{b}} + \epsilon_{\dot{a}\dot{b}}\Psi_a\Psi_b \tag{5.11}$$

which is nonvanishing if $\Psi_a$ is. [The converse is not necessarily possible, since not all tensors admit a decomposition of the form (5.11).]

We see that for all the SPIPS of Secs. 2 and 3 there exists an invariant, skew-symmetric tensor. For example, the subgroup generated by $\tilde{P}_{8,10}$ admits the nontrivial invariant spinor field (see Table II)

$$\Psi = (A\exp(ax/2),0). \tag{5.12}$$

From Eqs. (5.8) and (5.11), we find that the associated skew-symmetric tensor, expressed in terms of electric and magnetic components, is[32]

$$\mathbf{E} = (A\exp(ax), B\exp(ax), 0),$$

$$\mathbf{B} = (-B\exp(ax), A\exp(ax), 0), \tag{5.13}$$

which can also be deduced directly from Eqs. (2.24). However, this field diverges at infinity and does not satisfy the free Maxwell equations.

As a final comment, it is important to emphasize that the results obtained here relate to *strict* invariance of spinor fields, whereas the physically relevant criterion may only be invariant up to a *gauge* (i.e., phase) *transformation*, particularly in problems involving coupling to gauge fields possessing similar invariance properties. Such a criterion can only be applied, however, after a classification of the admissible gauge transformations corresponding to the given transformation group action has been obtained.[33]

It would be an interesting problem, for example, to use such symmetry requirements to simplify the equations coupling a spinor field to gauge fields assuming the spinor transforms nontrivially under the gauge group. Such problems will be studied in future work.

J. Beckers and P. Jasselette, with hearty and thankful feelings, dedicate the present work to Professor J. Serpe at the occasion of his retirement.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378, 1932 (1974); **16**, 1613 (1975); **17**, 717 (1976).
[2]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).
[3]J. Patera, P. Winternitz, and H. Zassenhaus, Lecture Notes in Physics, Vol. 79, Group Theoretical Methods in Physics, Tübingen 1977, p. 388.
[4]H. Bacry, Ph. Combe, and P. Sorba, Rep. Math. Phys. **5**, 145, 361 (1974).
[5]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[6]G. Burdet, J. Patera, M. Perrin, and P. Winternitz, J. Math. Phys. **19**, 1758 (1978).
[7]H. Bacry, Ph. Combe, and J.L. Richard, Nuovo Cimento A **67**, 267 (1970).
[8]Ph. Combe and P. Sorba, Physica (Utrecht) **80A**, 271 (1975).
[9]J. Beckers and G. Comte, Bull. Soc. R. Sc. Liége **45**, 279 (1976).
[10]C.P. Boyer, R.T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).
[11]J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. **18**, 72 (1977).
[12]J. Beckers and G. Comte (to be published, 1980).
[13]J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[14]J. Beckers and M. Jaminon, Physica (Utrecht) **A94**, 165 (1978).
[15]J. Beckers, M. Jaminon, and J. Serpe, J. Phys. A **12**, 1341 (1979).
[16]J. Harnad and L. Vinet, Phys. Lett. **B76**, 589 (1978).
[17]J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979).
[18]J.M. Levy-Leblond, Commun. Math. Phys. **12**, 64 (1969).
[19]Ph. Combe and J.L. Richard, Physica (Utrecht) **71**, 66 (1974).
[20]E. Cartan, *Leçons sur la Théorie des Spineurs* (Hermann, Paris, 1938).
[21]W.L. Bade and H. Jehle, Rev. Mod. Phys. **3**, 714 (1953).
[22]R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).
[23]H.A. Kastrup, Ann. Phys. (Leipzig) **9**, 388 (1962).
[24]G. Mack and A. Salam, Ann. Phys. (N.Y.) **53**, 174 (1969).
[25]S. Ferrara, R. Gatto, and A.F. Grillo, *Conformal Algebra in Space–Time*, Springer Tracts in Modern Physics (Springer, Berlin, 1973), Vol. 67.
[26]S. Weinberg, Phys. Rev. B **133**, 1318 (1964).
[27]J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw Hill, New York, 1964).
[28]S. Fubini, Nuovo Cimento A **34**, 521 (1976).
[29]V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **65**, 163 (1976); B **72**, 203 (1977).
[30]L. Infeld and B.L. Van der Waarden, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl., 1380 (1933).

[31]F.A.E. Pirani, "Introduction to Gravitational Radiation Theory", in *Lectures on General Relativity* (Brandeis, Prentice Hall, New Jersey, 1964), Vol. 1, Sec. 3, and references therein.

[32]In Ref. 8, Combe and Sorba do not list $\widetilde{P}_{8,10}$ among their "CEPS" (Poincaré subgroup leaving a skew-symmetric tensor invariant), contrary to our result (5.13). In brief, *all* SPIPS must be CEPS, though not necessarily the converse.

[33]See, for example, J. Harnad, S. Shnider, and L. Vinet, "Group Actions on Principal Bundles and Invariance Conditions for Gauge Fields" (to appear).

# Oscillators submitted to squared Gaussian processes

Christian Soize
*CTICM, 92807 Puteaux, France*

(Received 11 March 1980; accepted for publication 16 May 1980)

The paper is a study of oscillators governed by equations of the type $a\ddot{X}(t) + bX(t) + cX(t) = E(t)$, where $a$, $b$, $c$ are given constants and where $E(t)$ is for example the square of a Gaussian stationary process. A constructive and numerical method, using explicit expressions of Fourier transforms, are developed in order to compute the density of the distribution function (d.f.) of $X(t)$ and of the joint distribution of $X(t)$ and $\dot{X}(t)$. Hence the upcrossing rates of a given level and an approximation of the d.f. of $\max_{t\in T} X(t)$ can be computed. A numerical example is given.

## 1. INTRODUCTION

The purpose of this paper is to present a constructive approach which allows us to solve some random vibration problems for which the classical methods of the Markov process do not apply. Consider, for instance, a linear oscillator

$$a\ddot{x}(t) + b\dot{x}(t) + cx(t) = E(t), \tag{1}$$

with $t\in\mathbb{R}$; $a,b,c$ given constants of $\mathbb{R}^+{}^*$; and where the excitation is a stationary stochastic process $(E_t)$. Then the unknown solution is a stationary stochastic process, denoted by $(x_t)$. If $(E_t)$ is a Gaussian process, $(x_t)$ is obviously Gaussian also and the problem is a trivial one.

We consider here the case where $(E_t)$ is not Gaussian, but results from a quadratic transform of a Gaussian process which possesses *a priori* any one covariance operator. This problem cannot thus be reduced to a Markov case, and furthermore, the techniques of differential stochastic equations and of diffusion, leading, e.g., to a Fokker–Planck equation, do not apply. Prior to analyzing the general nonlinear transformation under consideration, let us give two particular example pertaining to the general case.

*First example*: Excitation $(E_t)$ is written as

$$E_t = \int_D (p_{11} y_1^2(t,u) + q_1(u) y_1(t,u) + r(u))\, du \tag{2}$$

with $t\in\mathbb{R}$; $u = (u_1,u_2,u_3)\in D\subset\mathbb{R}^3$; $du = du_1\, du_2\, du_3$; $q_1$ and $r$ given mappings: $D\to R$; $p_{11}$ a given real constant and where $y_1(t,u)$ is a stochastic Gaussian process denoted by $(y_{1tu})_{tu}$ with zero mean and stationary in time $t$. This model corresponds, for instance, to the effect of atmospheric turbulence on a deformable structure where the interaction of the fluid and the structure is ignored.[1,2]

*Second example*: This excitation is now a function of the unknown $(\dot{x}_t)$ and is written as

$$E_t = v[\epsilon y_1(t) - \dot{x}(t)]^2 + v'[\epsilon y_1(t) - \dot{x}(t)] + \epsilon r', \tag{3}$$

with $v'\in\mathbb{R}^+{}^*$; $v,r'\in\mathbb{R}^*$; $\epsilon$ a small real parameter and where $y_1(t)$ is a stochastic Gaussian process denoted by $(y_{1t})_t$ with zero mean and stationary in time $t$. This model corresponds to the same problem as in the first example, but taking the fluid-structure interaction into account.[2] The perturbed solution in $\epsilon$ of (1) and (3), restricted to the second-order: $x_\epsilon(t) = \epsilon x_1(t) + \epsilon^2 x_2(t)$ is sought for. All calculations being completed, it is found that $x_\epsilon$ is the solution of

$$a\ddot{x}_\epsilon(t) + b'\dot{x}_\epsilon(t) + cx_\epsilon(t) = \sum_{j,j'=1}^{2} p_{jj'}\, y_j(t)\, y_{j'}(t) + \sum_{j=1}^{2} q_j\, y_j(t) + r, \tag{4}$$

where $b' = b + v'$, $y_2(t) = \dot{x}_1(t)$; $p_{11} = p_{22} = \epsilon^2 v$; $p_{12} = p_{21} = -\epsilon^2 v$; $q_1 = \epsilon v'$; $q_2 = 0$; $r = \epsilon r'$; and where $x_1$ is the stochastic solution of:
$a\ddot{x}_1(t) + b'\dot{x}_1(t) + cx_1(t) = v' y_1(t) + r'$.

In the following consideration, our attention will be focused on a constructive method enabling the numerical calculation of some quantities related to trajectories of the process $(x_t)$. This is useful for the vibration analysis of systems. These quantities are

(i) The probability $D(\alpha,\alpha')$ such that:

$$D(\alpha,\alpha') = \text{Prob}(x_t\leqslant\alpha, \dot{x}_t\leqslant\alpha') = \int_{-\infty}^{\alpha}\int_{-\infty}^{\alpha'} w(a,b)\, da\, db, \tag{5}$$

where $w$ is the density of the joint distribution of $x_t$ and $\dot{x}_t$, which is independent of $t$ because $(x_t)$ and $(\dot{x}_t)$ are stationary processes.

(ii) The mean $\overline{N^+(\alpha,T)} = \mathscr{E}[N^+(\alpha,T)]$ of the random variable $N^+(\alpha,T)$ which represents the number of upcrossing of the level $\alpha\in\mathbb{R}$ by the trajectory of the $(x_t)$ stationary process, in time $T$ ($T$ being a bounded interval of $\mathbb{R}$). Using a classical result of Ref. 3, if $\mathscr{E}(\dot{x}_t^2) < \infty$, we have:

$$\overline{N^+(\alpha,T)} = T\int_0^{+\infty} b w(\alpha,b)\, db \tag{6}$$

(iii) The probability distribution $G$ of the extreme positive values of the random variable $X_{\max} = \max_{t\in T}(x_t)$. We suppose that for the high values $\beta\in\mathbb{R}^+$ of $X_{max}$, the sequence of instants of reaching level $\beta$ by increasing values, tends to become a Poisson process with a rate equal to $T\times[N^+(\alpha,T)]^{-1}$. Hence, for $\beta\to+\infty$

$$G(\beta) = \text{Prob}\Big(\max_{t\in T}(x_t)\leqslant\beta\Big)\simeq\exp\Big(-\overline{N^+(\alpha,T)}\Big) \tag{7}$$

and the mean $\overline{X}_{\max} = \mathscr{E}(X_{\max})$ is simply obtained by the expression

$$\overline{X}_{\max} = \int_0^{+\infty} \beta\, dG(\beta). \tag{8}$$

 0022-2488/80/102500-08$1.00  

One should note that the validity of Poisson assumption for the Gaussian case is proved by the Volkonsky–Rozanov's theorem.[4] Here, $(x_t)$ is not Gaussian and there is no result at the present time concerning this point. Nevertheless, we shall maintain this assumption, which seems to be physically correct.

Finally, we see that the effective computation of (5), (6), (7), and (8) requires us to determine the probability measure $W = w(a,b)\,da\,db$, $(x_t,\dot{x}_t)$ not being a Markov process. For this purpose we shall use a functional approach using the notion of measures on vector spaces.

It is nevertheless necessary to being with the definition of the various transformations between the known process $(y_s)$ and the process $(x_t)$ which will be further under considerations. The two examples which we have already given are particular cases of the transformations hereafter.

## 2. TRANSFORMATIONS UNDER CONSIDERATION

Let $S$ be a set of the space $\mathbb{R}^{n+1}$, with $n\in\mathbb{N}$; the generating point of $S$ is denoted by $s = (t',u_1,...,u_n) = (t',u)$; if $n = 0$, then $s = t'$. We shall denote by $C(S,\mathbb{R}^m)$ the set of all $\mathbb{R}^m$–valued continuous functions defined on $S$ with $m\in\mathbb{N}^*$. For $m = 1$, we shall simply write $C(S)$.

(i) The probabilistic data of the problem is an $m$–dimensional, real stochastic process on $S$: $(y_s) = (y_{1s},...,y_{ms})$ which is assumed to be Gaussian, with zero mean, stationary in time $t'$ and its trajectories are almost surely (a.s) continuous on $S$.

(ii) The process $(y_s)$ is transformed into a one-dimensional real stochastic process $(z_s) = (f(y_s))$ on $S$ by a nonlinear mapping $f$: $C(S,\mathbb{R}^m)\to C(S)$ such that $\forall y\in C(S,\mathbb{R}^m)$

$$y(s)\to z(s) = (fy)(s) = \sum_{jj'=1}^{m} p_{jj'}y_j(s)\,y_{j'}(s) + \sum_{j=1}^{m} q_j(s)\,y_j(s) + r(s), \tag{9}$$

where $(y_1,...,y_m)$ are the coordinates of $y$ on the canonical basis $\{\mathbf{b}_1,\mathbf{b}_2,...,\mathbf{b}_m\}$ of $\mathbb{R}^m$; $p_{jj'}$, is the component on $\mathbf{b}_j\otimes\mathbf{b}_{j'}$, of a symmetrical tensor $p\in\mathbb{R}^m\otimes\mathbb{R}^m$, $q_j$ is the component on $\mathbf{b}_j$ of $q\in C(S,\mathbb{R}^m)$; and $r\in C(S)$. The quantities $p$, $q$, $r$ are deterministic and known. It is assumed that for any fixed $u\in R^n$, $r(t',u)$ and $q(t',u)$ are independent with regard to $t'$ and that $r$ is such that $z_s = f(y_s)$ has a zero mean.

(iii) We consider lastly a linear filter with input $(z_s)$ and output the two-dimensional real process $(x_t,\dot{x}_t)$ on $\mathbb{R}^2$. For any fixed time $t$ or $\mathbb{R}$, this filter is defined by a linear operator $\beta_t: C(S)\to\mathbb{R}^2$. Then, $(x_t,\dot{x}_t) = \beta_t(z_s)$ and for any $z\in C(S)$, $\beta_t$ is such that

$$(x(t),\dot{x}(t)) = \beta_t(z) = \left(\int_S z(s)\,d\mu_1(t,s),\ \int_S z(s)\,d\mu_2(t,s)\right), \tag{10}$$

where $\mu_1(t,.)$ and $\mu_2(t,.)$ are real-valued measures with compact support on $S$, such that for $i\in\{1,2\}$, $du_1(t,s) = g_i(t,s)\,ds$, the density $s\to g_i(t,s)$ being a null function outside the compact subset $K_t$ of $S$ and continuous on $K_t$. It is assumed besides that $\mu_1$ and $\mu_2$ are such that $(x_t,\dot{x}_t)$ is stationary and has a zero mean. Then for the first example, $S = \mathbb{R}\times D$, $n = 3$, $m = 1$, Eqs. [(1) and (2)] let $x_t = \int_{t-\theta}^{t} h(t-t')\times\int_D (fy_1)(t',u)\,dt'\,du$, $\dot{x}_t = \int_{t-\theta}^{t}\dot{h}(t-t')\int_D (fy_1)(t_1,u)\,dt'\,du$, where $t\to h(t)$, is the restriction to $\mathbb{R}^+$ of the impulsional response of (1), and where $\theta\in\mathbb{R}^+$ is such that from numerical point of view, the integration interval $(-\infty,t)$ is replaced by the finite interval $(t-\theta,t)$, the system being damped. We are then led to the formulation (10) by putting $K_t = [t-\theta,t]\times D$; $g_1(t,s) = h(t-t')$ and $g_2(t,s) = \dot{h}(t-t')$ if $s\in K_t$; $g_1(t,s) = g_2(t,s) = 0$ and if $s\notin K_t$. In the same way, a similar formulation will be obtained for the second example.

(iv) We now have the following scheme

$$C(S,\mathbb{R}^m)\xrightarrow{f} C(S)\xrightarrow{\beta_t}\mathbb{R}^2$$

$$(y_s)_s\to(z_s) = (f(y_s))_s\to(x_t,\dot{x}_t) = \beta_t(z_s).$$

Here we are calculating explicitly the probability measure $W = w(a,b)\,da\,db$ of the random variable $(x_t,\dot{x}_t) = \beta_t[f(y_s)]$, $(y_s)$ being the process which is defined in (i) above. $W$ has a zero mean and is independent of the time $t$.

## 3. CONSTRUCTIVE METHOD USING MEASURES ON VECTOR SPACES

We know that a random variable with values in the finite-dimensional vector space $E = \mathbb{R}^n$ is represented by a probability measure $P$ on $\mathbb{R}^n$ equipped with the Borel $\sigma$–field, or, by the Fourier transform $\hat{P}$ of $P$ defined on the dual space $E'\simeq\mathbb{R}^n$ of $E$ and which is written: $\forall v\in E'$, $\hat{P}(v) = \int_{y\in E}\exp(-i\langle y,v\rangle)\,dP(y)$, where $\langle\cdot,\cdot\rangle$ is the bilinear form $E\times E'\to\mathbb{R}$ defining the duality between $E$ and $E'$. The functional method of analysis of a stochastic process consists in applying the same scheme, but space $E$ will be the space of trajectories of the process which will generally be a space of functions or distributions.[5] Thus, to represent a stochastic process $(y_s)$ by a probability measure $P$ on its space of trajectory $E$ equipped with an adequate $\sigma$–field, means that "the trajectories of $(y_s)$ are extracted randomly from $E$ according to the probability law $P$".

Let suppose $E$ be a locally convex Hausdorff space[6] and let be $E'$ its topological dual. The Fourier transform of the probability measure $P$ on $E$ equipped with the Borel $\sigma$–field is given by

$$\forall v\in E',\quad \hat{P}(v) = \int_{y\in E}\exp(-i\langle y,v\rangle)\,dP(y). \tag{11}$$

For $r\in E$, let us denote by $P$ the probability measure on $E$ deduced from the probability measure $P_0$ on $E$, by the translation of intensity $r$. Then the Fourier transform of $P$ is given by

$$\forall v\in E',\quad \hat{P}(v) = \exp(-i\langle r,v\rangle)\hat{P}_0(v). \tag{12}$$

Let $F$ be another locally convex Hausdorff space equipped with the Borel $\sigma$–field and let $F'$ be the topological dual. Let $l: E\to F$ be a continuous and measurable mapping. Then the Fourier transform of $Q = l(P)$ is

$$\forall v\in F',\quad \hat{Q}(v) = \int_{z\in F}\exp(-i\langle z,v\rangle)\,dQ(z) = \int_{y\in E}\exp(-i\langle l(y),v\rangle)\,dP(y). \tag{13}$$

In particular, if $l$ is a linear map, $\widehat{Q}(v) = \widehat{P}(l'v)$, where $l':F' \to E'$ is the transposed mapping of $l$. If $l$ is nonlinear, the calculation of (13) becomes much more difficult.

### A. Application to the calculation of the Fourier transform $\widehat{W}$ of the probability measure W defined in Sec. 2(iv)

We shall need the continuity and measurability properties of the mappings $f$ and $\beta_t$ defined respectively by (9) and (10). The spaces $E = C(S,\mathbb{R}^m)$, $F = C(S)$, and $\mathbb{R}^2$ are thus equipped with the Borel $\sigma$–field and with the topology of the compact convergence; their topological dual is denoted by $E' = M(S,\mathbb{R}^m)$ and $F' = M(S)$, where $E'$ and $F'$ are the set of $\mathbb{R}^m$-valued and $\mathbb{R}$-valued measures, respectively, with compact support on $S$. This topology is defined by the semi-norms $p_{K_i}:\varphi \mapsto p_{K_i}(\varphi) = \sup_{y\in K_i}|\varphi(y)|$, with $i$ describing the family of compact subsets $K_i$ of $S$. Then, $f:E \to F$ and $\beta_t:F \to \mathbb{R}^2$ are continuous and are measurable.

The process $(y_s)$ (defined in Sec. 2(i) defines a Gaussian probability measure $P_0$ on $E$, centered and invariant for temporal translation; $f:E \to F$ being continuous and measurable, $f$ transforms $P_0$ according to the probability measure $Q = f(P_0)$ on $F$ which represents the probability law of the process $(z_s) = f(y_s)$. In the same way, $\beta_t$ transforms $Q$ according to the probability measure $W = \beta_t(Q) = (\beta_t \circ f)(P_0)$ on $\mathbb{R}^2$. $W$ is the probability measure defined in Sec. 2(iv). To determine $W$, we shall decompose the nonlinear mapping $f$.

Let thus $E \odot E$ be the vector subspace of the symmetrical tensors of $E \otimes E$. Introduce the nonlinear mapping $B:E \to (E \odot E) \times E$ such that

$$\forall y \in E,\ y \to B(y) = (y \otimes y, y), \tag{14}$$

and the linear mapping $A:(E \odot E) \times E \to F$ such that $\forall y \otimes y \in E \odot E$, $\forall h \in E$,

$$((y \otimes y)(s,s'),h(s)) \to \Sigma_{jj'=1}^{m}\, p_{jj'}\, y_j(s)\, y_{j'}(s) + \sum_{j=1}^{m} q_j(s) h_j(s). \tag{15}$$

By taking into account (9), $f$ is written

$$\forall y \in E,\ y \to f(y) = (A \circ B)(y) + r. \tag{16}$$

We shall need a locally convex Hausdorff topology on $E \odot E$ such that $A$ and $B$ be continuous. For this purpose, we shall use the topologies $\epsilon$ and $\pi$ on tensor products.[7] Both topologies are defined hereafter.

The topology of $E$ being defined by a family of semi-norms $(p_j)_{j\in J}$, for each $j \in J$ the following semi-norms are defined on $E \otimes E$

$$\varphi \to \pi_j(\varphi) = \inf\left\{\sum_k p_j(\lambda_k) p_j(v_k) \quad \text{with } \varphi = \sum_k \lambda_k \otimes v_k\right\}$$

$$\varphi \to \epsilon_j(\varphi) = \sup\{|\langle \varphi, \lambda' \otimes v' \rangle| \quad \text{with } \lambda' \in V_j^0,\ v' \in V_j^0\},$$

where $V_j^0$ specifies the absolute polar set of $V_j = \{x \in E;\ p_j(x) < 1\}$. The topology $\pi$ (resp. $\epsilon$) on $E \otimes E$ is then defined by the semi-norms $(\pi_j)_{j\in J}$, (resp. $(\epsilon_j)_{j\in J}$). $E \odot E$ is equipped with the induced topology by each of the two topologies, noted by $E \odot_\pi E$ and $E \odot_\epsilon E$ respectively.

Then, the mappings $B:E \to (E \odot_\epsilon E) \times E$ and $A:(E \odot_\epsilon E) \times E \to F$ are continuous and measurable for the Borel $\sigma$–fields. By taking into account (16), the probability measure $Q = f(P_0)$ on $F$ is deduced from the probability measure $A \circ B(P_0)$ on $F$ by the translation of intensity $r \in F$. We write formally

$$Q = A \circ B(P_0) + r. \tag{17}$$

The mappings $A$ and $\beta_t$ being linear and transposable, and considering (11), (12), and (13) we obtain

$$\forall (a',b') \in \mathbb{R}^2,\quad \widehat{W}(a',b') = \widehat{Q}(\beta_t'(a',b')), \tag{18}$$

where $\beta_t':\mathbb{R}^2 \to F'$ is the continuous transposed of $\beta_t$, and $\widehat{Q}$ the Fourier transform of $Q$ which is given by

$$\forall \mu \in F',\quad \widehat{Q}(\mu) = \exp(-i\langle r,\mu \rangle) \times B(P_0)(A'(\mu)), \tag{19}$$

where $A':F' \to (E' \odot E') \times E' \subset (E \odot_\epsilon E)' \times E'$ is the continuous transposed mapping of $A$. There is no difficulty in determining $\beta_t'$ and $A'$; as a consequence, it only remains for us to determine the Fourier transform $\widehat{B(P_0)}(\lambda)$ of probability $B(P_0)$ for any $\lambda \in (E' \odot E') \times E'$. We shall do so in the next section, by considering a general case.

## 4. EXPLICIT EXPRESSION OF $\widehat{B(P)}$

Let $E$ be a locally convex Hausdorff space equipped with the Borel $\sigma$–field. We suppose always that any bounded and closed subset of $E$ is complete (quasi-complete space), and we shall consider only a probability measure $P_0$ having the following regularity property: For any Borel subset $\mathscr{B}$ of $E$, $\forall \epsilon$, $\exists$ a compact $K \subset \mathscr{B}$: $P_0(\beta \setminus K) < \epsilon$. This probability measure is sometimes called Radon measure; in particular, any measure on the $\sigma$–field of a separable metrizable complete space is a Radon measure. Then in order to do easy computation, we need, instead of $E$, a Hilbert space called "the reproducing Hilbert space of $E$".

*Theorem* 1: Let $P_0$ be a centered Radon probability measure on a Borel $\sigma$–field of a quasi-complete real locally convex Hausdorff space $E$, such that for any continuous semi-norm $p$ on $E$

$$\int_{x\in E} [p(x)]^2\, dP_0(x) < \infty. \tag{20}$$

Then there exists a linear continuous operator $C_0:E' \to E$ called the covariance operator of $P_0$ such that

$$\forall u,v \in E',\ \langle C_0 u, v \rangle = \int_{x\in E} \langle x,u \rangle\ \langle x,v \rangle\ dP_0(x). \tag{21}$$

*Proof*: $\forall u \in E'$, the mapping $x \to \langle x,u \rangle x$ is continuous. Then, for any compact subset $K$ of $E$, the weak integral $I_K(u) = \int_K \langle x,u \rangle x\, dP_0(x)$ is defines as an element of $E$. Hence, if $K$ describes the set $\mathscr{K}$ of all compact subsets of $E$, all $I_K(u)$ belongs to a bounded set of $E$. Let us introduce on $\mathscr{K}$ the filter with basis $(B_K)$ where $B_K = \{K' \in \mathscr{K},\ K' \supset K\}$. Then it can be shown that $(I_K(u))$ is the basis of a Cauchy filter on $\overline{B}_K$. But $\overline{B}_K$ is complete; hence the limit exists. Denoting this limit as $C_0 u$, we obtain expression (21).

By definition, a Hilbertian subspace of $E$ is a pair $(H,j)$, where $H$ is a Hilbert space, and $j$ is a linear continuous injection of $H$ into $E$.

*Corollary*: With the notations of Theorem 1, there exists

a Hilbertian subspace $H$ of $E$ such that $C_0 = j\rho j'$, where $j' : E' \to H'$ is the transposed mapping of $j$, and where $\rho^{-1}$ is the Riess isomophism of $H$ into its dual $H'$. This Hilbert space $H$ is called the reproducing Hilbert space of $E$.

*Proof*: $C_0$ is symmetric and positive. Hence the bilinear map

$$[C_0 u, C_0 v] = \langle C_0 u, v \rangle. \tag{22}$$

is a scalar product on the subspace $\mathrm{Im}C_0$ of $E$. The injection $j_c$ of $\mathrm{Im}C_0$ into $E$ is continuous for the weak topology because if $(x_n)_n \to 0$ in $\mathrm{Im}C_0$, then $[x_n, C_0 v] = \langle x_n, v \rangle \to 0$. Hence $j_c$ admits a continuous extension $j$ of the completion $H = \widehat{\mathrm{Im}C_0}$ of $\mathrm{Im}C_0$ to $E'^*$. In fact, $\mathrm{Im}j \subset E$ because any Cauchy sequence in $\mathrm{Im}C_0$ is a Cauchy sequence in the weak space $E$ and is bounded in $E$. $j$ is injective because of the continuity of Eq. (22). By an argument of topology, $j$ is continuous for the strong topology. We have $\forall x \in H$, $\forall v \in E'$, $[x, C_0 v] = \langle jx, v \rangle$. Taking in particular $x = \rho j'(u)$ with $u \in E'$, we obtain $\langle j\rho j'(u), v \rangle = [\rho j'(u), C_0 v] = \langle j'u, C_0 v \rangle_{H' \times H}$ $= \langle u, jC_0 v \rangle = \langle u, C_0 v \rangle$; hence $j\rho j' = C_0$.

*Remarks*: 1) If $C_0$ is injective, then $j'$ is injective; 2) if $P_0$ is a Gaussian measure and if $E$ is a Banach space, this corollary gives the result of Ref. 8; 3) if $E$ is metrizable, it may be proved as in the Banach case that the reproducing Hilbert space $H$ is separable; 4) if $P_0$ is a Gaussian measure, the condition (20) is automatically satisfied by the integrability theorem of X. Fernique.[9]

Let us apply this result to determine $\widehat{B(P)}$, where $B$ is the nonlinear mapping as previously defined, and where $P$ denotes the Gaussian probability measure on $E$, deduced from the centered Gaussian probability measure $P_0$ on $E$ by the translation of intensity $jm_H$, for any $m_H \in H$. The covariance operator $C_0$ of $P$ is given by (21). The continuous injection $j$ maps $H$ into $E$, hence the map $k = (j \otimes j, j)$ is continuous and maps $(H \odot_\epsilon H) \times H$ into $G = (E \odot_\epsilon E) \times E$. But $H$ being an Hilbert space, the dual of $H \odot_\epsilon H$ is $H \widehat{\odot}_\pi H$, where we have identified $H$ with its dual. Hence the transposed mapping $k' = (j' \otimes j') \times j'$ of $k$ maps $G'$ on $(H \widehat{\odot}_\pi H) \times H$; we have the following scheme:

$$(H \odot_\epsilon H) \times H \underset{k = (j \otimes j, j)}{\to} G = (E \odot_\epsilon E) \times E \overset{B}{\leftarrow} E$$

$$(H \widehat{\odot}_\pi H) \times H \underset{k' = (j' \otimes j', j')}{\to} G' \supset (E' \odot E') \times E'.$$

Let $T : G' \to L^1(E,P)$ be the linear mapping such that $\forall g' \in G'$, $T(g') = \langle (y \otimes y, y), g' \rangle$. Then, $\forall g' \in G'$, the Fourier transform of $B(P)$ is $\widehat{B(P)}(g') = \mathscr{E}\{\exp[-iT(g')]\}$. Under these conditions, we have the following theorem.

*Theorem* 2: Under the previous and notations, the Fourier transform $B(P)$ of probability measure $B(P)$ is the uniformly continuous function on $G'$ which is written for any $g'$ of $(E' \odot E') \times E'$:

$$B(P)(g') = \Phi(k'g'), \tag{23}$$

with $\Phi$ being the continuous function defined on $(H \widehat{\odot}_\pi H \times H$ such that $\forall (\theta_H, \mu_H) \in (H \widehat{\odot}_\pi H) \times H$

$$\Phi(\theta_H, \mu_H) = \{\det(1 + 2i\theta_H)\}^{-1/2}$$

$$\times \exp\{-\tfrac{1}{2}\langle m_H, m_H \rangle_H - \tfrac{1}{2}\langle (1 + 2i\theta_H)^{-1}$$

$$\times (\mu_H + im_H), \mu_H - im_H \rangle_{H^c}\}, \tag{24}$$

where $i = \sqrt{-1}$, $\det(1 + 2i\theta_H)$ is the Fredholm determinant[10] of the element $\theta_H \in H \widehat{\odot}_\pi H$ at the point $2i$, $\langle \cdot, \cdot \rangle_{H^c}$ the scalar product on the complexified space $H^c = H + iH$ of $H$ such that $\langle x_1, x_2 \rangle_{H^c} = \langle x_1, \bar{x}_2 \rangle_H$.

*Proof*: Let $R : H \to L^2(E,P)$ be the continuous linear mapping such that $\forall v \in E'$, $R(j'v) = \langle y, v \rangle$. Let $b$ be the bilinear symmetrical mapping of $H \times H$ in $L^1(E,P)$ such that $u_1, u_2 \to Ru_1, Ru_2$. Taking into account the universal property of the symmetrical tensor product, $b$ is factorized by the linear mapping $\tilde{b}$: $H \odot H \to L^1(E,P)$ denoted by $R \odot R$. Because $b$ is continuous, $R \odot R : H \widehat{\odot}_\pi H \to L^1(E,P)$ is continuous. Hence the linear mapping $S : (H \widehat{\odot}_\pi H) \times H \to L^1(E,P)$ such that $\forall \theta_H \in H \widehat{\odot}_\pi H$, $\forall \mu_H H$: $S(\theta_H, \mu_H) = (R \odot R)(\theta_H)$ $+ R(\mu_H)$ is continuous. A simple calculation shows that for $\forall g' \in (E' \odot E') \times E' \subset G'$, $T(g') = (S \circ k')(g')$. Because $S$ and $k'$ are continuous, $T$ is continuous and thus, the Fourier transform $\widehat{B(P)}$ is continuous also and is such that $\widehat{B(P)}(g') = \Phi(k'g')$ with $\Phi(u) = \mathscr{E}\{\exp[-iS(u)]\}$, which is continuous on $(H \widehat{\odot}_\pi H) \times H$. Let us now determine $\Phi$. We begin by assuming that $\theta_H \in H \odot H$. Let $H_q$ be a subspace of finite dimension of $H$, generated by the orthonormal basis $\{e_1, \ldots, e_q\}$ of dimension $q$ and such that $\theta_H \in H_q \odot H_q$ and $\mu_H \in H_q$. An orthonormal basis of $H_q \odot H_q$ is

$$e_k \odot e_p$$

$$= \{e_k \otimes e_k \text{ if } k = p \text{ or } \frac{1}{\sqrt{2}}(e_k \otimes e_p$$

$$+ e_p \otimes e_k) \text{ if } k < p\} \tag{25}$$

Decomposing $\theta_H$ on the basis defined by (25) and $\mu_H$ on the basis $\{e_k\}$, we obtain $\theta_H = \Sigma_{k<p} \theta^*_{kp} e_k \odot e_p$ with $\theta^*_{kp} = \theta_{kp}$ if $k = p$ and $\theta^*_{kp} = \sqrt{2}\theta_{kp}$ if $k < p$, the $\theta_{kp}$ being such that $\theta_H = \Sigma_{k,p} \theta_{kp} e_k \otimes e_p$, and $\mu_H = \Sigma_k \mu_k e_k$ with $\mu_k$ $= \langle \mu_H, e_k \rangle_H$. The restriction of $R \odot R$ (resp. $R$) to subspace $H_q \odot H_q$ (resp. $H_q$) is decomposed by the random variable $x \otimes x$ (resp. $x$) valued in $H_q \odot H_q$ (resp. $H_q$). Hence we have $(R \odot R)(\theta_H) = \langle x \otimes x, \theta_H \rangle_H$ and $R(\mu_H) = \langle x, \mu_H \rangle_H$. Then $\forall u = (\theta_H, \mu_H) \in (H_q \odot H_q) \times H_q$, we obtain

$$\Phi(\theta_H, \mu_H) = \mathscr{E}[\exp(-i\{\langle \Theta \mathbf{x}, \mathbf{x} \rangle + \langle \boldsymbol{\mu}, \mathbf{x} \rangle\}], \tag{26}$$

with $\Theta$ being the symmetrical real square matrix $(q \times q)$ of element $\Theta_{kp} = \theta_{kp}$, ${}^t\boldsymbol{\mu} = (\mu_1, \ldots, \mu_q)$ and ${}^t\mathbf{x} = (x_1, \ldots, x_q)$ where $x_k = \langle x, e_k \rangle_H$. Let $H_q^\perp$ be the orthogonal of $H_q$ into $H$. Then the restriction of $\rho$ to $H_q$ is the identity mapping of $H_q$ into $H / H_q^\perp \simeq H_q$. Taking $m_k = \langle m_H, e_k \rangle_H$ and ${}^t\mathbf{m} = (m_1, \ldots, m_q)$, (26) can be written as follows:

$$\Phi(\theta_H, \mu_H) = \int_{\mathbf{R}^q} (2\pi)^{-q/2} \exp\{-i(\langle \Theta \mathbf{x}, \mathbf{x} \rangle + \langle \boldsymbol{\mu}, \mathbf{x} \rangle)$$

$$- \tfrac{1}{2}\langle \mathbf{x} - \mathbf{m}, \mathbf{x} - \mathbf{m} \rangle\}\, d\mathbf{x},$$

with $d\mathbf{x} = dx_1 \times \cdots \times dx_q$. Grouping the terms and denoting $I_q$ as the unity matrix $(q \times q)$ we obtain

$$\Phi(\theta_H, \mu_H) = \{\det[I_q + 2i\Theta\,]\}^{-1/2} \exp\{-\tfrac{1}{2}\langle \mathbf{m}, \mathbf{m} \rangle$$

$$- \tfrac{1}{2}\langle [I_q + 2i\Theta\,]^{-1}(\boldsymbol{\mu} + i\mathbf{m}), \boldsymbol{\mu} + i\mathbf{m} \rangle\}$$

As a consequence, $\forall \theta_H \in H \odot H$ and $\forall \mu_H \in H$, we obtain in the limit

$$\Phi(\theta_H,\mu_H) = \{\det(1+2i\theta_H)\}^{-1/2} \times\exp\{-\tfrac{1}{2}\langle m_H,m_H\rangle_H - \tfrac{1}{2}\langle(1+2i\theta_H)^{-1} \times(\mu_H+\mathrm{im}_H),\mu_H-\mathrm{im}_H\rangle_{H'}\} \quad (27)$$

Recalling that $\Phi$ is continuous and as a consequence $\forall\theta_H\in H\hat{\odot}_\pi H$ and $\forall\mu_H\in H$. $\Phi$ is given by (27) and is obtained by continuous extension. For fixed $\lambda$ of $\mathbb{C}$, the mapping $\theta_H\to\det(1+\lambda\theta_H)$ is continuous of $H\hat{\odot}_\pi H$ in $\mathbb{C}$ (see Ref. 10); thus the exponential term of (27) is continuous. Finally, since $\widehat{B(P)}$ is continuous on $(E'\odot E')\times E'$, it is uniformly continuous and according to a theorem of topology, it extends by continuity to the closure of $(E'\odot E')\times E'$ in $G'$

## 5. EXPLICIT EXPRESSION OF $\hat{W}$

Let us apply the results of Sec. 4 for the explicit determination of $\hat{W}$ defined by (18) and (19). The probability measure $P_0$ on the Borel $\sigma$-field of $E = C(S,\mathbb{R}^m)$ defined by the process $(y_s)$, is Gaussian with zero mean, and satisfies the regular property and condition (20). On the other hand, we assume henceforth that the covariance operator $C_0$ of $P_0$ is injective. Taking into account (21) $\forall\mu = \Sigma_{j=1}^m \mu_j \mathbb{b}_j\in E'$ and $\forall v = \Sigma_{j=1}^m v_j\mathbb{b}_j\in E'$, we obtain

$$\langle C_0\mu,v\rangle = \int_S\int_S \sum_{j,j=1}^m R_{jj'}(s,s')\,d\mu_j(s)\,dv_{j'}(s'), \quad (28)$$

where $(s,s')\to R_y(s,s') = \Sigma_{j,j=1}^m R_{jj'}(s,s')\mathbb{b}_j\otimes\mathbb{b}_{j'}$, is the autocorrelation function defined on $S\times S$ with values in $\mathbb{R}^m\odot\mathbb{R}^m$ of the process $(y_s)$. We have $R_{jj'}(s,s') = \mathscr{E}[y_j(s)y_{j'}(s')]$ and since $(y_s)$ is time stationary we obtain for any $s=(t',u)$ and $s'=(t'',u')$ of $S$, $R_y(s,s') = R_y(t'-t'',u,u')$. Considering (18) and (19), it is enough to determine $\beta_t'$ and $A'$ and use (23) and (24) by putting $m_H = 0$, $P_0$ being centered. After completing the calculating, we obtain

$$\forall(a',b')\in\mathbb{R}^2,\ \hat{W}(a',b') = \hat{Q}(a'\mu_1 + b'\mu_2) \quad (29)$$

with $a'\mu_1 + b'\mu_2\in F' = M(S)$, $\mu_1$ and $\mu_2$ the measures defined in (10), and where $\hat{Q}$ is the Fourier transform of the probability measure $Q = f(P_0)$. We have for any $\mu$ of $F'$:

$$\hat{Q}(\mu) = \{\det[1+2i(j'\otimes j')(\tilde{\mu})]\}^{-1/2}\exp\{-i\langle\mu,r\rangle - \tfrac{1}{2} <[1+2i(j'\otimes j')(\tilde{\mu})]^{-1}j'(q\cdot\mu),j'(q\cdot\mu)\rangle_{H'}\}, \quad (30)$$

where $\tilde{\mu}\in E'\odot E'$ and is such that $\forall s,s'\in S$, $\tilde{\mu}(s,s') = p$. $(\mu(s)\otimes\delta_s(s'))$ with $p$, $q$, $r$ as defined in Sec. 2 and $\delta_s$ the Dirac measure at point $s$ of $S$.

## 6. THE APPROXIMATION PROCEDURE

In its actual form, expression (30) cannot be calculated numerically because it involves operators acting on space of infiite dimension. Let $(\beta_v)_v$ be a sequence of $L[C(S),\mathbb{R}^2]$, $\beta_v$ having a finite rank. For all $(a,b)\in\mathbb{R}^2$, $\beta_v(a,b)\to\beta_t(a,b)$ if $v\to\infty$. An approximation of $W = (\beta_t\circ f)(P_0)$ is given by $W_v = (\beta_v\circ f)(P_0) = \beta_v(Q)$ and then, $\forall(a',b')\in\mathbb{R}^2$, $\hat{W}_v(a',b') = \hat{Q}[\beta_v'(a',b')]$. By the dominant convergence theorem, we obtain $\hat{W}_v(a',b')\to\hat{W}(a',b')$ if $v\to\infty$. Then, $\hat{W}_v$ can be numerically calculated since $\beta_v'(a',b')$ belongs to a finite-dimensional subspace of $M(S)$. We now apply this approximation procedure. Taking into account point $C$ of Sec. 2, (10) can be written:

$$\beta_t(z) = \left(\int_{K_t} z(s)g_1(t,s)\,ds,\ \int_{K_t} z(s)g_2(t,s)\,ds\right). \quad (31)$$

The compact subset $K_t$ of $S$ is discretized in $v$ subsets $\Delta s_k$, $k\in\{1,v\}$. The center of $\Delta s_k$ is the point $s_k$ of $K_t$. If $v\to\infty$, $\max_k|\Delta s_k|\to 0$. For any $z\in C(S)$ we define

$$\beta_v(z) = \left[\sum_{k=1}^v v_k z(s_k),\ \sum_{k=1}^v \dot{v}_k z(s_k)\right], \quad (32)$$

with $v_k = g_1(t,s_k)|\Delta s_k|$, $\dot{v}_k = g_2(t,s_k)|\Delta s_k|$. It can be easily shown that $\beta_v(z)\to\beta_t(z)$ when $v\to\infty$ and that $\forall(a',b')\in\mathbb{R}^2$, $\beta_v'(a',b') = \Sigma_{k=1}^v(a'v_k + b'\dot{v}_k)\delta_{s_k}$, where $\delta_{s_k}$ is the Dirac measure at the point $s_k$ of $K_t\subset S$. It remains only to determine $\hat{Q}(\mu)$ given by the equation (30), with $\mu$ in form of $\mu = \Sigma_{k=1}^v\mu_k\delta_{s_k}$. The following result is then obtained for fixed $v\geqslant 1$ and $\forall\alpha' = (a',b')\in\mathbb{R}^2$

$$\hat{W}_v(\alpha') = \{\det[I_q + 2iL\tilde{\Theta}(\alpha')'L]\}^{-1/2} \times\exp\{-ie(\alpha') - \tfrac{1}{2}'[LN(\alpha')] \times[I_q + 2iL\tilde{\Theta}(\alpha')'L]^{-1}LN(\alpha')\}, \quad (33)$$

with:

$q = v\times m$ and $I_q$ the real unity matrix $(q\times q)$; (34)

$L$: the real triangular matrix $(q\times q)$ and $'L$ the transposed matrix such that $C_q = 'LL$, where $C_q$ is the real square matrix $(q\times q)$, symmetrical and positive definite and $\{C_q\}_{pp'} = \langle C_0e_p,e_{p'}\rangle$. $C_0$ is the covariance operator defined by (28) and

$$e_p = \delta_{s_k}\otimes\mathbb{b}_j,\ p=(k,j),\ k\in\{1,v\},\ j\in\{1,m\},\ \text{and}\ p\in\{1,q\}; \quad (35)$$

$$\mu_k = a'v_k + b'\dot{v}_k \quad (36)$$

$$e(\alpha') = \sum_{k=1}^v \mu_k r(s_k) \quad (37)$$

$\tilde{\Theta}(\alpha')$ the symmetrical real square matrix $(q\times q)$ of element $\{\tilde{\Theta}(\alpha')\}_{pp'} = \tilde{\tilde{\mu}}_{kk'}(\alpha')p_{jj'}$ such that $p=(k,j)$; $p'(k',j')$; $p,p'\in\{1,q\}$; $k,k'\in\{1,v\}$; $j,j'\in\{1,m\}$;

$$\tilde{\tilde{\mu}}_{kk'} = 0 \text{ if } k\neq k' \text{ and } \tilde{\tilde{\mu}}_{kk}(\alpha') = \mu_k \text{ if } k = k'. \quad (38)$$

$N(\alpha')$ = the column matrix of dimension $q$ of element $\{N(\alpha')\}_p = q_j(s_k)\mu_k$ such that

$$p = (k,j) \quad (39)$$

If $R_y: S\times S\to\mathbb{R}^m\odot\mathbb{R}^m$ is continuous, then $\langle C_0e_p,e_{p'}\rangle = R_{jj'}(s_k,s_{k'})$ with $p=(k,j)$,

$$p' = (k',j'),\ R_{jj'}\text{, being defined in (28)} \quad (40)$$

*Remark* 1: In order to obtain the result (33), we must demonstrate that $\forall\theta\in E'\odot E'\subset(E\odot_\epsilon E)'$ and denoting $\tilde{\theta}$ as the continuous linear operator of $E$ in $E'$ canonically associated with $\theta$, $j'\tilde{\theta}j\rho$ is the continuous linear operator of $H$ in $H$ canonically associated with $(j'\otimes j')(\theta)\in H\hat{\odot}_\pi H$. Denoting $M_\delta(S)$ as the space of finite linear combinations of Dirac

measures on $S$ generated by the basis $\{\delta_{s_1},\ldots,\delta_{s_N}\}$, we obtain the result by adopting the finite dimension subspace $H_q = M_\delta(S)\otimes\mathbb{R}^m$ and $E' = M(S,\mathbb{R}^m)$, the basis of $H_q$ being $\{e_p = \delta_{s_k}\otimes \mathbf{b}_j\}$, $k\in\{1,v\}$, $j\in\{1,m\}$ and $p\in\{1,q\}$ where $p$ is the index associated to the pair of indices $(k,j)$.

*Remark* 2: We should note that another basis could be used for $H_q$ which would be the eigenfunctions of the covariance operator $C_0$ of the process $(y_s)$. Thus let $\mu_\varphi = \varphi(s)\,ds \in M(S,\mathbb{R}^m)$, $\varphi(s)\equiv 0$ if $s\notin K_t\subset S$. The eigenfunctions of $C_0$ are solutions of $\varphi(s) = \lambda(C_0\mu_\varphi)(s)$, $s\in K_t$, where $\lambda$ is the characteristic number associated with $\varphi$. Using equation (28), $\varphi(s) = \sum_{j=1}^m \varphi_j(s)\mathbf{b}_j$ is the solution of

$$\varphi(s) = \lambda\int_{K_t}\sum_{j,j'=1}^{m} R_{jj'}(s,s')\varphi_j(s')\mathbf{b}_j\,ds', \quad s\in K_t \tag{41}$$

The kernel $R_y$ is symmetrical, real, and positive definite on $K_t\times K_t$ (because $C_0$ is injective), all $\lambda_1,\lambda_2,\cdots$ are strictly positive, and the associated eigenfunctions $\varphi^{(1)},\varphi^{(2)},\cdots$ form an orthonormal basis of $L^2(K_t,\mathbb{R}^m)$. We have $\langle\varphi^{(p)},\varphi^{(p')}\rangle = \delta_{pp'}$ with $\delta_{pp'}$ the Kronecker symbol. Then the orthonormal basis $\{e'_p\}$ of $H_q$ is such that $e'_p = \mu_{\varphi^{(p)}}$; thus $e'_p = \varphi^{(p)}(s)\,ds$ if $s\in K_t$ and $e'_p = 0$ if $s\notin K_t$; we have $\langle C_0 e'_p, e'_{p'}\rangle = \delta_{pp'}\lambda_p$. Under these conditions, expression (33) for $\hat{W}_v$ as well as (34), (35), (36), and (37) are unchanged, but (38), (39), and (40) must be replaced by Eqs. (42), (43), and (44) respectively.

$$\{\widetilde{\Theta}(\alpha')\}_{pp'} = \left\langle p\sum_{k=1}^{v}\mu_k\delta_{s_k}\otimes\delta_{s_k}, e'_p\otimes e'_{p'}\right\rangle = \sum_{j,j'=1}^{m} p_{jj'}\sum_{k=1}^{v}\mu_k\varphi_j^{(p)}(s_k)\varphi_{j'}^{(p')}(s_k), \tag{42}$$

$$\{N(\alpha')\}_p = \left\langle\sum_{k=1}^{v}\sum_{j=1}^{m} q_j(s_k)\mu_k\delta_{s_k}\otimes\mathbf{b}_j, e'_p\right\rangle = \sum_{k=1}^{v}\sum_{j=1}^{m} q_j(s_k)\mu_k\,\varphi_j^{(p)}(s_k), \tag{43}$$

$$\{L\}_{pp'} = \delta_{pp'}\sqrt{\lambda_p}. \tag{44}$$

However, generally there is no analytical solution for Eqs. (41) and furthermore, from a numerical point of view, the intergation with respect to $K_t\subset S = \mathbb{R}^{n+1}$ is a $(n+1)$-tuple integral.

*Remark* 3: Expression (33) therefore enables a numerical calculation of an approximation of $v$-order of the probability density $w$. We shall denote by $w_1$ the probability density of the law of the random variable $x_t$. Approximations of $v$-order will be denoted by $w_v$ and $w_{1v}$ respectively. Since we know $\hat{W}_v$ given by (33) for any $(a',b')$ of $\mathbb{R}^2$, we obtain for any $(a,b)$ of $\mathbb{R}^2$:

$$w_{1v}(a) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(iaa')\hat{W}_v(a',0)\,da', \tag{45}$$

$$w_v(a,b) = \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp[i(aa'+bb')]\times\hat{W}_v(a',b')\,da'\,db'. \tag{46}$$

A direct numerical integration of (45) and (46) can be considered, but it leads to very complicated calculations because using expression (33) for $\hat{W}_v$, it will be necessary to compute a very large number of inversions of complex non-Hermitian matrices of order $v$. We have therefore worked out expressions much more rapidly computed, which allow us to solve a restricted number of eigenvalue problems of a real symemtrical matrix. We shall not give the proof which is lengthy but not difficult.[2] Let $L$ be the matrix defined in (35). For $\theta\in[0,2\pi]$, we define the real matrices $(q\times q)$: $U(\theta)$, $V(\theta)$, $\Psi(\theta)$ and $\Omega(\theta)$ such that $q = v\times m$; $\{U(\theta)\}_{pp'} = \delta_{kk'}(v_k\cos\theta + \dot{v}_k\sin\theta)p_{jj'}$, $V(\theta) = 2LU(\theta)'L$; $\Psi(\theta)$ the matrix of eigenvectors and $\Omega(\theta)$ the diagonal matrix of eigenvalues of $V(\theta)$. In that case, for $q>4$, we have for any $a$ and $b$ of $\mathbb{R}$

$$w_{1v}(a) = \frac{1}{\pi}\int_0^{+\infty}\gamma_2(\rho,0)\cos(\gamma_1(\rho,0)+\rho a)\times\exp[-\gamma_3(\rho,0)]\,d\rho, \tag{47}$$

$$w_v(a,b) = \frac{1}{2(\pi)^2}\int_0^{\pi}\int_0^{+\infty}\rho\gamma_2(\rho,\theta)\cos(\gamma_1(\rho,\theta)+\rho(a\cos\theta+b\sin\theta))\exp[-\gamma_3(\rho,\theta)]\,d\rho\,d\theta. \tag{48}$$

with:

$$\gamma_1(\rho,\theta) = -\tfrac{1}{2}\sum_{p=1}^{q}\arctan[\rho\Omega_{pp}(\theta)] - \rho e(\theta) + \tfrac{1}{2}\rho^3\mathbb{D}(\theta)\Psi(\theta)[I_q+\rho^2\Omega^2(\theta)]^{-1}{}'\Psi(\theta)\times V(\theta)\mathbb{D}(\theta),$$

$$\gamma_2(\rho,\theta) = \prod_{p=1}^{q}[1+\rho^2\Omega_{pp}^2(\theta)]^{-1/4},$$

$$\gamma_3(\rho,\theta) = \tfrac{1}{2}\rho^2\mathbb{D}(\theta)\Psi(\theta)[I_q+\rho^2\Omega^2(\theta)]^{-1}{}'\Psi(\theta)\mathbb{D}(\theta),$$

where $\mathbb{D}(\theta) = LN(\theta)$, whose $p$th element of the column matrix $N(\theta)$ of dimension $q$ is $\{N(\theta)\}_p = q_j(s_k)(v_k\cos\theta + \dot{v}_k\sin\theta)$ $e(\theta)$ is the element of $\mathbb{R}$ such that $e(\theta) = \sum_{k=1}^{v} r(s_k)(v_k\cos\theta + \dot{v}_k\sin\theta)$. The expression of $w_v$ thus allows a numerical computation of (5), (6), (7), and (8). Nevertheless, the direct numerical computation of (8), can be very difficult. We shall thus given an approximate expression for $\bar{X}_{\max}$ in analytical form which leads to reasonably rapid numerical computations.

## 7. MEAN OF POSITIVE EXTREME VALUES OF THE PROCESS $(x_t)$

Let $M_{pq}$ be the moment of order $(p,q)$ of the random variables $x_t$ and $\dot{x}_t$; $\forall p\geqslant 0$, $\forall q\geqslant 0$, we have

$$M_{pq} = \mathscr{E}(x_t^p\dot{x}_t^q). \tag{49}$$

Then the mean of the positive extreme values $\bar{X}_{\max}$ of the stationary centered process $(x_t)$ over the time $T$ is given by the following proposition:

*Proposition*: Assuming the hypothesis (7) and taking the Hermite polynomial expansion restricted to the 4th order of $w$, a numerical approximation $\bar{X}^*_{\max}$ of $\bar{X}_{\max}$ is given by

$$\bar{X}'_{\max} = gM_{20}^{1/2}, \tag{50}$$

$$g = (2\ln v_N T)^{1/2} + (2\ln v_n T)^{-1/2}[\gamma + \ln(1+\varphi)],$$

with $\gamma = 0.577\cdots$ the Euler constant,

$$v_n = (2\pi)^{-1} M_{02}^{1/2} M_{20}^{-1/2}; \varphi = \sum_{j=0}^{4} a_j (2 \ln v_N T)^{j/2}$$

$$a_0 = \tfrac{1}{24}(3M_{40}M_{20}^{-2} - 6M_{22}M_{20}^{-1}M_{02}^{-1} - M_{04}M_{02}^{-2})$$

$$a_1 = \tfrac{1}{2}(M_{12}M_{20}^{-1/2}M_{02}^{-1} - M_{30}M_{20}^{-3/2})$$

$$a_2 = \tfrac{1}{4}(M_{22}M_{20}^{-1}M_{02}^{-1} - M_{40}M_{20}^{-2} + 2) \qquad (51)$$

$$a_3 = \tfrac{1}{6}M_{30}M_{20}^{-3/2}$$

$$a_4 = \tfrac{1}{24}(M_{40}M_{20}^{-2} - 3)$$

*Comments*:

(1) We shall not give here the very long and much complicated proof; the interested reader is referred to Ref. 2. The principal of this proof consists in developing in Hermite polynomials the probability density $w_R$ of the joint law of the normalized random variables $X_R = M_{20}^{-1/2} x_t$ and $\dot{X}_R = M_{02}^{-1/2}\dot{x}_t$. We obtain to the 4th order $w_R(\eta,\xi)\, d\eta\, d\xi = \alpha(\eta)\alpha(\xi)[1 + \Sigma_{p,q=1}^{4} c_{pq} H_p(\eta) H_q(\xi)]\, d\eta\, d\xi$, with $\alpha(x) = (2\pi)^{-1/2} \exp(-x^2/2)$; $d^p\alpha(x)/dx^p = (-1)^p H_p(x)\,\alpha(x)$ and where the coefficients $c_{pq}$ are expressed as functions of the cumulant $k_{pq}$.

(2) If the transformation $f$ of the problem under consideration were linear, then the processes $(x_t)$ and $(\dot{x}_t)$ would be Gaussian; under these conditions, we should have the classical expression $g = (2 \ln v_N T)^{1/2} + \gamma(2 \ln v_N T)^{-1/2}$. Here, $f$ is nonlinear and as a consequence, the term $(2\ln v_N T)^{-1/2} \ln(s + \varphi)$ of (50) occurs as a corrective term of the linear case because the process $(x_t)$ is not Gaussian.

(3) Lastly, one should note that using (50) requires the determination of the moments $M_{pq}$, $p,q \in \{1,4\}$. Although it is possible to proceed without greatly difficulty to an analytical calculation of $M_{20}$ and $M_{02}$, the direct analytical determination without using $w$ of the other moments may be impossible. We shall therefore calculate the approximate moments $M_{pq}^{*}$ from the approximation for $\hat{W}_v$ given by (33). Using the approximation procedure it is obvious that $M_{pq}^{*} \to M_{pq}$ when $v \to \infty$. Values $c_j(\theta)$ for $j \in \{1,4\}$ are defined such that $\forall\theta \in [0,2\pi]$

$$c_1(\theta) = \tfrac{1}{2}\mathrm{Tr}[\Omega(\theta)] + e(\theta) = \mathrm{Tr}[U(\theta)C_q] + e(\theta),$$

$$c_2(\theta) = {}^t\mathbb{D}(\theta)\mathbb{D}(\theta) + \tfrac{1}{2}\mathrm{Tr}[\Omega^2(\theta)]$$
$$= {}^tN(\theta)C_q N(\theta) + 2\,\mathrm{Tr}([(U(\theta)C_q)^2]), \qquad (52)$$

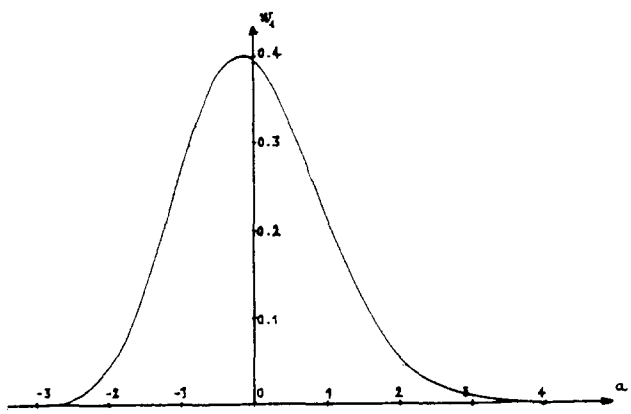


FIG. 1.

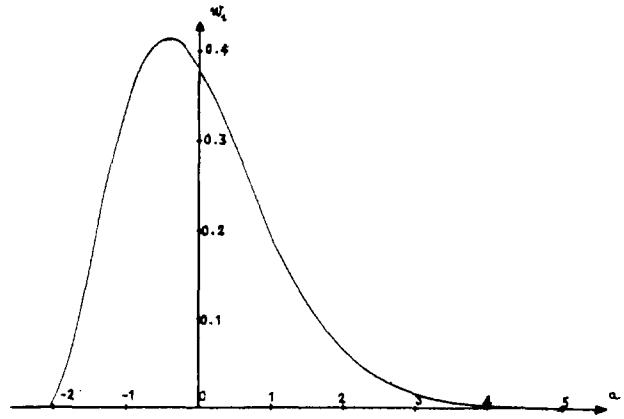


FIG. 2.

$$c_3(\theta) = {}^t\mathbb{D}(\theta)V(\theta)\mathbb{D}(\theta) + \tfrac{1}{3}\mathrm{Tr}[\Omega^3(\theta)]$$
$$= 2{}^tN(\theta)C_q U(\theta)C_q N(\theta) + \tfrac{8}{3}\mathrm{Tr}([U(\theta)C_q]^3),$$

$$c_4(\theta) = {}^t\mathbb{D}(\theta)V^2(\theta)\mathbb{D}(\theta) + \tfrac{1}{4}\mathrm{Tr}[\Omega^4(\theta)]$$
$$= 4{}^tN(\theta)C_q[U(\theta)C_q]^2 N(\theta) + 4\,\mathrm{Tr}([U(\theta)C_q]^4).$$

where $C_q$ is defined in (35), $U(\theta)$, $N(\theta)$, $e(\theta)$, $\Omega(\theta)$, $\mathbb{D}(\theta)$, $V(\theta)$ are defined in Sec. 6, and Tr[·] is the trace of matrix [·]. Let $X_1(\theta)$, $X_2(\theta)$, $X_3(\theta)$ be such that, $\forall\theta \in [0,2\pi]$

$$X_1(\theta) = c_1^2(\theta) + c_2(\theta),$$
$$X_2(\theta) = c_1^3(\theta) + 3c_1(\theta)c_2(\theta) + 3c_3(\theta), \qquad (53)$$
$$X_3(\theta) = c_1^4(\theta)c_2(\theta) + 6c_1^2(\theta)c_2(\theta) + 12c_1(\theta)c_3(\theta) + 3c_2^2(\theta) + 14c_4(\theta).$$

In that case, the approximation $M_{pq}^{*}$ of order $v$ of the moments $M_{pq}$ is[2]

$$M_{10}^{*} = c_1(0);\ M_{01}^{*} = c_1(\pi/2);\ M_{20}^{*} = X_1(0);$$
$$M_{02}^{*} = X_1(\pi/2);$$
$$M_{30}^{*} = X_2(0);\ M_{40}^{*} = X_3(0);\ M_{04}^{*} = X_3(\pi/2); \qquad (54)$$
$$M_{12}^{*} = -\tfrac{1}{3}M_{30}^{*} + (\sqrt{2}/3)[X_2(\pi/4) - X_2(3\pi/4)],$$
$$M_{22}^{*} = (1/3)[X_3(\pi/4) + X_3(3\pi/4)] - (1/6)[M_{40}^{*} + M_{04}^{*}].$$

## 8. NUMERICAL EXAMPLE

Let $(y_s)$ be a centered Gaussian process on $S = [0,1]$ with a given covariance function $R_y(s,s') = \mathscr{E}[y(s)y(s')] = r\exp(-1.065 s^{-0.5}|s - s'|)$ for $s \geqslant s'$. The formula (47) can be applied to compute an approximation $w_{1v}$ of the density $w_1$ of the distribution of the $\mathbb{R}$-valued random variables $X$

$$X = \int_0^1 (y_s^2 + q(s)y_s - r)\, g(s)\, ds.$$

The results are represented by Figs. 1 and 2 and correspond to the following values of $g$, $q$ and $r$:

Fig. 1: $g(s) = 6.086s$, $q(s) = 2.480s^{0.15}$, $r = 0.02657$.

Fig. 2: $g(s) = 1.823s$, $q(s) = 3.305s^{0.35}$, $r = 0.19981$.

## 9. CONCLUDING REMARKS

The writer has developed a general computer program that permits the calculation of $w_{1v}$, $w_v$ given by (47) and (48), the various $M^*_{pq}$ moments given by (54), and $\bar{X}^*_{max}$ given by (50). This computer program has been used to study the vibrations of structures due to effects of turbulent wind; the numerical and practical results obtained on this point have been published in Ref. 1.

We have presented in this paper an approach which is constructive from a numerical point of view for solving stochastic problems to which Markovian methods do not apply. This method is general and can be used for other kinds of stochastic problems.


## ACKNOWLEDGMENT

I wish to thank very much P. Kree of the University of Paris for the help which he has given me while carrying out this work.



[1]C. Soize, "Gust Loading Factors with Nonlinear Pressure Terms", J. Struct. Div., ASCE **104** No. ST6 991–1007 (1978).
[2]C. Soize, "Resolution de deux problémes de méchanique aléatoire", Thèse de Doctorat d'Etat ès Sciences physiques, Univ. Pierre et Marie Curie, Paris VI (1979).
[3]M. R. Leadbetter, "On Crossing of Levels and Curves by a Wide Class of Stochastic Processes", Annales of Math. Stat., Vol. 37 (1966).
[4]H. Cramer, and M. R. Leadbetter, *Stationary and Related Stochastic Processes* (Wiley, New York, 1967).
[5]P. Kree, *Methods of Functional Analysis in the Study of Random Phenomena* (*North–Holland*, 1979).
[6]K. Yosida, *Functional Analysis*, Third edition (Springer-Verlag, Berlin, 1971).
[7]A. Grothendieck, "Produits tensoriels topologiques et espaces nucléaire", Memoirs Amer. Math. Soc. Providence, **16** (1955).
[8]H. Sato-Nagoya, Math. J. 36, pp. 65–81 (1969).
[9]X. Fernique, Compte rendu Acad. Sci. Paris **270**, pp. 1698–9 (1970).
[10]A. Grothendieck, Bull. Soc. Math. Fr. **84**, 319–84 (1956).

# Factorization of operators I. Miura transformations

Allan P. Fordy and John Gibbons
*School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland*

(Received 27 December 1979; accepted for publication 15 February 1980)

The method of factorization of operators, which has been used to derive the Miura transformation of the KdV equation, is here extended to some third-order scattering operators, and transformations between several fifth-order nonlinear evolution equations are derived. Further applications are discussed.

## 1. INTRODUCTION

It has been known for some time[1,2] that the Miura transformation[3]

$$u = v_1 - v^2 \tag{1.1}$$

(where $v_n = [\partial^n/\partial x^n]v$), which relates solutions of the Korteweg–de Vries (KdV) and modified Korteweg–de Vries (MKdV) equations

$$\begin{aligned} u_t &= u_3 + 6uu_1 \\ v_t &= v_3 - 6v^2v_1 \end{aligned} \tag{1.2}$$

may be derived by factorizing the scattering operator for the KdV equation. From this factorization one may also derive Wadati's[4] scattering problem for the MKdV equation, as will be shown below.

In this paper we use this approach to investigate two fifth-order nonlinear evolution equations, deriving Miura transformations between them and a single "modified" equation, and a Lax pair of operators for the modified equation.

The Hamiltonian structures of these equations are shown to be related by the Miura transformation, and this approach permits a verification of the validity of the Miura transformation which is simpler than the direct method.

Of course, the original development of the inverse scattering transform proceeded via a linearization of the Ricatti equation (1.1) to give the Schrödinger operator.[5] We show here, however, that it is not always possible to proceed in this manner, the relationship between Miura transformations and scattering operators being more complicated in general.

## 2. THE SCHRÖDINGER OPERATOR

To illustrate the method of factorization we consider the operator

$$L = \partial^2 + u, \tag{2.1}$$

where

$$\partial = \partial_x = \frac{\partial}{\partial x}.$$

We require that this should be equal to

$$L' = (\partial - v)(\partial + v). \tag{2.2}$$

An elementary computation gives

$$u = v_1 - v^2, \tag{2.3}$$

which is the transformation discovered by Miura.[3]

### A. Hamiltonian structures

To derive the transformed equation, we consider the Hamiltonian structure of the KdV equation.[6] The equation possesses infinite sets, both of conserved quantities and of Poisson brackets

$$\{F,G\}_n = \int_{-\infty}^{\infty} \frac{\delta F}{\delta u} \Pi_n \frac{\delta G}{\delta u}\, dx, \tag{2.4}$$

where the skew adjoint operator

$$\Pi_n = \left(\partial^2 + 4u + 2u_1 \int_{-\infty}^{x}\right)^n \partial \tag{2.5}$$

is the "Poisson operator". The conserved quantities are in involution with respect to any one of the $\Pi_n$.

The KdV equation may be generated by the Hamiltonian

$$H = \int_{-\infty}^{\infty} \tfrac{1}{2}u^2 dx \tag{2.6}$$

and the Poisson operator $\Pi_1$

$$u_t = \Pi_1 \frac{\delta H}{\delta u} = u_3 + 6uu_1. \tag{2.7}$$

Now $v$ satisfies some Hamiltonian equation

$$v_t = \Pi_v \frac{\delta K}{\delta v}, \tag{2.8}$$

where the Poisson operator $\Pi_v$ and the Hamiltonian $K$ are to be determined. The natural construction for $K$ is to express the Hamiltonian $H$ in terms of $v$

$$K[v] = H[v_1 - v^2] = \int_{-\infty}^{\infty} \tfrac{1}{2}(v_1^2 + v^4)\, dx. \tag{2.9}$$

Now, quite generally, a Miura transformation of the form

$$u = F[v] \tag{2.10}$$

implies

$$\begin{aligned} u_t &= F'[v]v_t \\ &= F'[v]\Pi_v \frac{\delta K}{\delta v} \\ &= F'[v]\Pi_v (F'[v])^\dagger \big|_{u=F[v]} \frac{\delta H}{\delta u} \\ &= \Pi_u \frac{\delta H}{\delta u}, \end{aligned} \tag{2.11}$$

where the operator $F'[v]$ is the usual Fréchet derivative of $F[v]$.

For all the systems discussed in this paper, it is a re-

 0022-2488/80/102508-03$1.00  

markable fact that there exist *differential* operators $\Pi_u$ and $\Pi_v$ which are related in this fashion, for if we take $\Pi_v$ as $-\partial$, then the Poisson operator $\Pi_u$ must be

$$\begin{aligned}\Pi_u &= (\partial - 2v)(-\partial)(\partial - 2v)^\dagger\big|_{u=v_1-v^2} \\ &= [\partial^3 + 4(v_1 - v^2)\partial + 2(v_2 - 2vv_1)\,]\big|_{u=v_1-v^2},\end{aligned} \tag{2.12}$$

which is just the operator $\Pi_1$.

The relationship between the factorizations (2.2) and (2.12), of the scattering operator and the Poisson operator respectively, possibly deserves further investigation. Equation (2.8) now becomes

$$v_t = v_3 - 6v^2v_1. \tag{2.13}$$

### B. The eigenvalue problem

The operator (2.2) gives rise to the second eigenvalue problem

$$(\partial - v)(\partial + v)\psi = \zeta^2\psi. \tag{2.14}$$

This may be decomposed into the pair of first-order equations

$$\begin{aligned}(\partial + v)\psi &= \zeta\phi \\ (\partial - v)\phi &= \zeta\psi.\end{aligned} \tag{2.15}$$

By introducing $\psi_+ = \psi + \phi$ and $\psi_- = \psi - \phi$, we would get

$$\partial\begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix} = \begin{pmatrix}\zeta & v\\ v & -\zeta\end{pmatrix}\begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix}, \tag{2.16}$$

which is the Zakharov–Shabat eigenvalue problem, first introduced in this context by Wadati.[4] The form (2.15), however, is more convenient for the purpose of this paper.

The time evolution of $\psi$ is clearly given by

$$\begin{aligned}\psi_t &= (4\partial^3 + 6u\partial + 3u_1)\psi \\ &= (4\partial^3 + 6(v_1 - v^2)\partial + 3(v_2 - 2vv_1))\psi,\end{aligned} \tag{2.17}$$

while that of $\phi$ is given by

$$\phi_t = [4\partial^3 - 6(v_1 + v^2)\partial - 3(v_2 + 2vv_1)\,]\phi. \tag{2.18}$$

Hence

$$\partial_t\begin{pmatrix}\psi\\ \phi\end{pmatrix} = \begin{pmatrix}4\partial^3 + 6(v_1 - v^2)\partial + 3(v_2 - 2vv_1) & 0\\ 0 & 4\partial^3 - 6(v_1 + v^2)\partial + 3(v_2 + 2vv_1)\end{pmatrix}\begin{pmatrix}\psi\\ \phi\end{pmatrix}, \tag{2.19}$$

where the matrix on the right may clearly be rewritten as a matrix of polynomials in $\zeta$, $v$, and its derivatives, which is often more convenient for applications.

## 3. SOME THIRD ORDER OPERATORS

Two fifth order nonlinear evolution equations which are of some interest are

$$u_t = u_5 + 5uu_3 + 5u_1u_2 + 5u^2u_1, \tag{3.1a}$$

$$w_t = w_5 + 10ww_3 + 25w_1w_2 + 20w^2w_1. \tag{3.1b}$$

The first of these is due to Sawada and Kotera[7] and Gibbon[8] while the second is due to Kupershmidt.[9] They both have Lax representations

$$L_t = [L,G\,], \tag{3.2a}$$

$$M_t = [M,K\,], \tag{3.2b}$$

where

$$\begin{aligned}L &= \partial^3 + u\partial, \\ G &= 9\partial^5 + 15u\partial^3 + 15u_1\partial^2 + (5u^2 + 10u_2)\partial\end{aligned} \tag{3.3a}$$

and

$$\begin{aligned}M &= \partial^2 + 2w\partial + w_1, \\ K &= 9\partial^5 + 30\,w\partial^3 + 45\,w_1\partial^2 \\ &\quad + (20w^2 + 35\,w_2)\partial + (10w_3 + 20\,ww_1).\end{aligned} \tag{3.3b}$$

We factorize the scattering operators $L$ and $M$ as follows:

$$L = (\partial - v)(\partial + v)\partial, \tag{3.4a}$$

$$M = (\partial + v)\partial(\partial - v), \tag{3.4b}$$

where the choice of the same notation, $v$, for the new variables in the two cases will be justified below.

The Miura transformations are seen to be

$$u = v_1 - v^2 = P[v], \tag{3.5a}$$

$$w = -v_1 - \tfrac{1}{2}v^2 = Q[v]. \tag{3.5b}$$

The Poisson operators for the systems (3.1a) and (3.1b) are

$$\Pi_u = \partial^3 + 4u\partial + 2u_1, \tag{3.6a}$$

$$\Pi_w = \partial^3 + 2w\partial + w_1. \tag{3.6b}$$

They generate Eqs. (3.1a) and (3.1b) from the Hamiltonians

$$H_G = \int_{-\infty}^{\infty} \tfrac{1}{6}(u^3 - 3u_1^2)\,dx \tag{3.7a}$$

$$H_K = \int_{-\infty}^{\infty} \tfrac{1}{6}(8w^3 - 3w_1^2)\,dx. \tag{3.7b}$$

Now the Poisson operators may be factorized, as in Sec. 2, giving

$$\begin{aligned}\Pi_u &= (\partial - 2v)(-\partial)(-\partial - 2v) \\ &= (P')(\Pi_v)(P')^\dagger,\end{aligned} \tag{3.8a}$$

$$\Pi_w = (-\partial - v)(-\partial)(\partial - v) = (Q')(\Pi_v)(Q')^\dagger. \tag{3.8b}$$

It may be observed that the transformed Poisson operator $\Pi_v$ is the same in both cases. The Hamiltonians $H_G$ and $H_K$ both transform into a single functional

$$H_v = \int_{-\infty}^{\infty} -\tfrac{1}{6}(3v_2^2 + 5v_1^3 + 15v_1^2v^2 + v^6)\,dx. \tag{3.9}$$

Hence the transformed equation is

$$\begin{aligned}v_t &= \Pi_v\frac{\delta H_v}{\delta v} \\ &= (-\partial)(-v_4 + 5v_1v_2 + 5v\,v_1^2 + 5v^2v_2 - v^5) \\ &= v_5 - 5(v_1v_3 + v_2^2 + v_1^3 + 4vv_1v_2 + v^2v_3 - v^4v_1).\end{aligned} \tag{3.10}$$

That the factorizations (3.4a) and (3.4b) lead to a single evo-

evolution equation in $v$ justifies the notation adopted in equations (3.4). The validity of the Miura transformations may be verified by direct substitution, but such a calculation is much more cumbersome than the method used here.

### A. The eigenvalue problem

The system of three first-order equations

$$\begin{aligned}(\partial - v)\phi &= \zeta\chi,\\ \partial\chi &= \zeta\psi,\\ (\partial + v)\psi &= \zeta\phi,\end{aligned} \tag{3.11}$$

may, by elimination, be reduced to any one of

$$\begin{aligned}(\partial + v)\partial(\partial - v)\phi &= \zeta^3\phi,\\ (\partial - v)(\partial + v)\partial\chi &= \zeta^3\chi,\\ \partial(\partial - v)(\partial + v)\psi &= \zeta^3\psi,\end{aligned} \tag{3.12}$$

which are, respectively, the eigenvalue problems of the Kupershmidt operator $M$, the Gibbon operator $L$, and the operator $(-L^{\dagger})$.

The existence of only a single fifth-order isospectral flow, (3.10), for the operators (3.4a) and (3.4b), is thus seen to be a consequence of these being merely different scalar representations of the single system (3.11).

The time evolution of the column vector $(\phi,\chi,\psi)^T$ is given by

$$\partial_t \begin{pmatrix}\phi\\ \chi\\ \psi\end{pmatrix} = \begin{pmatrix}K & 0 & 0\\ 0 & G & 0\\ 0 & 0 & -G^{\dagger}\end{pmatrix}\begin{pmatrix}\phi\\ \chi\\ \psi\end{pmatrix}, \tag{3.13}$$

where $G$ and $K$ are given by (3.3a) and (3.3b), respectively, but expressed in terms of $v$ by means of Eqs. (3.5a) and (3.5b).

The integrability condition for Eqs. (3.11) and (3.13) is just Eq. (3.10), as may be verified directly.

## 4. CONCLUSIONS

The equations considered in Sec. 3 do not, at first sight, seem to be connected. The factorization method, however, brings out the relationship between them quite simply, and can further be used to derive a Backlund transformation[10] between Eqs. (3.1a) and (3.1b).

Using these methods,[11] we have also studied the general scalar third-order differential operator

$$B = \partial^3 + u\partial + w, \tag{4.1}$$

for which the Boussinesq equation[12] is an isospectral flow.

Further generalizations, including the case when the scattering operator is of higher order, are being considered.


## ACKNOWLEDGMENTS

We would like to thank B.A. Kupershmidt for bringing Eq. (3.1b) to our attention, and J.D. Gibbon for several useful discussions on these equations.



[1] R. Abraham and J.E. Marsden, *Foundations of Mechanics*, 2nd edition, (Benjamin, London, 1978), p. 465.
[2] M. Adler and J. Moser, Communs. Math. Phys. **61**, 1 (1978).
[3] R.M. Miura, J. Math. Phys. **9**, 1202 (1968).
[4] M. Wadati, J. Phys. Soc. Jpn., **32**, 1681 (1972).
[5] R.M. Miura, SIAM Rev. **18**, 412 (1976), and references therein.
[6] C.S. Gardner, J. Math. Phys. **12**, 1548 (1971); V.E. Zakharov and L.D. Faddeev, Func. Anal. Appl. **5**, 280 (1972).
[7] K. Sawada and T. Kotera, Prog. Theor. Phys. **51**, 1355 (1974).
[8] P.J. Caudrey, R.K. Dodd and J.D. Gibbon, Proc. R. Soc. Lond. A **351**, 407 (1976); R.K. Dodd and J.D. Gibbon, Proc. R. Soc. Lond. A **358**, 287 (1977).
[9] B.A. Kupershmidt, private communication.
[10] A.P. Fordy and J. Gibbons, Phys. Lett. **A75**, 325 (1980).
[11] A.P. Fordy and J. Gibbons, DIAS preprint.
[12] V.E. Zakharov and A.B. Shabat, Func. Anal. Appl. **8**, 226 (1974); H.C. Morris, J. Math. Phys. **17**, 1867 (1976).

# A modified Bars–Durgut equation with polynomial eigenfunctions

E. Onofri
*Istituto di Fisica dell'Università di Parma, 43100 Parma, Italy*
*and*
*I.N.F.N., Sezione di Milano, Milano, Italy*

(Received 8 January 1980; accepted for publication 30 May 1980)

A singular integral equation in two degrees of freedom is defined. Its structure is similar to Bars–Durgut equation for baryons in two dimensions. It admits polynomial eigenfunctions and its spectrum can be studied exactly. A comparison with numerical data available for the baryon equation shows strong similarities. The asymptotic behavior of the eigenvalues for high quantum numbers is studied in the semiclassical approximation and it is found to be in good agreement with the exact spectrum. A peculiar feature of this model is the presence of a transition from a region of periodic classical orbits with constant frequency (straight Regge trajectories in the spectrum) to a regime of aperiodic orbits (nearly parabolic trajectories).

## INTRODUCTION

A singular integral equation for the bound states of three quarks (baryons) in two space-time dimensions was introduced by Bars[1] in the context of the quantum string theory and later derived by Durgut[2] as a planar approximation of QCD. Webber[3] gave a numerical solution of this equation in the special case of vanishing quark masses.

The object of the present paper is to introduce a modified equation which admits polynomial eigenfunctions and whose spectrum can be studied exactly. The interest of such an equation is twofold. First of all, this soluble model can be taken as a rough approximation of the Bars–Durgut equation: it is plausible that this latter can be studied more easily in the basis where our modified operator is diagonal. Our operator may also provide a "reference" operator in terms of which one could prove that the Bars–Durgut operator has a compact inverse, hence a purely discrete spectrum, as was done by Federbush and Tromba[4] for 't Hooft's meson equation. Secondly, our model has some mathematical interest of its own: it gives an example of a pseudodifferential operator whose classical analogue (principal symbol) exhibits a transition from a regime of periodic orbits with constant frequency to another regime of aperiodic orbits. Our analysis shows that the asymptotic spectrum is similarly characterized by two regimes, one with straight Regge trajectories and one with parabolic trajectories. Both regimes can be understood in terms of a semiclassical approximation, but a detailed description fo the boundary layer between the two regions seems to be beyond the reach of the semiclassical approximation.

The paper is organized as follows. In Sec. 1 the modified integral equation is introduced and it is proved that it admits a basis of polynomial eigenfunctions. It is then shown that the equation is equivalent to the eigenvalue equation for the operator $\frac{1}{2}|M_1| + \frac{1}{2}|M_2| + \frac{1}{2}|M_3|$, where $M_i$ are the components of the standard (orbital) angular momentum operator acting on a suitable subset of harmonic polynomials. This fact provides a simple group-theoretical method to calculate the spectrum, which turns out to be qualitatively very close to that of the Bars–Durgut equation (Sec. 2). The asymptotic behavior of eigenvalues has been explored numerically (we computed the first 2556 eigenvalues); this is discussed in Sec. 3A. The semiclassical approximation is derived in Sec. 3B and compared with the exact spectrum. Details about the calculation of the action variables are given in the Appendix.

All the numerical work was done by my colleague and friend Mario Casartelli on the CDC7600 of CINECA (Bologna). A relevant portion of this paper has been made possible by his generous collaboration.

## 1. THE MODEL

### A. The modified Bars–Durgut equation

The Bars–Durgut equation describing three-quark bound states can be written as follows:

$$\left(\frac{\alpha_1}{x}+\frac{\alpha_2}{y}+\frac{\alpha_3}{z}\right)\phi(x,y,z) - \frac{1}{2}\left(\frac{\partial}{\partial x}\right)_z \fint_0^{1-z} \frac{\phi(\omega,1-\omega-z,z)}{\omega-x}\,d\omega - \frac{1}{2}\left(\frac{\partial}{\partial y}\right)_x \fint_0^{1-x} \frac{\phi(x,\omega,1-\omega-x)}{\omega-y}\,d\omega - \frac{1}{2}\left(\frac{\partial}{\partial z}\right)_y \fint_0^{1-y} \frac{\phi(1-\omega-y,y,\omega)}{\omega-z}\,d\omega = \mu^2\phi(x,y,z), \tag{1}$$

where $x, y, z$ are confined to the triangle $x\geqslant 0, y\geqslant 0, z\geqslant 0$, $x+y+z=1$; $(\partial/\partial x)_z$ means partial derivative at $z$ constant; $\fint$ denotes Cauchy principal parts; $\mu^2$ is the eigenvalue (squared baryon mass in units $8g^2/3\pi$). $\phi(x,y,z)$ must vanish on the boundary (except for $\alpha_1=\alpha_2=\alpha_3=-1$) and its norm is defined by

$$\|\phi\|^2 = \fint dx\,dy\,dz\,\delta(x+y+z-1)\,|\phi(x,y,z)|^2. \tag{2}$$

We shall study the following modified equation

$$(S\psi)(x,y,z) = (s+3)\psi(x,y,z), \tag{3}$$

where

$$S = S_1 + S_2 + S_3 \tag{4}$$

$$(S_1\psi)(x,y,z) = -\frac{1}{\pi}\sqrt{yz}\left(\frac{\partial}{\partial y}\right)_x \fint_0^{1-x} \frac{\psi(x,\omega,1-\omega-x)}{\omega-y}\,d\omega \tag{4a}$$

 0022-2488/80/102511-10$1.00  

$$(S_2\psi)(x,y,z)$$

$$= -\frac{1}{\pi}\sqrt{zx}\left(\frac{\partial}{\partial z}\right)_y \fint_0^{1-y} \frac{\psi(1-\omega-y,y,\omega)}{\omega - z}\, d\omega \tag{4b}$$

$$(S_3\psi)(x,y,z)$$

$$= -\frac{1}{\pi}\sqrt{xy}\left(\frac{\partial}{\partial x}\right)_z \fint_0^{1-z} \frac{\psi(\omega,1-\omega-z,z,)}{\omega - x}\, d\omega, \tag{4c}$$

with the same conventions as above, except for the new $L_2$–norm

$$\|\psi\|^2 = \frac{1}{2\pi}\int \frac{dx\,dy\,dz}{(xyz)^{1/2}}\,\delta(x+y+z-1)\,|\psi(x,y,z)|^2. \tag{5}$$

(The eigenvalue is denoted by $s+3$ in order to have $s=0$ as the ground state). It is easily shown that $S$ is Hermitian and positive definite (as a quadratic form) in the domain $\bar{D}_S$ which is the (Friedrich) extension of

$$D_S = \{\psi | \psi = (xyz)^{1/2} \times \text{polynomial}\}.$$

In fact, by writing $\psi = (xyz)^{1/2}\sigma(x,y,z)$, we find

$$(\psi, S_3\psi)$$

$$= \|\psi\|^2 + \frac{1}{4\pi}\int_0^1 dz\,\sqrt{z}\int_0^{1-z} d\omega \sqrt{\omega(1-\omega-z)}$$
$$\times \int_0^{1-z} d\omega' \sqrt{\omega'(1-\omega'-z)}\,\frac{1}{\pi^2}$$
$$\times\left[\frac{\sigma(\omega,1-\omega-z)-\sigma(\omega',1-\omega'-z,z)}{\omega-\omega'}\right]^2$$
$$\geqslant \|\psi\|^2, \tag{6}$$

and analogous relations are valid for $S_1$ and $S_2$. Since $S$ is bounded from below, a self-adjoint extension of $S$ exists (Friedrich's extension). The following theorem essentially says that we do not have to worry too much about functional-analytic details.

*Theorem* 1: $S$ admits a basis of eigenfunctions of the form $\psi(x,y,z) = (xyz)^{1/2}P(x,y,z)$, with $P$ a homogeneous polynomial.

*Proof*: we shall make use of the well-known identity

$$\frac{1}{\pi}\fint_0^1 \frac{dy}{y-x}\,[y(1-y)]^{1/2}y^p$$
$$= -\sum_{k=0}^{p+1}\binom{\frac{1}{2}}{k}(-)^k x^{p+k-1}, \tag{7}$$

which can be easily proven by contour deformation in the complex $y$–plane. Now we shall prove that $(xyz)^{-1/2}S(xyz)^{1/2}P(x,y,z)$ is a polynomial of degree not exceeding the degree of $P$. To do this it is sufficient to consider the special case $P = x^a y^b z^c$ and to calculate the action of $S_3$, the other two terms being obtained by permutation of variables. We have

$$S_3(xyz)^{1/2}x^a y^b z^c$$

$$= -\frac{1}{\pi}\sqrt{xy}\left(\frac{\partial}{\partial x}\right)_z \fint_0^{1-z} \frac{d\omega}{\omega - x}\,\omega^{a+1/2}$$
$$\times (1-\omega-z)^{b+1/2} z^{c+1/2}$$

$$= -\frac{1}{\pi}\sqrt{xyz}\,z^c(1-z)^{a+b+1}\left(\frac{\partial}{\partial x}\right)_z$$
$$\times \fint_0^1 \frac{d\xi\,[\xi(1-\xi)]^{1/2}}{\xi - x(1-z)^{-1}}\,\xi^a(1-\xi)^b, \tag{8}$$

where we have defined $\omega = (1-z)\xi$. By Eq. (7), the integral in Eq. (8) is given by a polynomial of degree $a+b+1$ in the variable $x(1-z)^{-1}$. The factor $(1-z)^{a+b+1}$ reduces it to a homogeneous polynomial of degree $a+b+1$ in $x$ and $(1-z)$. The derivative with respect to $x$ gives a homogeneous polynomial of degree $a+b$ and the final result is a homogeneous polynomial of degree $a+b+c$ times $(xyz)^{1/2}$. Since $S$ is Hermitian it follows that it is diagonalizable in each subspace, $\{(xyz)^{1/2}P_n(x,y,z)\}$, with $P_n$ homogeneous of degree $n$ ($n = 0,1,2,\ldots$).

### B. Connection with orbital angular momentum

The eigenvalue problem for $S$ is now reduced to an *algebraic* one. The degree $n$ is a "good quantum number," while it is only an "approximate" one for the Bars–Durgut equation, as found by Webber.[3] However we have to take into account the fact that $x$, $y$ and $z$ are not independent, which establishes an equivalence relation $P \sim Q$ if $P-Q$ vanishes for $x+y+z=1$. In Ref. 3 the choice was made to eliminate one variable as a function of the others, but this destroys the manifest cyclic symmetry of the triangle. We shall instead make a change of variables which will uncover a simpler geometrical description of $S$ and will provide a straightforward method for its diagonalization.

Let us map the triangle $x+y+z=1$ onto an octant of the unit sphere by introducing new variables $x_1, x_2, x_3$ as follows:

$$\begin{cases} x = x_1^2 = (\sin\theta\cos\varphi)^2 \\ \qquad\qquad 0 \leqslant \theta \leqslant \frac{1}{2}\pi \\ y = x_2^2 = (\sin\theta\sin\varphi)^2 \\ \qquad\qquad 0 \leqslant \varphi \leqslant \frac{1}{2}\pi. \\ z = x_3^2 = \cos^2\theta \end{cases} \tag{9}$$

The norm of $\psi$ is given by

$$\|\psi\|^2 = \frac{1}{2\pi}\int \frac{dx\,dy}{[xy(1-x-y)]^{1/2}}\,|\psi(x,y,1-x-y)|^2$$
$$= \frac{2}{\pi}\int_0^{\pi/2}\sin\theta\,d\theta\int_0^{\pi/2} d\varphi\,|\Psi(\theta,\varphi)|^2, \tag{10}$$

which means that $\Psi$ is normalized with respect to the ordinary Lebesgue measure on the first octant.

Let us calculate the action of $S_3$ in the new variables; since $(\partial/\partial x)_z = -(\sin 2\varphi \sin^2\theta)^{-1}\partial/\partial\varphi$, we have

$$(S_3\Psi)(\theta,\varphi)$$

$$= -\frac{1}{2\pi}\frac{\partial}{\partial\varphi}$$
$$\times \fint_0^{\sin^2\theta} \frac{\psi(\eta,1-\eta-\cos^2\theta,\cos^2\theta)}{\eta - \sin^2\theta\cos^2\varphi}\,d\eta$$
$$= \frac{1}{2\pi}\frac{\partial}{\partial\varphi}$$
$$\times \fint_0^1 \frac{\Psi(\theta,\varphi')}{\cos^2\varphi' - \cos^2\varphi}\,d(\cos^2\varphi'), \tag{11}$$

having defined $\eta = \sin^2\theta \cos^2\varphi$. By introducing $\xi = \cos 2\varphi$, $\zeta = \cos 2\varphi'$, Eq. (11) is finally reduced to the form

$$(S_3\Psi)(\theta,\varphi) = -(1-\xi^2)^{1/2}\frac{\partial}{\partial\xi}\int_{-1}^{1}\frac{\Psi(\theta,\frac{1}{2}\cos^{-1}\zeta)}{\zeta-\xi}\,d\zeta/\pi, \quad (12)$$

with $\Psi$ vanishing at $\xi = \pm 1$. Eq. (12) is well known to admit the Chebyshev polynomials of the II kind as eigenfunctions.[5,6] It follows that $S_3$ satisfies the eigenvalue equation

$$S_3[\lambda(\theta)\sin(2(n+1)\varphi)] = (n+1)[\lambda(\theta)\sin(2(n+1)\varphi)], \quad (13)$$

with arbitary $\lambda(\theta)$. We can then identify $S_3$ with the pseudodifferential operator $\frac{1}{2}|\partial/\partial\varphi| = \frac{1}{2}[-(\partial/\partial\varphi)^2]^{1/2}$ with vanishing boundary conditions at $\varphi = 0$ and $\varphi = \frac{1}{2}\pi$. Since $-i\,\partial/\partial\varphi$ is the angular momentum operator $M_3 = +i(x_2\,\partial/\partial x_1 - x_1\,\partial/\partial x_2)$, it is obvious by symmetry that we shall find:

*Theorem* 2: The singular integral operator $S$ is unitarily equivalent to the pseudodifferential operator $\frac{1}{2}|M_1| + \frac{1}{2}|M_2| + \frac{1}{2}|M_3|$ acting on the unit sphere with vanishing boundary conditions for $x_1x_2x_3 = 0$.

Now, by Theorem 1 we know that the eigenfunctions of $S$ are of the form

$$\Psi(x) = x_1x_2x_3P(x_1^2,x_2^2,x_3^2). \quad (14)$$

It is then convenient to consider the operator $S$ defined on the whole unit sphere, provided that we restrict its domain to the linear manifold $\mathscr{B} \subset L_2(S^2)$ spanned by polynomials of this kind. Let us observe that $\mathscr{B}$ is not invariant under rotations, but it is invariant under $M^2 = M_1^2 + M_2^2 + M_3^2$ which commutes with $S$. We can then find a common orthonormal basis by requiring $\Psi$ to be harmonic. We now recognize that the "good quantum number" $n$ can be interpreted in terms of angular momentum. The subspace of harmonic polynomials of the type given in Eq. (14) with $P$ a polynomial of degree $n$ corresponds to $M^2 = l(l+1)$ with $l = 2n+3$. We shall see later that the Bars–Durgut Hamiltonian can also be expressed in terms of angular momentum operators, but it contains also the coordinates $x_i$, and it does not commute with $M^2$.

TABLE I. $O_h$ multiplicities in $\mathscr{B} \subset \mathscr{D}^{(l)}$.

| $n$ | $l$ | $\mu(S\|l)$ | $\mu(A\|l)$ | $\mu(M\|l)$ |
|---|---|---|---|---|
| 0 | 3 | 1 | 0 | 0 |
| 1 | 5 | 0 | 0 | 1 |
| 2 | 7 | 1 | 0 | 1 |
| 3 | 9 | 1 | 1 | 1 |
| 4 | 11 | 1 | 0 | 2 |
| 5 | 13 | 1 | 1 | 2 |

$$\mu(S|l+12) = \mu(S|l) + 1$$
$$\mu(A|l+12) = \mu(A|l) + 1$$
$$\mu(M|l+12) = \mu(M|l) + 2$$

## 2. GROUP THEORETICAL CALCULATION OF THE SPECTRUM

A good deal of information about the spectrum can be obtained by symmetry arguments. The operator $S = \frac{1}{2}\Sigma|M_i|$ breaks rotational invariance, but it has nonetheless some symmetry; precisely the full rotation group O(3) is reduced to the discrete symmetry $O_h$ of the octahedron. It follows that each subspace $\mathscr{D}^{(l)}$ of total angular momentum $l$ must be reduced with respect to $O_h$ and only the representations contained in $\mathscr{B}$ [i.e., compatible with Eq. (14)] retained. This is an elementary problem in group theory (see Ref. 7) which was solved by Bethe fifty years ago and applied to the splitting of atomic levels in a cubic crystal. First of all we notice that only the representations $A_{1u}$, $A_{2u}$ and $E_u$ (in chemists' notation) are compatible with Eq. (14). In fact, the reflections $\sigma_1(x \to -x)$, $\sigma_2(y \to -y)$ and $\sigma_3(z \to -z)$ are represented by minus the identity in $\mathscr{B}$, which rules out all other representations. The dimension $2l+1$ of $\mathscr{D}^{(l)}$ is reduced to $\frac{1}{2}(l-1)$ (recall that only odd values of $l$ are allowed in $\mathscr{B}$). Actually $O_h$ is "too big" for $S$, the maximal symmetry group being in fact the permutation group in three objects. What we gain in considering $O_h$ is that the reduction from O(3) to $O_h$ can be readily found in the literature. The multiplicities $\mu(\cdot|l)$ of $A_{1u}$, $A_{2u}$ and $E_u$ can now be found by standard methods (see Table I). *These are actually the same multiplicities* found by Webber for the Bars–Durgut equation, if we identify $A_{1u} \equiv A$, $A_{2u} \equiv S$, $E_u \equiv M$, with notations of Ref. 3 (which we shall adopt from now on).

Now we can calculate the first few eigenvalues of $S$, with almost no more effort. In fact, let $\Psi = \Sigma a_k Y_k^l$ belong to the representation $S$ or $A$ and suppose that the representation has multiplicity *one* in $\mathscr{D}^{(l)}$. By Schur's lemma $\Psi$ is an eigenfunction of $S$ and the eigenvalue $s+3$ is calculated as follows:

$$s+3 = \frac{(\Psi,S\Psi)}{(\Psi,\Psi)} = \frac{1}{2}\sum_i\frac{(\Psi,|M_i|\Psi)}{(\Psi,\Psi)} = \frac{3}{2}\frac{(\Psi,|M_3|\Psi)}{(\Psi,\Psi)} = \frac{3}{2}\sum|k|\,|a_k|^2\Big/\sum|a_k|^2. \quad (15)$$

If $\Psi_1$, $\Psi_2$ belong to $M$ and $\mu(M|l) = 1$, $\Psi_1$ and $\Psi_2$ are degenerate eigenfunctions and the eigenvalue can be found by taking half the trace of $S$. In this way we find the first few eigenvalues as reported in Table II. The basis functions were taken from Bradley and Cracknell.[8] The case $n = 4$ already needs some deeper analysis, because $M$ occurs twice.

The spectrum calculated so far is already sufficient to allow a comparison with Webber's numerical results [see

TABLE II. The low energy spectrum of $S$ obtained through the reduction $O(3)/O_h$.

| $n$ | $l$ | $s$ | $a$ | $b$[¶] |
|---|---|---|---|---|
| 0 | 3 | 0 | S | $\|3,1\rangle$ |
| 1 | 5 | 3/2 | M | $\|5,1\rangle$ ; $\|5,2\rangle$ |
| 2 | 7 | 22/8 | S | $\cos\alpha\ \|7,1\rangle + \sin\alpha\ \|7,3\rangle$ |
| | | 25/8 | M | $-\sin\alpha\ \|7,1\rangle + \cos\alpha\ \|7,3\rangle$ ; $\|7,2\rangle$ |
| 3 | 9 | 67/16 | M | $\cos\beta\ \|9,2\rangle + \sin\beta\ \|9,4\rangle$ ; $\cos\gamma\ \|9,1\rangle + \sin\gamma\ \|9,3\rangle$ |
| | | 76/16 | A | $-\sin\beta\ \|9,2\rangle + \cos\beta\ \|9,4\rangle$ |
| | | 78/16 | S | $-\sin\gamma\ \|9,1\rangle + \cos\gamma\ \|9,3\rangle$ |
| 4 | 11 | 694/128 | S | $\rho\ \|11,1\rangle + \sigma\|11,3\rangle + \tau\|11,5\rangle$ |
| | | 736/128 | M | . . . |
| | | 837/128 | M | . . . |

¶ $|l,k\rangle = \sqrt{\frac{1}{2}}\,(Y^l_{2k} - Y^l_{-2k})$ ; $\cos\alpha = \sqrt{13/24}$ ; $\cos\beta = \sqrt{7/24}$ ; $\cos\gamma = \sqrt{13/16}$ ;

$\sigma = \sqrt{81/384}$ ; $\tau = \sqrt{133/384}$ ; $\rho = \sqrt{170/384}$ ;

$a$: symmetry type; $b$: eigenfunctions.

Figs. 1(a) and 1(b) which contain data from Ref. 3; eigenvalues with $n \geqslant 5$ are taken from Table III of next section]. We notice that the spectrum can be interpreted in terms of "Regge trajectories" which connect eigenvalues belonging to the same $O_h$ representation; only even or odd values of $n$ belong to a given trajectory. From this point of view the structure of the spectrum is very similar to that of the Bars–Durgut equation. We believe that our model should be an even better approximation for the baryon equation with no potential term ($\alpha_i = 0$), but we do not have explicit results in this direction. The only qualitative difference is given by a lower spreading of multiplets in our model, which means that our operator breaks O(3) invariance to a lesser extent; as a consequence there is no crossing of levels $4S < 3A$. Let us observe that the Bars–Durgut equation can also be studied on the sphere $S^2$. In terms of angular momentum operators it is given by

$$\left(\frac{\alpha_1}{x_1^2} + \frac{\alpha_2}{x_2^2} + \frac{\alpha_3}{x_3^2}\right)\phi(x_1,x_2,x_3) + (x_1x_2x_3)^{-1/2}\sum x_i\,|M_i|\,(x_1x_2x_3)^{-1/2}\phi = \mu^2\phi. \quad (16)$$

An approximate solution of this equation starting on this representation is now being tried. Results (if any) will be reported elsewhere.

## 3. THE ASYMPTOTIC SPECTRUM

### A. Numerical calculation of the spectrum

To calculate higher eigenvalues group theory is not sufficient any more, because of multiplicities higher than one. We shall then resort to a numerical calculation which is based on the following preliminaries:

(i) an orthonormal basis in $\mathscr{H}$ is given by

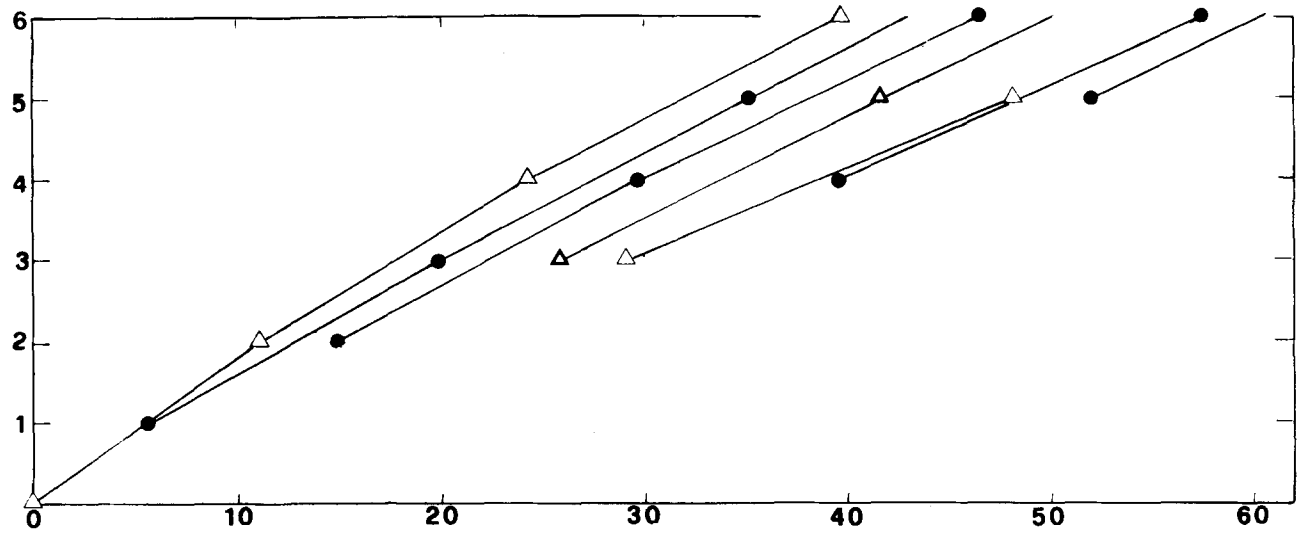
$$\mathscr{Y}^l_k = \frac{1}{\sqrt{2}}(Y^l_{2k} - Y^l_{-2k}), \quad l = 2n+3, \quad 1 \leqslant k \leqslant n+1; \quad (17)$$

(ii) $M_i^2$ are represented by the matrices

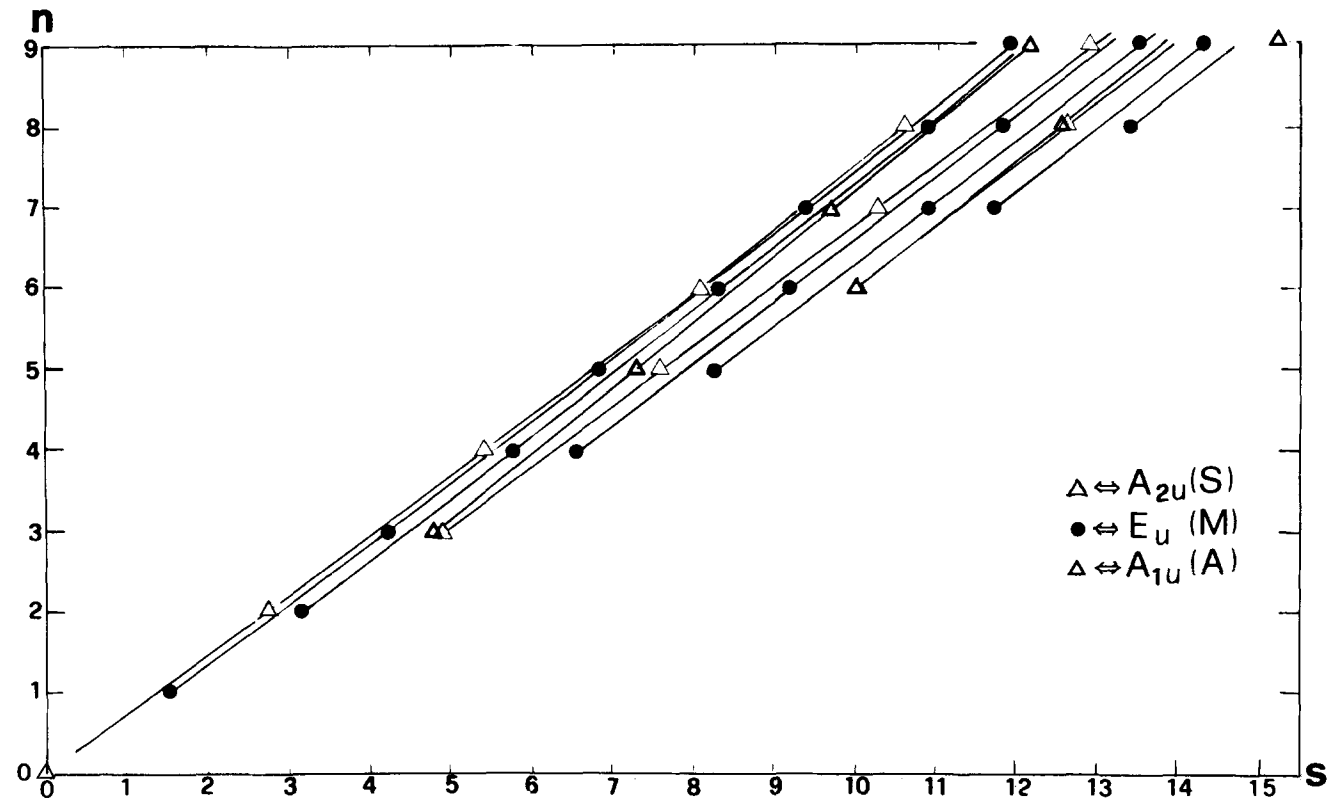
$$(\mathscr{Y}^{l'}_{k'}, M_3^2 \mathscr{Y}^l_k) = 4\delta_{ll'}\delta_{kk'}k^2 \quad (18a)$$

$$(\mathscr{Y}^{l'}_{k'}, M_1^2 \mathscr{Y}^l_k) = \tfrac{1}{2}\delta_{ll'}\delta_{k',k+1}\,[\tfrac{1}{2}l(l+1) - k(2k+1)\,] \times [\tfrac{1}{2}l(l+1) - (k+1)(2k+1)\,] + (k \longleftrightarrow k') + \tfrac{1}{2}[\,l(l+1) - 4k^2\,]\delta_{ll'}\delta_{kk'}. \quad (18b)$$

($M_2^2$ has the same diagonal as $M_1^2$ and opposite off-diagonal terms). We omit the proofs of (i) and (ii), which are elementary. The problem is now reduced to a purely algebraic one, which can be easily solved numerically. This was done on a computer up to $n = 70$.[9] For lack of space we report only the eigenvalues for $n \leqslant 12$ in Table 3 and a typical example for high $n$ in Table 4. The numerical calculation seems to be very accurate; no instabilities arise even for $n = 70$. A simple check is given by comparing the average computed eigenvalue with the exact average: $\langle \tfrac{1}{2}\Sigma M_i \rangle = \tfrac{3}{2}\langle\,|M_3|\,\rangle = 3(n+2)/2$. The numerical value agrees with this to all




(a)




(b)

FIG. 1.(a) Low energy spectrum of Bars–Durgut equation (from Ref. 3). (b) Low energy spectrum of $S$.

significant figures (see Table IV).

In columns A and B of Table IV the eigenvalues are written in units of $\frac{1}{2}\sqrt{3}$. This shows that the highest eigenvalues are almost equally spaced with a gap $\frac{1}{2}\sqrt{3}$. This is easy to understand: these states are very sharply peaked along the direction (1,1,1) in $M$ space, which implies that $S$ cannot be distinguished from $\frac{1}{2}\sqrt{3}|M\cdot\hat{n}|$ [where $\hat{n} = 3^{-1/2}(1,1,1)$]. Going to lower eigenvalues this is no longer true and the spectrum is more difficult to understand. We notice that the symmetry type of the eigenstates changes periodically as *S-M-M-A-S-M-M-A-S*····. Neighbor *A-S* states tend to become degenerate when the eigenvalue is high enough (for a given $n$) while for low eigenvalues there is a tendency to form almost degenerate *S-M* or *M-A* triplets. To have an idea of the

TABLE III. The eigenvalues of $S$ up to $n = 12$.

| $n$ | $l$ | $s$ | |
|---|---|---|---|
| 0 | 3 | 0 | S |
| 1 | 5 | 1.5 | M |
| 2 | 7 | 2.75 | S |
| | | 3.125 | M |
| 3 | 9 | 4.1875 | M |
| | | 4.75 | A |
| | | 4.875 | S |
| 4 | 11 | 5.421875 | S |
| | | 5.75 | M |
| | | 6.5390625 | M |
| 5 | 13 | 6.8192335 | M |
| | | 7.2656250 | A |
| | | 7.5703125 | S |
| | | 8.2627977 | M |
| 6 | 15 | 8.0375448 | S |
| | | 8.3242959 | M |
| | | 9.1649619 | M |
| | | 9.9843750 | A |
| | | 9.9995646 | S |
| 7 | 17 | 9.3996024 | M |
| | | 9.7480469 | A |
| | | 10.2646484 | S |
| | | 10.8708908 | M |
| | | 11.7231592 | M |
| 8 | 19 | 10.6027029 | S |
| | | 10.8464670 | M |
| | | 11.8036489 | M |
| | | 12.5664062 | A |
| | | 12.6211130 | S |
| | | 13.4547730 | M |
| 9 | 21 | 11.9312363 | M |
| | | 12.2011545 | A |
| | | 12.9499258 | S |
| | | 13.4827015 | M |
| | | 14.3238705 | M |
| | | 15.1858694 | A |
| | | 15.1874338 | S |
| 10 | 23 | 13.1211911 | S |
| | | 13.3231967 | M |
| | | 14.4466523 | M |
| | | 15.1358032 | A |
| | | 15.2581500 | S |
| | | 16.0539443 | M |
| | | 16.9186345 | M |
| 11 | 25 | 14.4186363 | M |
| | | 14.6282278 | A |
| | | 15.6208142 | S |
| | | 16.0926378 | M |
| | | 16.9283504 | M |
| | | 17.7814478 | A |
| | | 17.7889491 | S |
| | | 18.6506560 | M |
| 12 | 27 | 15.5978308 | S |
| | | 15.7630009 | M |
| | | 17.0840932 | M |
| | | 17.6908998 | A |
| | | 17.9085642 | S |
| | | 18.6546207 | M |
| | | 19.5169432 | M |
| | | 20.3826171 | A |
| | | 20.3827723 | S |

structure of the whole spectrum see Fig. 2: a 45° oblique axis has been chosen in order to be able to draw the spectrum up to $n = 50$.

The general tendency of eigenvalues for growing $n$ is to arrange themselves into a regular lattice on the right of the diagram, with a lattice spacing $\frac{1}{2}\sqrt{3}$. In the other region (on the left) there is a tendency to form degenerate trajectories *S-M-M-A* which are approximately parabolic. *Any* given Regge trajectory starts as a straight line with a slope $4\sqrt{3}/g$ (referred to orthogonal axes) but is bound to merge with other trajectories and become parabolic. It should also be noticed that distinct trajectoreis never intersect. This asymptotic structure will be (almost completely) understood by the semiclassical approximation. We should stress that the asymptotic spectrum of the Bars–Durgut equation is probably totally different; we expect straight trajectories (for $\alpha_i = 0$) with a slope $[(\sqrt{2})\pi^2]^{-1}$ ($\pi^{-2}$ is the slope for the meson trajectory).

### B. The semiclassical approximation

The classical analogue of our singular integral equation is most simply found from its expression in terms of angular momentum operators. We have to interpret $M_i$ as classical functions on the phase space $T^*S^2$, i.e., the phase space of a point particle constrained to the surface of the unit sphere in $R^3$. We ignore for the moment the presence of boundaries; this will be taken into account later by suitably identifying points on the sphere. The relevant classical dynamical variables are the position $\mathbf{x}$, the momentum $\mathbf{p}$ and the angular momentum $\mathbf{M}$ related by the following equations

TABLE IV. Sample eigenvalues for high $n$.

| $n = 50$ | | $n = 51$ | $A$ | $B$ |
|---|---|---|---|---|
| 59.2640887 | S | | | |
| 59.2789999 | M | | | |
| 63.6473144 | M | 60.3789627 | | |
| 63.6870082 | A | 60.3934048 | | |
| 66.5811683 | S | 64.8005563 | | |
| 66.6648279 | M | 64.8383140 | | |
| 68.7574642 | M | 67.8193050 | | |
| 68.9475348 | A | 67.9001087 | ... | |
| 70.2235470 | S | 69.9569926 | 84.55 | |
| 70.5749412 | M | 70.1251827 | 84.96 | ... |
| 71.5703533 | M | 71.6247572 | 86.106 | |
| 72.2536107 | A | 72.0087148 | 86.895 | 86.413 |
| 72.4639982 | S | 72.6911082 | 87.138 | 87.401 |
| 73.2167246 | M | 73.2356125 | 88.007 | 88.029 |
| 74.0781501 | M | 74.0977934 | 89.002 | 89.025 |
| 74.9392183 | A | 74.9186671 | 89.996 | 89.973 |
| 74.9460459 | S | 74.9733688 | 90.004 | 90.036 |
| 75.8083787 | M | 75.8093200 | 91.00007 | 91.001 |
| 76.6743484 | M | 76.6745518 | 92.00001 | 92.0002 |
| 77.5403366 | A | 77.5399159 | 92.99997 | 92.9995 |
| 77.5403918 | S | 77.5408850 | 93.00003 | 93.0006 |
| 78.4063882 | M | 78.4063942 | 94.0000003 | 94.000007 |
| 79.2724134 | M | 79.2724143 | 95.0000000 | 95.000001 |
| 80.1384387 | A | 80.1384366 | 95.9999999 | 95.999998 |
| 80.1384388 | S | 80.1384411 | 96.0000000 | 96.000003 |
| 81.0044642 | M | 81.0044642 | 97.0000000 | 97.0000000 |
| 81.8704896 | M | 81.8704896 | 98.0000000 | 98.0000000 |
| 82.7365150 | A | 82.7365150 | 99.0000000 | 99.0000000 |
| 82.7365150 | S | 82.7365150 | 99.0000000 | 99.0000000 |
| 83.6025404 | M | 83.6025404 | 100.0000000 | 100.0000000 |
| 84.4685658 | M | 84.4685658 | 101.0000000 | 101.0000000 |
| 85.3345912 | A | 85.3345912 | 102.0000000 | 102.0000000 |
| 85.3345912 | S | 85.3345912 | 102.0000000 | 102.0000000 |
| 86.2006166 | M | 86.2006166 | 103.0000000 | 103.0000000 |
| | M | 87.0666420 | | 104.0000000 |
| | A | 87.9326674 | | 105.0000000 |
| | S | 87.9326674 | | 105.0000000 |

A: $(s+3)/\omega$ , $n=50$ , $\omega = \frac{1}{2}\sqrt{3}$

B: $(s+3)/\omega$ , $n=51$

$$\begin{aligned} &\mathbf{x}\cdot\mathbf{x} = 1 \\ &\mathbf{p}\cdot\mathbf{x} = 0 \\ &\mathbf{M} = \mathbf{x}\wedge\mathbf{p}, \end{aligned} \tag{19}$$

which imply

$$\begin{aligned} &\mathbf{p} = \mathbf{M}\wedge\mathbf{x} \\ &\mathbf{p}\cdot\mathbf{p} = \mathbf{M}\cdot\mathbf{M}. \end{aligned} \tag{20}$$

If we fix $\mathbf{M}$, the point in phase space is identified up to an angle (see Fig. 3).

The Hamiltonian is given by $H_s = \frac{1}{2}\Sigma\,|M_i|$ and there is a constant of the motion $M^2 = \Sigma M_i^2$. The motion must then take place in the intersection of the two surfaces $H_s = E$, $M^2 = l^2$, where the numerical values of $E$ are confined to the range $\frac{1}{2}l \leqslant E \leqslant \frac{1}{2}(\sqrt{3})l$. The energy surface has the shape of an

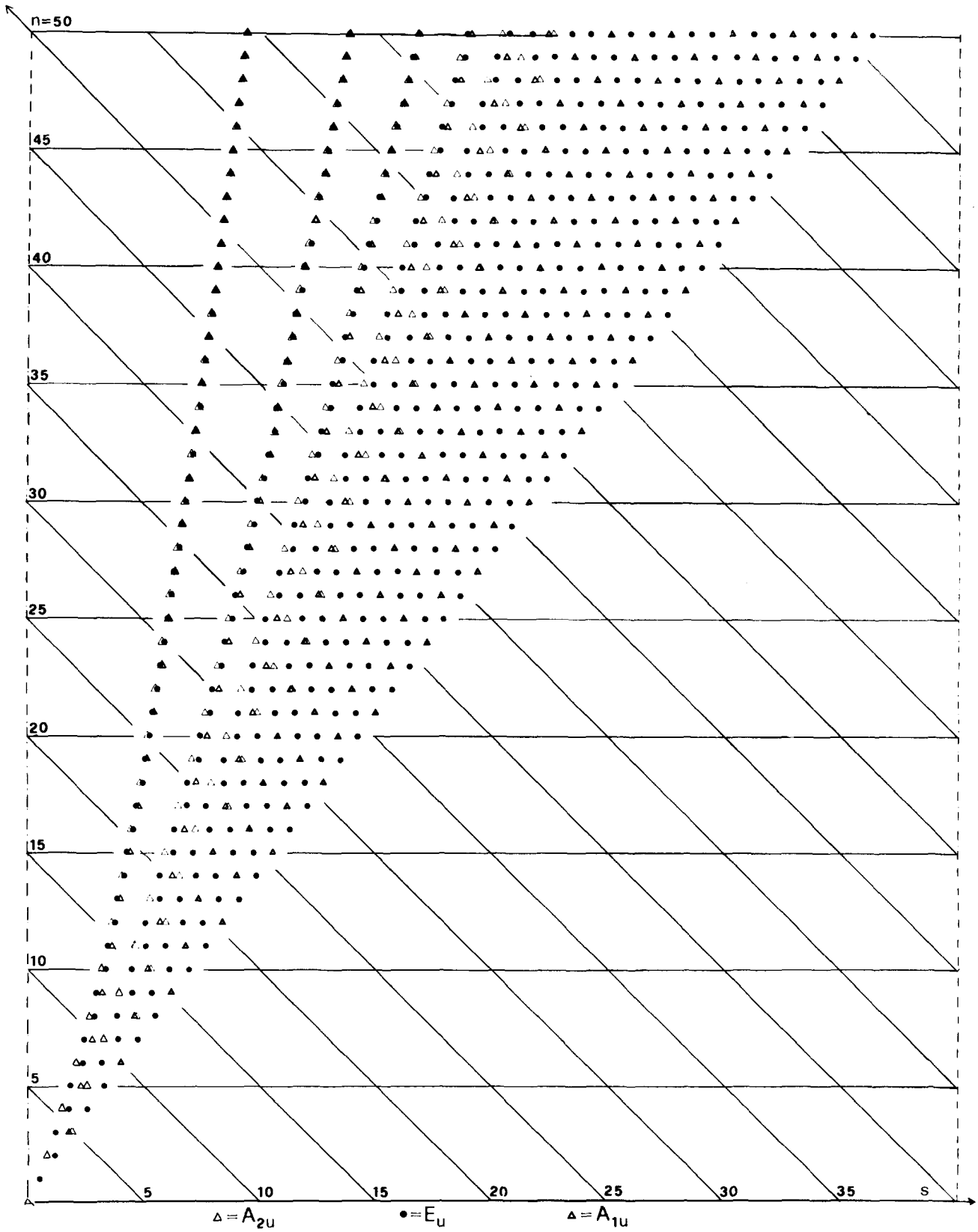



FIG. 2. The spectrum of $S$ up to $n = 50$.

octahedron in $M$ space and its intersection $\Sigma_{E,l}$ with the sphere $M^2 = l^2$ can be of two kinds (see Fig. 4):

(1) $\frac{1}{2}l < E < \frac{1}{2}(\sqrt{2})l$: $\Sigma_{E,l}$ has six components, each connected component being a cusped loop around the coordinate axis;

(2) $\frac{1}{2}l\sqrt{2} < E < \frac{1}{2}l\sqrt{3}$: $\Sigma_{E,l}$ has eight components, each

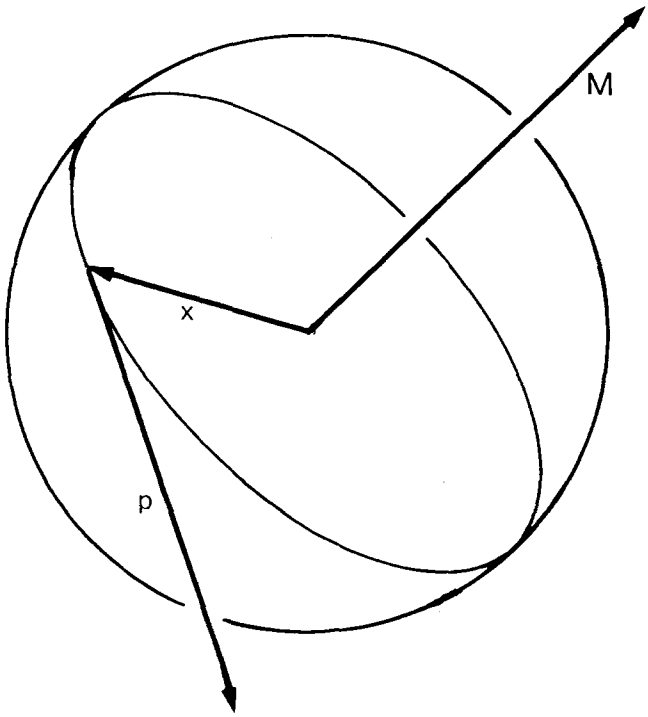



FIG. 3. Classical variables on the unit sphere.

connected component being a circle surrounding one of the directions ( $\pm 1, \pm 1, \pm 1$).

The two regions in phase space corresponding to (1) and (2) have rather different dynamical properties. In region (1) the components of **M** change sign during the motion, while in region (2) they do not. As a consequence in region (2) the Hamiltonian coincides with the component of **M** along one of the directions ( $\pm 1, \pm 1, \pm 1$). The Hamiltonian flow in this region is then given by a uniform rotation along a fixed axis with constant angular velocity $\tilde{\omega} = \frac{1}{2}\sqrt{3}$. The action along a trajectory is easily calculated:

$$\dot{\mathbf{x}} = \tfrac{1}{2}\boldsymbol{\epsilon} \wedge \mathbf{x} \quad (\epsilon_i = \operatorname{sgn} M_i)$$
$$\oint \mathbf{p}\cdot d\mathbf{x} = \int_0^T \mathbf{p}\cdot\dot{\mathbf{x}}\, dt = \tfrac{1}{2}\int_0^T \boldsymbol{\epsilon}\cdot\mathbf{M}\, dt = 2\pi\frac{E}{\tilde{\omega}}, \tag{21}$$

which gives a quantization condition $E = m\tilde{\omega}$ ($m = l, l-1, l-2, \ldots$). This estimate should work for $(2/3)^{1/2} l < m \leqslant l$. In the region where this formula applies we expect degeneracy of energy levels (with same $m$ but different $l$). This is in very good agreement with the observed spectrum: it corresponds to the region of straight Regge trajectories. The part of the spectrum characterized by parabolic trajectories will now be shown to be related to the other region in classical phase space (1). Here the orbits cannot be calculated so simply, because the components $M_i$ change sign; actually we only need to calculate the action variables and this can be done explicitly (see Appendix). The result (taking into account the fact that **x** is confined to the first octant) is the following:

$$J_2 = \frac{4E}{\pi\sqrt{3}} \tan^{-1}\sqrt{3}\,\frac{E - (\frac{1}{2}l^2 - E^2)^{1/2}}{E + 3(\frac{1}{2}l^2 - E^2)^{1/2}} - \frac{2l}{\pi}\tan^{-1}\frac{(E + \frac{1}{2}l)[E - (\frac{1}{2}l^2 - E^2)^{1/2}]}{l^2 + \frac{1}{2}lE - E^2 + (\frac{3}{2}l + E)(\frac{1}{2}l^2 - E^2)^{1/2}}. \tag{22}$$

Bohr–Sommerfeld quantization gives then $l$ = integer and $J_2(E,l) = m + \nu$ ($m = 0,1,2,\cdots$). For very high $l$ we find

$$E_{m,l} = \tfrac{1}{2}l + \tfrac{1}{2}(2\pi(m+\nu)l^{1/2} + \beta_m, \tag{23}$$

which is in excellent agreement with the observed spectrum if $\nu$ and the additive constant $\beta_m$ are chosen appropriately. This quantization condition should work in the range $\frac{1}{2}l < E < \frac{1}{2}l\sqrt{2}$, but it is not clear where exactly it breaks down. While classically there is a sharp transition from region (1) to (2), we do not know exactly where and how the transition takes place in the spectrum. Apparently there is a smooth transition from one region to the other which is triggered by the *A-S* trajectories which at a certain point bifurcate. This is displayed in Fig. 5. Strictly speaking, of course, *A-S* states are never degenerate, and it seems to be a difficult problem to identify the transition point and to understand this picture theoretically. Another problem is to find the theoretical value of the constants entering in Eq. (23): $\nu$ (the "Maslov index") seems to be very close to $5/(2\pi)$. Finally, the semiclassical approximation gives a good description of the average Regge trajectory, but it seems to be unable to give the splitting between *S-M* and *M-A* trajectories. Hope-

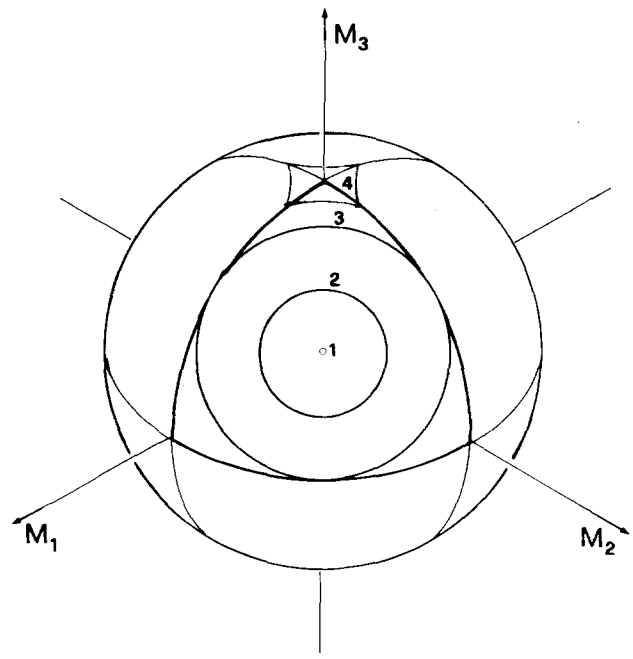



FIG. 4. The intersection of the energy surface with the surface $M^2 = l^2$ as seen from the direction (1,1,1): 1) $E \simeq l\sqrt{3}/2$; 2) $l/\sqrt{2} < E < l\sqrt{3}/2$; 3) $E = l/\sqrt{2}$; 4) $\frac{1}{2}l < E < l/\sqrt{2}$.

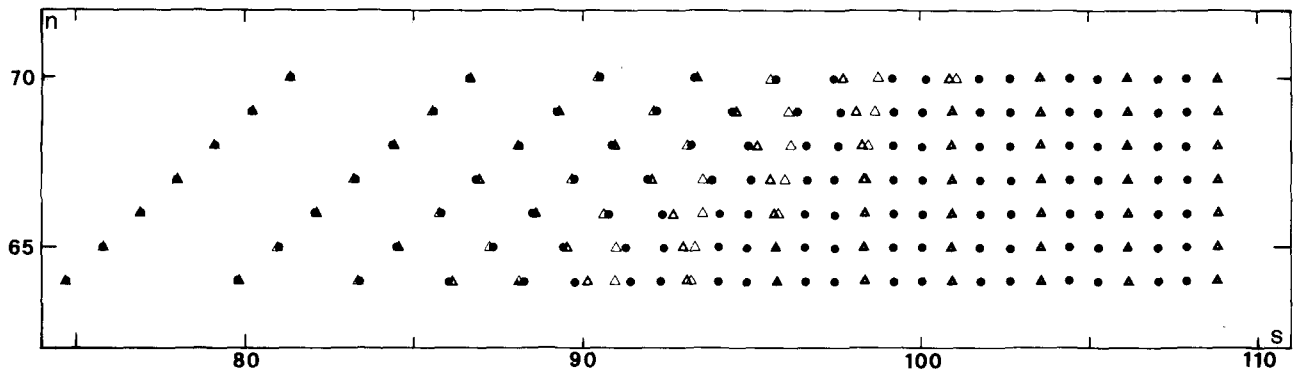



FIG. 5. The transition from straight to parabolic Regge trajectories.

fully the general W.K.B. expansion developed by Voros[10] will enable one to fill some of the gaps left open in this paper. It is hoped that our example will stimulate some progress in this field.

## APPENDIX

We want to calculate the action variables $J_i(E,l)$ for the tori $\Sigma_{E,l} = \{H_s = E, M^2 = l^2\}$ in the region $\frac{1}{2}l < E < \frac{1}{2}l\sqrt{2}$. We identify two independent cycles on the torus $\Sigma_{E,l}$ in this way: the first cycle $\gamma_1$ is defined by keeping **M** fixed and rotating **x** and angle $2\pi$. It is obvious that

$$J_1 = \frac{1}{2\pi}\oint_{\gamma_1} \mathbf{p}\cdot d\mathbf{x} = l. \tag{A1}$$

The second cycle $\gamma_2'$ is defined by varying $M$ along the cusped loop around $M_3$ keeping **x** on the (1-2)-plane; actually $\gamma_2'$ is *not* a fundamental cycle, but it holds $\gamma_2' = \gamma_1^{-1}\gamma_2^2, \gamma_1$ and $\gamma_2$ being a basis of fundamental 1-cycles. This implies that

$$J_2 = \frac{1}{2\pi}\oint_{\gamma_2} \mathbf{p}\cdot d\mathbf{x} = \tfrac{1}{2}\left(J_1 + \frac{1}{2\pi}\oint_{\gamma_2'} \mathbf{p}\cdot d\mathbf{x}\right). \tag{A2}$$

To calculate $J_{2'}$, let $M = (l\sin\beta\cos\alpha, l\sin\beta\sin\alpha, l\cos\beta)$; since we have $x_3 = 0$ on $\gamma_2'$, we find $\mathbf{x} = (\sin\alpha, -\cos\alpha, 0)$ and $\mathbf{p} = \mathbf{M}\wedge\mathbf{x} = (l\cos\beta\cos\alpha, l\cos\beta\sin\alpha, -l\sin\beta)$. The loop $\gamma_2'$ is composed of four arcs, each one giving the same contribution to the action. The integral is calculated by inserting $\mathbf{M}(\theta) = R(\hat{n},\theta)\mathbf{M}(0)$, $\mathbf{M}(0) = (0, l\sin\beta_0, l\cos\beta_0)$, $R(\hat{n},\theta)$ denoting the rotation matrix of an angle $\theta$ around the unit vector $\hat{n} = 3^{-1/2}(1,1,1)$. **x** and **p** are given explicitly as a function of $\theta$ and we have to integrate from $\theta = 0$ to $\theta = \theta_0(E,l)$ which is the smallest $\theta$ for which $M_2(\theta_0) = 0$:

$$\theta_0(E,l) = 2\tan^{-1}\left[\sqrt{3}\sin\beta_0/(2\cos\beta_0 - \sin\beta_0)\right] \tag{A3}$$

$$\beta_0 = \sin^{-1}(\sqrt{2}E/l) - \pi/4. \tag{A4}$$

Some tedious but straightforward algebra leads to

$$J_2' = \frac{4E}{\pi\sqrt{3}}\theta_0(E,l) - \frac{2l\sqrt{3}}{\pi}\int_0^{\theta_0/2} \frac{2\rho - (\frac{3}{2} - 2\rho^2)^{1/2}\cos\eta}{\frac{9}{4} - [\rho + (\frac{3}{2} - 2\rho^2)^{1/2}\cos\eta]^2}\,d\eta \tag{A5}$$

The integration can be performed in closed form to yield the result [Eq. (22)].

*Note added in proof*: The asymptotic behavior of the eigenvalues in the region of parabolic Regge trajectories can be calculated by applying the Holstein–Primakov representation of $M_i$. It turns out that Eq. (23) is good for $m \gg 0$ with $\nu = 3/4$ and $\beta_m = (4m+3)\pi/12 - 1/4$. Also it is probable that the corrected Bohr–Sommerfeld quantization of $J_1$ in Eq. (22) is given by $J_1 = l + \frac{1}{2}$. Details will be given in a paper in preparation.


[1] I. Bars, Phys. Rev. Lett. **36**, 1521 (1976); Nucl. Phys. B **111**, 413 (1976).
[2] M. Durgut, Nucl. Phys. B **116**, 233 (1976).
[3] B. R. Webber, Nucl. Phys. B **153**, 455 (1979).
[4] P. Federbush and A. Tromba, Phys. Rev. D **15**, 2913 (1977).
[5] F. Calogero and A. M. Perelomov, Lett. Nuovo Cimento **23**, 650 (1978).
[6] M. Casartelli, G. Marchesini and E. Onofri, J. Phys. A: Math. Gen. **13**, 1217 (1980).
[7] R. S. Knox and A. Gold, *Symmetry in the Solid State* (Benjamin, New York, 1964).
[8] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[9] In this way we find the symmetry-adapted surface harmonics up to $l = 143$.
[10] A. Voros, Ann. Inst. Henri Poincaré A **26**, 343 (1977).

# Variational principles for particles and fields in Heisenberg matrix mechanics

Abraham Klein, Ching Teh Li,[a] and Moyez Vassanji[b]
*Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

(Received 14 November 1979; accepted for publication 20 June 1980)

For many years we have advocated a form of quantum mechanics based on the application of sum rule methods (completeness) to the equations of motion and to the commutation relations, i.e., to Heisenberg matrix mechanics. Sporadically we have discussed or alluded to a variational foundation for this method. In this paper we present a series of variational principles applicable to a range of systems from one-dimensional quantum mechanics to quantum fields. The common thread is that the stationary quantity is the trace of the Hamiltonian over Hilbert space (or over a subspace of interest in an approximation) expressed as a functional of matrix elements of the elementary operators of the theory. These parameters are constrained by the kinematical relations of the theory introduced by the method of Lagrange multipliers. For the field theories, variational principles in which matrix elements of the density operators are chosen as fundamental are also developed. A qualitative discussion of applications is presented.

## I. INTRODUCTION

A special scheme of calculation of value in a wide class of problems in quantum mechanics and the many-body problem emerges from Heisenberg's matrix mechanics when matrix elements of products of elementary operators appearing in equations of motion and in commutation relations are evaluated by physically chosen approximations to the completeness relation (sum over intermediate states). This method, first introduced for the study of nuclear collective motion,[1] has been applied in recent years to problems in particle quantum mechanics[2-4] and solitons in quantum field theory.[5,6]

Almost from the start of this development, we have been aware that the formulation possessed a variational content, and we attempted some time ago to pin down this content within the nuclear framework.[7] More recently, we have alluded to this aspect on a number of different occasions[3,8,9] and even utilized it as a theoretical tool.[6] In this paper we give a more systematic and thorough account of this property than we have hitherto attempted.[10]

The unifying aspect in this approach is the construction of a stationary expression in which the variational parameters are matrix elements (ultimately between energy eigenstates) of the elementary operators of the theory. In particle quantum mechanics these are the $x$'s and $p$'s, in field theory the particle creation and annihilation operators, though ultimately we show how composite operators may also be utilized. For each case surveyed, we demonstrate that the appropriate stationary expression is the *trace* of the Hamiltonian over the space of states studied, subject to constraints which remind us that the variations of the $x$'s and $p$'s or their equivalents are restricted by their commutation relations. The structure of the variational principle suggests that the natural setting for this approach is indeed that class of problems where the need is to characterize the properties of a special band of states selected from a much larger space (as opposed to a *single* state).

The inclusion of constraints in the variational treatment also suggests that these may be recovered from the variational principles as a bonus by carrying out suitably chosen special variations. This turns out to be the case with carefully stated restrictions. In consequence a complete theory can be extracted from the variational principle.

In the following sections we treat in sequence one-dimensional quantum mechanics (Sec. II), many-particle quantum mechanics (Sec. III), nonrelativistic field theory for bosons (Sec. IV), and nonrelativistic field theory for fermions (Sec. V). Only parts of Secs. II and V have been given previously. The variational principles for the density operators in Secs. IV and V are new results which generalize known (though not widely known) results from Hartree–Fock theory.

In Sec. VI we comment on the classical limit, on applications past and (possibly) future, and on further theoretical possibilities.

## II. QUANTUM MECHANICS—ONE DEGREE OF FREEDOM

We focus attention initially on the system described by the Hamiltonian

$$H = \tfrac{1}{2}p^2 + V(x), \tag{2.1}$$

with equations of motion

$$[x,H] = ip, \tag{2.2}$$

$$[ip,H] = dV/dx \equiv V', \tag{2.3}$$

derived by utilization of the commutation relation ($\hbar = 1$)

$$[x,p] = i. \tag{2.4}$$

In practice,[3] we have been concerned particularly with the matrix elements of (2.2)–(2.4) in the representation in which $H$ is diagonal with eigenvalues $E_n$, namely,

$$(E_n - E_m)x_{mn} = ip_{mn},$$
$$(E_n - E_m)ip_{mn} = (V')_{mn}, \tag{2.5}$$

[a]Present address: Dept. of Physics, Brown University, Providence, Rhode Island 02912.
[b]Present address: Theoretical Physics Branch, Chalk River Nuclear Laboratories, Chalk River, Ontario, KOJIJO, Canada.

 0022-2488/80/102521-07$1.00  

and

$$[x,p]_{nm} = i\delta_{nm}. \tag{2.6}$$

We have shown how the energy differences and the matrix elements $x_{mn}$, $p_{mn}$ can be obtained from Eqs. (2.5) and (2.6). The eigenvalues themselves can be found by the direct evaluation of the expectation values

$$E_n = H_{nn} = \langle n|H|n\rangle = \sum_{n'} \tfrac{1}{2}|p_{nn'}|^2 + \langle n|V(x)|n\rangle, \tag{2.7}$$

using sum rules as illustrated here for the kinetic energy.

A natural question is whether Eqs. (2.5) possess any of the attributes associated with a variational formulation. We wish to treat the matrix elements of $x$ and $p$ as variables in the variational statement $\delta E_n = 0$. There are, however, two obstacles to such an endeavor: (i) The matrix elements are not all independent. (ii) The same matrix elements appear in different energy functionals. Thus $p_{nn'}$ occurs both in $H_{nn}$ and in $H_{n'n'}$. For which is it to be a variational parameter?

A solution to the second problem posed is to form an average of the stationary functionals. In this paper, we shall study only the most symmetrical possible average, namely, the trace. (See Sec. VI for some further comment.) Thus we require

$$\delta \sum_n H_{nn} = \delta \,\mathrm{Tr} H = 0. \tag{2.8}$$

A solution to the first problem is to impose all the possible kinematical constraints, namely,

$$\delta\,[x,p]_{nn'} = 0. \tag{2.9}$$

Multiplying (2.9) by a Lagrange multiplier matrix $(-i)\Lambda_{n'n}$ ($\Lambda$ is Hermitian), we add the result to (2.8) and are thus lead to a master variational principle

$$\begin{aligned}\delta\,&\mathrm{Tr}\{H - i\Lambda\,[x,p]\}\\ &= \delta\,\mathrm{Tr}\{H - ip[\Lambda,x]\}\\ &= \delta\,\mathrm{Tr}\{H + ix[\Lambda,p]\} = 0.\end{aligned} \tag{2.10}$$

The several forms are equivalent because of the assumed cyclic invariance of the trace. (This is certainly unobjectionable in practice where the trace is taken over a finite dimensional vector space.)

Carrying out the unconstrained variation (2.10) with respect to the matrix elements $x_{n'n}$ and $p_{n'n}$, keeping $\Lambda$ fixed, and using the explicit form (2.1) of $H$, we obtain the equations

$$p_{nn'} = -i[x,\Lambda\,]_{nn'}, \tag{2.11}$$

$$(V')_{nn'} = i[p,\Lambda\,]_{nn'}. \tag{2.12}$$

Because of the invariance of the trace with respect to choice of basis, the representation $|n\rangle$ is, at this point, arbitrary. The most convenient immediate choice is the one in which the Hermitian operator $\Lambda$ is diagonal. By comparing with the *known* equations of motion, we then identify $\Lambda$ as the Hamiltonian. Thus we may write (2.10) as

$$\delta\,\mathrm{Tr}\{H - p\dot{x}\} = -\delta\,\mathrm{Tr}L = 0, \tag{2.13}$$

where $L$ is the Lagrange operator. The result is a quantum Hamilton's principle, with unconstrained variation of the coordinates and momenta, which has been derived from a Rayleigh–Ritz principle with constraints.

The structure of the variational expression as a trace suggests an additional inquiry, utilizing the first form of (2.10).[9] We carry out an infinitesimal change of basis, resulting in alterations of commensurate size in the matrix elements of $x,p$ and of $\Lambda$ ($=H$). The variational principle (2.10) implies, however, that in this change of basis only the contribution from explicit variation of $\Lambda$ contributes. Thus we conclude

$$\mathrm{Tr}\{\delta\Lambda\,[x,p]\} = 0. \tag{2.14}$$

The matrix elements $\delta\Lambda_{nn'}$ are, however, themselves not all independent. Let us for the moment suppose that the trace refers to an $N$-dimensional (finite) vector space. Then we know that there are $N$ independent polynomials in the eigenvalues of $\Lambda$ which are invariant under a unitary transformation. We may, for convenience, choose these as $\mathrm{Tr}\Lambda^p$, $p = 1,\ldots,N$ and impose these invariance conditions by subtracting from (2.14) the expression

$$0 = \delta\,\mathrm{Tr}\sum_{p=1}^{N} \frac{1}{p}\lambda_{p-1}\Lambda^p = \mathrm{Tr}\sum_{p=1}^{N} \lambda_{p-1}\Lambda^{p-1}\delta\Lambda, \tag{2.15}$$

where the $\lambda_{p-1}$ are numerical Lagrange multipliers. We deduce that

$$[x,p] = \sum_{p=0}^{N-1} \lambda_p \Lambda^p. \tag{2.16}$$

Passing to the limit $N\to\infty$, we conclude that the commutator is diagonal in the representation in which $\Lambda$ is diagonal, i.e., it is a function of $\Lambda$.

If matters appear to be getting complicated, we can restore simplicity by asking the right question. What additional statements, if any, must be adjoined to the variational principle, with $\Lambda \equiv H$ so that the full dynamical scheme (2.2)–(2.4) is obtained? It suffices to require that the commutator $[x,p]$ be a *kinematical* quantity. Thus the only function of $H$ it can be is then the constant function, i.e., a multiple of the unit operator. The correct multiple, the imaginary unit, is chosen by studying the equations of motion for the free particle, e.g.,

$$[x,\tfrac{1}{2}p^2] = ip = \tfrac{1}{2}p[x,p] + \tfrac{1}{2}[x,p]\,p. \tag{2.17}$$

The desired conclusion depends, however on having *established* that the commutator is a $c$-number.

We finally note a simple alternative to the argument following (2.14). If we make the unitary transformation alluded to above starting from the representation in which $\Lambda = H$ is diagonal, then the additional constraints are simply $\delta H_{nn} = 0$, i.e., the eigenvalues are unchanged. The remaining variations—all off-diagonal elements—may be considered independent and thus we conclude, as in (2.16) that the commutator has only nonvanishing diagonal elements.

## III. QUANTUM MECHANICS—MANY-PARTICLE PROBLEM

As a first generalization of the results of the preceding section, we study $n$ particles described by the Hamiltonian

$$H = \sum_{i=1} \frac{p_i^2}{2m_i} + \frac{1}{2}\sum_{i\neq j} V_{ij}(|x_i - x_j|), \tag{3.1}$$

where we consider motion on the line only to simplify notation. The equations of motion

$$[x_i,H] = ip_i, \tag{3.2}$$

$$[ip_i,H] = \sum_{j\neq i} \partial_i V_{ij}(|x_i - x_j|), \tag{3.3}$$

follow from (2.18) and the commutation relations

$$[x_i,x_j] = [p_i,p_j] = 0, \tag{3.4}$$

$$[x_i,p_j] = i\delta_{ij}. \tag{3.5}$$

We turn to the variational formulation. The first step is to establish a stationary expression with identified Lagrange multipliers. We choose to study the variations of the functional, $F$, given by the expression

$$F = \mathrm{Tr}\{H - i\sum_i \Lambda_i[x_i,p_i] - i\sum_{i\neq j}\Lambda_{ij}[x_i,p_j]\}. \tag{3.6}$$

The constraints (3.4) have been omitted for reasons explained below. Carrying through the required variations utilizing a consistent operator notation, we have,

$$\frac{\delta F}{\delta p_i} = 0 = \frac{p_i}{m_i} - i[\Lambda_i,x_i] - i\sum_{j\neq i}[\Lambda_{ji},x_j], \tag{3.7}$$

$$\frac{\delta F}{\delta x_i} = 0 = \sum_{j\neq i}\partial_i V_{ij}(|x_i - x_j|) + [\Lambda_i,p_i] + i\sum_{j\neq i}[\Lambda_{ij},p_j]. \tag{3.8}$$

We can exhibit three different solutions to (3.7) and (3.8) which satisfy the requirement that they reproduce (3.2) and (3.3):

$$\text{(i)}\ \Lambda_{ij} = 0, \quad \Lambda_i = \Lambda_j = H, \tag{3.9}$$

$$\text{(ii)}\ \Lambda_{ij} = 0, \quad \Lambda_i = \frac{p_i^2}{2m} + \sum_{j\neq i} V_{ij}(|x_i - x_j|), \tag{3.10}$$

$$\text{(iii)}\ \Lambda_i = \frac{p_i^2}{2m}, \quad \Lambda_{ij} = V_{ij}(|x_i - x_j|). \tag{3.11}$$

In addition, any Lagrange multipliers associated with Eq. (3.4) are zero for these solutions.

Further investigation indicates that (3.11) must be disregarded, since the associated terms in $F$, Eq. (3.6), vanish, as one verifies using the cyclic invariance of the trace.

Let us focus on solution (ii). To the now definite functional we adjoin the assumption that the fundamental commutators are kinematic quantities and that only the commutators $[x_i,p_i]$ are nonvanishing. The argument leading to (2.16) can then be generalized simply by placing a subscript $i$ on each quantity involved. The only (temporary) doubt concerning the permissibility of this step is the fact that the various $\Lambda_i$ do not commute and thus their commutators represent additional kinematical constraints. But since we are carrying out a unitary change of basis, these additional constraints are automatically satisfied.

Continuing the argument, we insist that the commutators in question must be $c$-numbers, and we fix the scale from simple examples such as the free-particle case.

If we choose solution (i), we can only prove, by similar reasoning, that $\Sigma_i\ [x_i,p_i]$ is a $c$-number. For identical particles, an additional requirement of permutation symmetry will allow the same final conclusions.

## IV. NONRELATIVISTIC FIELD THEORY. BOSONS

Let $\psi^\dagger(x)$, $\psi(x)$ be boson creation and annihilation operators satisfying the commutation relations

$$[\psi(x),\psi^\dagger(y)] = \delta(x - y). \tag{4.1}$$

As Hamiltonian we consider the structure

$$H = \int \psi^\dagger(x)h(x|y)\psi(y) + \tfrac{1}{2}\int \psi^\dagger(x)\psi^\dagger(x')V(xx'|yy')\psi(y')\psi(y), \tag{4.2}$$

where the functions $h$ and $V$ are real and satisfy

$$V(xx'|yy') = V(x'x|yy') = V(xx'|y'y), \tag{4.3}$$

$$h(x|y) = h(y|x),\ V(yy'|xx') = V(xx'|yy'). \tag{4.4}$$

We have

$$[\psi(x),H] = \int h(x|y)\psi(y) + \int \psi^\dagger(x')V(xx'|yy')\psi(y')\psi(y). \tag{4.5}$$

In analogy with (3.6), we seek a variational principle in the form

$$\delta F/\delta\psi^\dagger(x) = \delta F/\delta\psi(x) = 0, \tag{4.6}$$

where $F$ is of the structure

$$F = \mathrm{Tr}\{H + \int \Lambda(y|x)[\psi(x),\psi^\dagger(y)]\}. \tag{4.7}$$

Equations (4.6) become, for instance,

$$\frac{\delta F}{\delta\psi^\dagger(x)} = [\psi(x),H] + \int [\Lambda(x|y),\psi(y)]\,dy, \tag{4.8}$$

where by the first term we mean the right-hand side of (4.5). The only general solution we have been able to find to this equation is

$$\Lambda(x|y) = \delta(x - y)H, \tag{4.9}$$

analogous to the one-dimensional problem, and this provides a suitable variational principle.

In the formulation just given the variables are the matrix elements of the single-particle creation and annihilation operators. We shall also uncover an alternative formulation in which the variables are the matrix elements of the density operators

$$\rho(x|y) = \psi^\dagger(y)\psi(x). \tag{4.10}$$

We rewrite the Hamiltonian in the form

$$H = \int h'(x|y)\rho(y|x) + \tfrac{1}{2}\int V(xx'|yy')\rho(y'|x')\rho(y|x), \tag{4.11}$$

where

$$h'(x|y) = h(x|y) - \tfrac{1}{2}\int V(xx'|yx')\,dx'. \tag{4.12}$$

The density operators satisfy the commutation relations

$$[\rho(y|x),\rho(y'|x')] = \rho(y'|x)\delta(x'-y) - \rho(y|x')\delta(x-y'). \tag{4.13}$$

We are thereby enabled to derive the equations of motion

$$[\rho(y|x),H] = \int \mathscr{H}(y|z)\rho(z|x) - \rho(y|z)\mathscr{H}(z|x) = [\mathscr{H},\rho](y|x) = 0, \tag{4.14}$$

where

$$\mathscr{H}(y|x) = h'(y|x) + v(y|x), \tag{4.15}$$

$$v(y|x) = \int V(xx'|yy')\rho(y'|x'). \tag{4.16}$$

We now seek a variational principle which yields (4.14). The form of this principle is not as obvious as the forms utilized previously in this paper have become. For one thing, Eq. (4.13) is not the appropriate kinematical restriction since it holds equally for bosons and for fermions and thus fails to distinguish the two. By rearranging the expression

$$\psi^\dagger(y)[\psi(x),\psi(x')]\psi^\dagger(y') = 0 \tag{4.17}$$

with the help of the commutation relations, we find, however, that

$$\rho(x|y)\rho(x'|y') - \rho(x'|y)\rho(x|y') + \rho(x|y)\delta(x'-y') - \rho(x'|y)\delta(x-y') = 0. \tag{4.18}$$

Setting $x = y$ and integrating, we obtain the equation

$$N\rho(y|x) - \int \rho(y|z)\,dz\,\rho(z|x) + N\delta(y-x) - \rho(y|x) = 0, \tag{4.19}$$

where $N$ is the number operator

$$N = \int \psi^\dagger(x)\psi(x)\,dx. \tag{4.20}$$

We write (4.19) in operator form as

$$\rho^2 + \rho - N\rho - N = 0, \tag{4.21}$$

and this will prove to be the constraint sought.

We are now prepared to establish the stationary property of the functional

$$F = \mathrm{Tr}\{H - \Omega[\rho^2 + \rho - N\rho - N] - \mu N\}, \tag{4.22}$$

where we have introduced two Lagrange multiplier operators $\Omega$, which is also a matrix in the space of $x$, and $\mu$, which does not carry any free spatial coordinates. The condition which we wish to verify (with suitable choice of $\Omega$ and $\mu$) is

$$\delta F/\delta\rho = 0. \tag{4.23}$$

With attention to the several kinds of product involved in our definitions, (4.23) yields [with $\mathscr{H}$ defined by (4.15)]

$$\mathscr{H} = \Omega\rho + \rho\Omega + \Omega - \Omega N - \mathrm{tr}(\Omega)1 - \mathrm{tr}\Omega\,1 + \mu 1, \tag{4.24}$$

where tr is a trace only in the space of $x$ and

$$(x|1|y) = \delta(x-y). \tag{4.25}$$

From (4.24) we form the commutator with $\rho$ and find

$$[\rho,\mathscr{H}] = -\Omega(\rho^2 + \rho - N\rho) + (\rho^2 + \rho - N\rho)\Omega + [N\,\mathrm{tr}\Omega,\rho] + [\mathrm{tr}\Omega,\rho] - [\mu,\rho], \tag{4.26}$$

or utilizing (4.21),

$$[\rho,\mathscr{H}] = [N,\Omega] - [\rho,N\,\mathrm{tr}\Omega] - [\rho,\mathrm{tr}\Omega] + [\rho,\mu]. \tag{4.27}$$

This last equation must agree with (4.14). This can be achieved if we require

$$[N,\Omega] = 0, \tag{4.28}$$

$$H = N\,\mathrm{tr}\Omega, \tag{4.29}$$

$$\mu = \mathrm{tr}\Omega. \tag{4.30}$$

Since, in practice, we determine $\rho$ from the solution of (4.14) and (4.21), the Lagrange multiplier $\Omega$ here serves only as an intermediary quantity which, in principle, is determined by (4.24).

The further question of the extent to which the kinematical constraints can be reconstructed from the variational principle will not be pursued, since the considerations are analogous to those given in previous sections.

## V. NONRELATIVISTIC FIELD THEORY. FERMIONS

Now let $\psi^\dagger(x)$, $\psi(x)$ be fermion creation and annihilation operators satisfying the anticommutation relations

$$\{\psi(x),\psi^\dagger(y)\} = \delta(x-y). \tag{5.1}$$

As Hamiltonian we choose the form (4.2) with the properties (4.4), except that (4.3) is replaced by

$$V(xx'|yy') = -V(x'x|yy') = -V(xx'|y'y). \tag{5.2}$$

The equations of motion are of the form (4.5),

$$[\psi(x),H] = \int h(x|y)\psi(y) + \int \psi^\dagger(x')V(xx'|yy')\psi(y')\psi(y). \tag{5.3}$$

At this point the natural impulse is to imitate Eq. (4.7), multiplying the Lagrange multiplier $\Lambda(y|x)$ by the anticommutator. The maneuver fails, however, if pursued heedlessly, to produce a viable variational principle, for a reason which will illuminate the way to a solution of the difficulty. To understand the source of the problem, we must make a distinction between systems composed of an even number of fermions which, as composites, have bosonic properties and systems with an odd number of fermions, which retain fully their fermionic character. We denote states of the former by $|n\rangle$, and states of the latter by $|i\rangle$ or $|j\rangle$, according to context. Thus we write

$$\langle n|\psi^\dagger(y)\psi(x)|n'\rangle = \sum_i \langle n|\psi^\dagger(y)|i\rangle\langle i|\psi(x)|n'\rangle \equiv \sum_i \phi_i^\dagger(yn)\phi_i(xn'), \tag{5.4}$$

and

$$\langle n|\psi(x)\psi^\dagger(y)|n'\rangle = \sum_j \langle n|\psi(x)|j\rangle\langle j|\psi^\dagger(y)|n'\rangle \equiv \sum_j \chi_j(xn)\chi_j^\dagger(yn'). \tag{5.5}$$

In the boson variational principle of the previous section, we identified the amplitudes $\phi_i$ and $\chi_j$ before variation, which should be correct to order $N^{-1}$ for a boson system, but is qualitatively wrong for a system of fermions. One need only think of the Hartree–Fock limit to see this. Here

$|n\rangle = |n'\rangle =$ | Slater determinant of occupied orbitals⟩. Then $\phi_i(x)$ are the orbitals of the occupied single-particle states, $\chi_j(x)$ the orbitals of the unoccupied single-particle states. These must be varied independently subject to the constraints which follow from (5.1), namely,

$$\begin{aligned}\langle n|\{\psi(x),\psi^\dagger(y)\}|n'\rangle &= \delta(x-y)\delta_{nn'} \\ &= \sum_i \phi_i^\dagger(yn)\phi_i(xn') + \sum_j \chi_j(xn)\chi_j^\dagger(yn') \\ &\equiv \rho(xn'|yn) + \tau(xn'|yn). \end{aligned} \tag{5.6}$$

(Notice the peculiar association of indices in the definition of $\tau$).

We now propose the following functional as the basis for a variational principle,

$$F = \mathrm{Tr}\{\tfrac{1}{2}h'(\rho-\tau+1) + \tfrac{1}{8}V(\rho-\tau+1)(\rho-\tau+1) - \tfrac{1}{2}\Lambda(\rho+\tau)\}, \tag{5.7}$$

which is indeed equivalent to what was proposed initially in this section except that we have symmetrized the terms of the Hamiltonian by means of (5.6).

Variation of (5.7) with respect to $\phi_i^\dagger(xn)$ and $\chi_j(yn)$ respectively yields the equations

$$\int \mathscr{H}(xn|yn')\phi_i(yn') = \int \langle n'|\Lambda(x|y)|n\rangle\phi_i(yn'), \tag{5.8}$$

$$-\int \chi_j^\dagger(xn')\mathscr{H}(xn|yn') = \int \chi_j^\dagger(xn')\langle n'|\Lambda(x|y)|n\rangle, \tag{5.9}$$

where

$$\mathscr{H}(xn|yn') = h'(x|y)\delta_{nn'} + \int V(xx'|yy')\rho(y'n|x'n'). \tag{5.10}$$

For (5.8) and (5.9) to be correct, they must agree with the matrix elements of (5.3) and its Hermitian conjugate. These equations are

$$\int \mathscr{H}(xn|yn')\phi_i(yn') = (E_n - E_i)\phi_i(xn), \tag{5.11}$$

$$-\int \chi^\dagger\mathscr{H}(xn|yn') = (E_n - E_j)\chi_j^\dagger(yn), \tag{5.12}$$

where $E_n$ and $E_i$ are the energies of the states $|n\rangle$ and $|i\rangle$, respectively, or comparing with (5.8) and (5.9),

$$\int \langle n'|\Lambda(x|y)|n\rangle\phi_i(yn') = (E_n - E_i)\phi_i(yn), \tag{5.13}$$

$$\int \chi_j^\dagger(xn)\langle n'|\Lambda(x|y)|n\rangle = (E_n - E_j)\chi_j^\dagger(yn). \tag{5.14}$$

These equations provide a determination of $\Lambda$. But more important for our purposes they allow us to simplify the variational principle. By combining (5.13) and (5.14) after the formation of suitable bilinear expressions we obtain for the combination that appears in (5.7),

$$\begin{aligned}\mathrm{Tr}\{\Lambda(\rho+\tau)\} &= \sum_i\sum_n \int [E_n|\phi_i(xn)|^2 - E_i|\phi_i(xn)|^2] \\ &\quad + \sum_j\sum_n \int [E_n|\chi_j(xn)|^2 - E_j|\chi_j(xn)|^2]. \end{aligned} \tag{5.15}$$

The required identity of the two sides of (5.15) permits us to start from a revised and simplified variational principle in which $E_i$, $E_j$, and $E_n$ play the role of Lagrange multipliers.

We now ask: What are the constraints for which these serve as multipliers, constraints which are equivalent to the totality of constraints otherwise and originally expressed by (5.6)? According to (5.15) the energy $E_i$, e.g., is associated with a condition

$$\delta\sum_n \int dx\,|\phi_i(xn)|^2 = 0. \tag{5.16}$$

This is the condition that the "wave function," eigenvalue labeled by $i$ and components labeled by $(xn)$, should have a fixed norm in the variation. Similarly for $E_j$

$$\delta\sum_n \int dx\,|\chi_j(xn)|^2 = 0. \tag{5.17}$$

One should not be beguiled into setting the norms equal to unity, however, since this value is correct only in the extreme Hartree–Fock limit. Otherwise the values must be determined from (5.6) or in the case of (5.16) from the nonlinear equation for $\rho$ given below in Eq. (5.22).

The remaining constraints implicit in (5.15), for which $E_n$ serves as Lagrange multiplier, can be understood in part as conservation of particles,

$$\langle n|N|n\rangle = \sum_i \int dx\,|\phi_i(xn)|^2. \tag{5.18}$$

However, the sign of the other term involving $E_n$ is more difficult (in our opinion) to understand on *a priori* grounds. Thus a symmetrization of the condition (5.18), using (5.16) would have led to the opposite sign.

At this juncture, we are ready to notice that it is possible to construct a variational principle for $\rho(xn|x'n')$ or its ingredients $\phi_i(x'n')$ alone. A form which upon variation yields the correct equations of motion is

$$G = \mathrm{Tr}\{h'\rho + \tfrac{1}{2}V\rho\rho\} - \sum_{i,n}\int (E_n - E_i)|\phi_i(xn)|^2. \tag{5.19}$$

The main point, of course, is that the "Hamiltonian" $\mathscr{H}$ which determines $\rho$ is a functional only of $\rho$ itself and thus, in analogy with Hartree–Fock theory, we might think it possible to determine $\rho$ first and $\tau$ afterwards. There is one remaining objection to the theoretical feasibility of this program, namely, we need a condition for the normalization of $\rho$ which is independent of $\tau$, so that we can dispense with (5.6). Such a condition is provided below in Eq. (5.22), which is the analog for fermions of Eq. (4.21) for bosons.

Our previous remarks should be clearer once we show that a variational principle analogous to (4.22) for bosons can be formulated. Indeed all we require is an analog of (4.21). Starting from the expression

$$\psi^\dagger(y)\{\psi(x),\psi(x')\}\psi^\dagger(y') = 0, \tag{5.20}$$

we derive

$$\rho(x|y)\rho(x'|y') + \rho(x'|x)\rho(x|y')$$

$$= \rho(x'|y)\delta(x-y') + \rho(x|y)\delta(x'-y'). \tag{5.21}$$

Setting $x = y$ and integrating we find

$$\rho^2 - \rho + N\rho - N = 0, \tag{5.22}$$

which is the constraint sought.

We therefore study the functional

$$F = \mathrm{Tr}\{H - \Omega(\rho^2 - \rho + N\rho - N) - \mu N\}. \tag{5.23}$$

The condition $\delta F/\delta\rho = 0$ yields

$$\mathscr{H} = \Omega\rho + \rho\Omega - \Omega + \Omega + (\mathrm{tr}\Omega)N1 + (\mathrm{tr}\Omega)1 + \mu 1, \tag{5.24}$$

or utilizing (5.22) and the assumption that $\Omega$ commutes with $N$, we have

$$[\rho,H] = [\rho,\mathrm{tr}\Omega N] + [\rho,\mathrm{tr}\Omega] + [\rho,\mu]. \tag{5.25}$$

This is the correct equation of motion if

$$H = -N\,\mathrm{tr}\Omega, \tag{5.26}$$

$$\mu = -\mathrm{tr}\Omega. \tag{5.27}$$

Finally we consider the important question of the extent to which (5.22) determines the norm of the amplitudes $\phi_i(xn)$. From the equation of motion (5.11) we can conclude in the standard way [because $\mathscr{H}(xn|yn') - E_n\delta_{nn'}\delta(x-y)$ is an Hermitian operator] that

$$\sum_n \int dx\, \phi_i^\dagger(xn)\phi_{i'}(xn) = \delta_{ii'}\mathscr{N}_i, \tag{5.28}$$

where $\mathscr{N}_i$ is a normalization condition to be determined. On the other hand, Eq. (5.22) when written out is

$$\int dx' \sum_{ii''} \phi_{i''}^\dagger(x'n)\phi_{i''}(xn'')\phi_i^\dagger(yn'')\phi_{i'}(x'n') - \sum_i \phi_i^\dagger(yn)\phi_{i'}(xn') = N\left[\delta(x-y)\delta_{nn'} - \sum_i \phi_i^\dagger(yn)\phi_{i'}(xn')\right]. \tag{5.29}$$

By forming obvious scalar products so as to utilize (5.28), and introducing the definition

$$P_{ii'}(n|n') \equiv \int dx\, \phi_i^\dagger(xn)\phi_{i'}(xn'), \tag{5.30}$$

we find a result that can be written in the form

$$\mathscr{N}_i^3 - \mathscr{N}_i^2 = N(\mathscr{N}_i - \mathscr{N}_i^2) + \mathscr{N}_i^3 - \sum_{n,n',n'',i'i''} P_{ii''}(n'|n'')P_{i''i'}(n|n')P_{i'i}(n''|n). \tag{5.31}$$

In the Hartree–Fock limit when there is no sum on $n$ values, the product of $P$'s becomes $\mathscr{N}_i^3$ and we obtain the usual normalization condition $\mathscr{N}_i = 1$. In many cases of interest it should be possible to solve (5.31) by iteration starting from this extreme value.

This argument also shows that the variational principle based on (5.19) provides a complete scheme of calculation—without our having to consider at the same time the $\chi$ amplitudes. This constitutes a considerable simplification.

## VI. DISCUSSION

The purpose of this exposition has been to state and prove a number of variational principles associated with Heisenberg matrix mechanics. Our discussion has been purposely monolithic, but we have thereby failed to touch several issues of some importance, which we will survey in our final remarks. For these remarks let us initially refer to the specific model of one-dimensional quantum mechanics.

The variational principle given for this simple case involved the trace of the Hamiltonian $H$. The variation then yielded the Heisenberg equations of motion. In practice one solves a set of nonlinear algebraic equations[3] derived as approximations to the equations of motion and the commutation relations. We can prove a theorem, but will not do so here, that in this scheme $H$ is diagonal in the same subspace of Hilbert space used to derive the approximate scheme. Because the equations expressing the commutation relations were part of the computation scheme, we have thereby also found a representation of the algebra—to the same approximation.

It is the theorem quoted above about the diagonalization of the Hamiltonian which selects the trace as the favored average for formation of a variational principle. According to the Rayleigh–Ritz principle, however,

$$\delta\langle n|H|n\rangle = 0, \tag{6.1}$$

any weighted average should be stationary and a suitably constrained weighted average should, upon variation, yield a valid set of equations. If these equations are adjoined to the matrix elements of the commutation relations so that a sufficient number of equations is available to determine the matrix elements of $x$, of $p$, and of the Lagrange multipliers that appear in the scheme, we determine thereby a basis in a subspace, but one in which $H$ is *not*, in general, diagonal. On the other hand the matrix of $H$ can be computed by sum rule methods and the diagonalization carried out as an independent second step. In the past, we have, in fact, stumbled onto such formulations and applied them to the study of various nuclear models.[11,12] One may even think of circumstances which call for the use of a weighted average, e.g., a situation where one favors accurate knowledge about one or more states over others in the subspace considered.

Until now, the variational principles have not been used directly, except for theoretical development. We can envision future applications in which trial values (with parameters) of the matrix elements of the fundamental operators are entered directly into the variational principles.

We mention finally two other aspects of the problem for further theoretical study. The first is the study of the classical limit where the following results emerge rather easily: The variational principles involving canonical variables (commutation relations!) go over into Hamilton's principle. The principles involving the density matrix go over first into a Hartree or Hartree–Fock formulation and eventually into a classical density matrix formulation.

To explain the second aspect, we start by reducing the complexity of the field-theory problem by introducing a "shell model," that is, by expanding in a complete set $f_\alpha(x)$,

$$\psi(x) = \sum a_\alpha f_\alpha(x), \tag{6.2}$$

and retaining only a finite number of terms. At this juncture the $a_\alpha^\dagger a_\alpha$ play the role played previously by $\rho(x|y)$ and there is nothing new to be said. However in a further simplification, we sometimes form closed Lie algebras from combinations

$$\sum_{\alpha,\beta} a_\alpha^\dagger (j_i)_{\alpha\beta} a_\beta \equiv J_i, \quad i = 1,\cdots, \tag{6.3}$$

and approximate $H$ as a polynomial in the $J_i$ (*not* the angular momentum in general). One may at this point ask for a variational principle. In so far as one now drops the definition (6.3) and defines the $J_i$ by their commutation relations

$$[J_i, J_j] = f_{ijk} J_k, \tag{6.4}$$

with structure constants $f_{ijk}$, and the representation by the values of the Casimir invariants, one finds *new* and challenging forms of the variational principle.

We plan to develop these remarks in future work.


## ACKNOWLEDGMENT

This work was supported in part by the DOE through contract No. EY-76-C-02-3071.



[1] A. K. Kerman and A. Klein, Phys. Lett. **1**, 185 (1962).
[2] F. R. Halpern, J. Math. Phys. **14**, 219 (1973).
[3] C. T. Li, A. Klein, and F. Krejs, Phys. Rev. D **12**, 2311 (1975).
[4] C. A. Uzes, Phys. Rev. D **14**, 3362 (1976).
[5] J. Goldstone and R. Jackiw, Phys. Rev. D **11**, 1486 (1975).
[6] A. Klein and A. Weldon, Phys. Rev. D **17**, 1009 (1978).
[7] G. Do Dang, G. J. Dreiss, R. M. Dreizler, A. Klein, and Chi-Shiang Wu, Nucl. Phys. A **114**, 501 (1968).
[8] A. Klein and F. Krejs, Phys. Rev. D **12**, 3112 (1975).
[9] A. Klein and C. T. Li, Physica (The Hague) **96A**, 243 (1979); Section 3 of this paper describing the derivation of commutation relations is seriously in error—a conceptual and an algebraic error cooperate to yield the result sought. A correct exposition is found in Sec. II and in Sec. III of the present account.
[10] The only published paper on this general subject, written by someone outside our group, of which we are aware, is by D. Janssen, Yad. Fiz. **28**, 935 (1978) [Sov. J. Nucl. Phys. **28**, 479 (1978)]. This work applies a method described in Ref. 7.
[11] C. Dasso, F. Krejs, and P. K. Chattopadhyay, Nucl. Phys. A **210**, 429 (1973).
[12] F. Krejs and A. Klein, J. Math. Phys. **14**, 1155 (1973).

# Dynamical symmetry and magnetic charge

Jonathan F. Schonfeld
*Laboratoire de Physique Théorique de l'Ecole Normale Supérieure,* [a] *24 rue Lhomond, 75231 Paris-cedex 05, France*
*and*
*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455*

(Received 20 February 1980; accepted for publication 6 June 1980)

By adding to the force between an electric and a magnetic point charge a central force arising from a specially chosen potential, one can construct a system known to have the same SO (3,1) and/or SO (4) dynamical symmetry algebra as the Kepler system. We derive projective changes of variables under which the classical orbits of any such system are put in one-to-one correspondence with SO (3,1)- and/or SO (4)-invariant sets of curves on similarly invariant surfaces. This extends results hitherto established only for the Kepler system. This is surprising in that there is a sense in which the phase space of such a magnetic system is a truncation of the Kepler phase space and so one might have expected such global properties not to generalize. Our transformations apparently do not permit transcription of the corresponding Schrödinger equation into a manifestly SO (3,1)- and/or SO (4)-symmetric form, in contrast to the pure Kepler case. Such magnetic systems play roles in the theory of quantum fields in Taub–NUT space-times, and in the theory of quantum-mechanical fluctuations about extended magnetic monopoles in supersymmetric gauge theories. In passing, we use the properties of the magnetic systems to formulate a very short and direct proof that the classical orbits of the Kepler system are conic sections.

## 1. INTRODUCTION

According to a general theorem[1] that can easily be verified by direct calculation, if the 3-vector $\mathbf{r}$ satisfies the equation of motion

$$m\ddot{\mathbf{r}} = \frac{q\dot{\mathbf{r}}\times\mathbf{r}}{r^3} - \nabla\left[V(r) + \frac{q^2}{2mr^2}\right], \tag{1.1}$$

then the related vector $\mathbf{R}$, defined by

$$\mathbf{R} \equiv [\mathbf{r} + qr\mathbf{L}/L^2](1 - q^2/L^2)^{-\frac{1}{2}}, \tag{1.2}$$

satisfies the equation

$$m\ddot{\mathbf{R}} = -\nabla V(R), \tag{1.3}$$

and both orbits have the same energy (which we shall call $E$). The vector in (1.2) is simultaneously the conserved angular momentum for both systems (1.1) and (1.3),

$$\mathbf{L}_q \equiv m\mathbf{r}\times\dot{\mathbf{r}} - q\hat{r} = m\mathbf{R}\times\dot{\mathbf{R}} \equiv \mathbf{L}, \tag{1.4}$$

and $L$ is its length, just as $r$ and $R$ are the lengths of $\mathbf{r}$ and $\mathbf{R}$ respectively. The symbol $\hat{r}$ stands for $\mathbf{r}/r$.

From this one might be tempted to conclude that systems (1.1) and (1.3) are really the same system expressed in two different coordinate schemes, but this is naive for two reasons:

First, one notices that, by virtue of (1.4), $L \geqslant |q|$ because

$$\|m\mathbf{r}\times\dot{\mathbf{r}} - q\hat{r}\|^2 = \|m\mathbf{r}\times\dot{\mathbf{r}}\|^2 + q^2 \geqslant q^2. \tag{1.5}$$

Thus there are no orbits of system (1.1) that correspond under transformation (1.2) to those orbits of system (1.3) satisfying $\|m\mathbf{R}\times\dot{\mathbf{R}}\| < |q|$.

Second, one observes that the transformation (1.2) is formally singular when $L = |q|$. This is an indirect reflection of a basis mismatch between the two systems. If $L$ is fixed at $|q|$, then, for fixed $\hat{L}$ and $E$, there is only one solution of (1.1), whereas there are in most cases infinitely many such orbits of (1.3) (any one is obtained from any other by rotation about $\hat{L}$). Thus the singularity in (1.2) indicates that when $L = |q|$ the correspondence between orbits of (1.1) and (1.3) is not one-to-one but one-to-many.

In summary, the transformation (1.2) determines a one-to-one correspondence between the phase space of system (1.1) and a set formed from the phase space of system (1.3) by excision of the set $m\|\mathbf{R}\times\dot{\mathbf{R}}\| < |q|$ and by the performance of certain topological identifications at the boundary $m\|\mathbf{R}\times\dot{\mathbf{R}}\| = |q|$. (Incidently, a similar phenomenon has recently been observed in a set of dynamical systems of fields. See Ref. 2.)

We raise this issue because the quantum mechanical version of a system of type (1.1) with a high degree of dynamical symmetry has recently turned up in two distinct settings in quantum field theory. (We shall describe them in detail at the end of this section.) The potential involved in either case is of the general form

$$V(r) = \frac{1}{2m}\left(\frac{mk}{q} + \frac{q}{r}\right)^2 - \frac{q^2}{2mr^2}, \tag{1.6}$$

where $k$ is a constant. It is clear that with the potential given by (1.6), system (1.3) has the Kepler form, and therefore has a conserved Runge–Lenz vector, as well as a conserved energy and angular momentum. Through transformation (1.2) this property carries over to system (1.1) as well. The expression for the Runge–Lenz vector of system (1.3) in terms of the coordinate $\mathbf{r}$ is

$$\begin{aligned}\mathbf{M} &\equiv m\dot{\mathbf{R}}\times\mathbf{L} + k\hat{R} \\ &= (1 - q^2/L^2)^{-1/2}[\mathbf{M}_q + (kq/L^2)\mathbf{L}_q],\end{aligned} \tag{1.7}$$

[a] Laboratoire propre du CNRS associé à l'Ecole Normale Supérieure et à l'Université de Paris-Sud.

 0022-2488/80/102528-06$1.00  

where the conserved vector $\mathbf{M}_q$ is defined by

$$\mathbf{M}_q \equiv m\dot{\mathbf{r}} \times \mathbf{L}_q + k\hat{r}. \tag{1.8}$$

One can check[3-6] directly that, just as in the pure Kepler case, either classically (using Poisson commutation) or quantum-mechanically (using operator commutation, provided the magnetic monopole coupling $q$ is quantized appropriately), the components of $\mathbf{L}_q$ and $(m/2|\epsilon|)^{1/2}\mathbf{M}_q$ form the Lie algebra of O (3,1) for $\epsilon > 0$ and of O (4) for $\epsilon < 0$, where the conserved quantity $\epsilon$ is defined by

$$\epsilon \equiv E - \frac{mk^2}{2q^2} \equiv \frac{1}{2}m\|\dot{\mathbf{r}}\|^2 + \frac{1}{2m}\left(\frac{q}{r} + \frac{mk}{q}\right)^2 - \frac{1}{2m}\left(\frac{mk}{q}\right)^2. \tag{1.9}$$

(The importance of $\epsilon$ is that, unlike $E$, it has a smooth limit as $q \to 0$.) In the quantum-mechanical case, for any energy eigenvalue of the Hamiltonian the corresponding eigenstates fill out complete representations of this O (3,1) or O (4). Henceforth, as we shall always be referring to $\mathbf{L}_q$ and $\mathbf{M}_q$, we shall omit the subscript "$q$." The momentum canonically conjugate to $\mathbf{r}$ is

$$\mathbf{p} \equiv \dot{\mathbf{r}} + q\mathbf{A}(\mathbf{r}), \tag{1.10}$$

where the necessarily singular vector potential $\mathbf{A}$ is defined up to Abelian gauge transformation by

$$\nabla \times \mathbf{A} = \mathbf{r}/r^3. \tag{1.11}$$

We shall use the name "$q$–Kepler" to indicate systems of the kind shown in (1.1) with potential given by (1.6), and "Kepler" or "pure Kepler" or "$q = 0$" to indicate the corresponding systems of the kind shown in Eq. (1.3). The results in the present paper extend to the $q$–Kepler systems some statements about the geometrical significance of the O (3,1) and O (4) symmetry algebras that had been previously formulated only in the pure Kepler case.[7]

In the classical version of the pure Kepler case, for $\epsilon < 0$ (when the sign of $k$ permits) there are known to be two separate projections from spheres in four-dimensional Euclidean space—one (stereographic) to three-dimensional velocity space and one (vertical) to three-dimensional position space—that establish one-to-one correspondences between physical orbits and great circles on the spheres.[8] Of course any two great circles on the same sphere can be transformed into one another by O (4) rotations. For $\epsilon > 0$ there are analogous correspondences between the set of physical orbits and O (3,1)-invariant sets of paths on mass-hyperboloids in four-dimensional Minkowski space.

Our main results, detailed in the next section, are the generalizations of these projections to the $q$–Kepler systems. The appropriate generalizations of the vertical projections turn out to be straightforward. The generalizations of the stereographic projections are not. In the pure Kepler case the center of the well-known projection is located on the sphere or hyperboloid being projected; when $q$ is nonzero the center of the projection turns out to be located off the sphere or hyperboloid. An important consequence is the following distinction: when $q = 0$ the projections into velocity space (understood to include a point at infinity) are onto and one-to-one, whereas when $q \neq 0$ there are regions of velocity space excluded from the ranges, and the projections are two-to-one in large regions of the domain.

The existence of these maps, an open question until now (and conjectured against in Ref. 4), is somewhat surprising to us. We had expected that the spheres and hyperboloids of the pure Kepler system would suffer some sort of damage on account of the excisions and topological identifications implicit in the passage (1.2) to the $q$–Kepler systems. Evidently this is not so, at least classically.

The quantum-mechanical situation is different. In the pure Kepler case, under the stereographic projections discussed above, the momentum-space Schrödinger eigenvalue equation is known to transform into an equation manifestly covariant under the O (4) or O (3,1) symmetry group of the associated sphere or hyperboloid.[7] For several reasons there appears, in contrast, to be no such quantum-mechanical role for the stereographic projections of the general $q$–Kepler system. First, because when $q$ is nonzero the three components of kinetic momentum, to which the projections refer, do not commute as operators and therefore cannot be used as Fourier transformation variables. Second, because when $q$ is nonzero the projections are, as noted above, neither one-to-one nor onto, so that even if one could sensibly define a wavefunction on physical velocity space one could not thereby transfer it to the appropriate sphere or hyperboloid without conspicuous loss of information. We shall return to this issue in the third section.

As mentioned earlier, the quantum-mechanical version of the $q$–Kepler system,

$$E\psi = \left[\frac{-1}{2m}(\nabla - iq\mathbf{A})\cdot(\nabla - iq\mathbf{A}) + \frac{1}{2m}\left(\frac{mk}{q} + \frac{q}{r}\right)^2\right]\psi \tag{1.12}$$

has appeared in two distinct settings of current interest in quantum field theory.

Most recently Page[9] observed that Eq. (1.12) describes the components of angular frequency $q$ in the Green's function of a scalar quantum field propagating through the self-dual Taub–NUT spacetime, provided one makes the identifications $2m \equiv 1$ and $k \propto q$.

Equation (1.12) also corresponds to an equation that determines (up to boundary conditions[10]) the spectrum of small quantum-mechanical fluctuations about a particular singular solution to the Bogomolny monopole equations in a supersymmetric gauge theory.[11] The appropriate eigenvalue equation determining (up to boundary conditions) the spectrum of fluctuations about the general solution is

$$-D_j D_j P + g^2[S,[S,P]] = \omega^2 P, \tag{1.13}$$

where $P$ is a fluctuating pseudoscalar field, $S$ is the vacuum expectation of a scalar field, $\omega$ is the angular frequency of the fluctuation, and $g$ is a coupling constant. Both $P$ and $S$ take values in the Lie algebra of the gauge group, and the action of the covariant derivative is defined by

$$D_j P \equiv \partial_j P - ig[\mathscr{A}_j, P], \tag{1.14}$$

where $\mathscr{A}_j$, itself an element of the Lie algebra, is the expectation of the gauge field. Bogomolny's equation for $S$ and the $\mathscr{A}_j$ is

$$D_j S = \tfrac{1}{2}\epsilon_{jkl} F_{kl}, \tag{1.15}$$

where $F$ is the non-Abelian field strength determined by $\mathscr{A}$. To obtain (1.12) as a special, albeit singular, case of (1.13)–(1.15), let $T$ be an arbitrary element of the Lie algebra, let $\{t_a\}$ be the set of solutions to the eigenvalue equation

$$[T,t] = \lambda t, \tag{1.16}$$

and define the functions $\psi_a$ by

$$P = \sum_a t_a \psi_a. \tag{1.17}$$

Then, with $\mathscr{A}_j$ defined as in (1.11), the fields

$$A_j \equiv \mathscr{A}_j T, \quad S \equiv -(1/r + \text{const})T \tag{1.18}$$

solve (1.15); and when (1.18) is substituted into (1.13) the functions $\psi_a$ satisfy (1.12), with the identifications $2m \equiv 1$, $q \equiv g\lambda_a$, $k \equiv 2q^2 \times \text{const}$ and $E \equiv \omega^2$. [In this context the familiar quantization condition $q = n/2$ with $n$ integral, for every value of the index $a$, amounts to the condition that $\mathscr{A}_j$ and $S$ as given by (1.18) can be gauge transformed into a nonabelian spherically symmetric configuration having no string singularity.]

A natural question to ask at this point is whether the high dynamical symmetry of Eq. (1.13) is present only for the substitution (1.18) (which has a gauge-invariant singularity at the origin), or whether at least part of the dynamical symmetry might be active under more general conditions—for example assuming only (1.15) and rotational symmetry. In particular, does Eq. (1.13) exhibit any dynamical symmetry when one substitutes for $\mathscr{A}$ and $S$ any of the explicit nonsingular spherically symmetric solutions to (1.15) derived in Refs. 12–14? At present this remains speculative.

It may be useful to note in passing that at large distance the nonsingular Prasad–Sommerfield[12] solution to (1.15) turns out to approach a gauge transform of (1.18) with a negative constant, corresponding to an attractive Kepler term in (1.12). Thus one expects that in such a background the fluctuation equation (1.13) should have infinitely many bound states, with a high degree of near-degeneracy corresponding to the SO (4) multiplets of Eq. (1.12).

We now turn to a detailed description of the projections that make explicit the dynamical symmetry of the classical $q$–Kepler systems. As an amusing tangent, we formulate at the end of Sec. 2 a very brief and direct proof that every classical orbit of every $q$–Kepler system is a conic section. The paper will conclude with the third section, in which we discuss the quantum-mechanical situation in more depth.

## 2. PROJECTIONS OF THE CLASSICAL $q$–KEPLER SYSTEMS THAT MAKE THE DYNAMICAL SYMMETRIES MANIFEST

The algebra that follows generalizes the calculations of Ref. 8. We imagine in what follows that the physical position and velocity vectors move in three linear dimensions of a four-dimensional linear space. The generic 4-vector will be called $u$, represented in coordinates as $(u_0,\mathbf{u})$. We define the inner product of two such vectors $u$ and $u'$ to be

$$u\cdot u' \equiv u_0 u_0' - (\text{sgn}\,\epsilon)\mathbf{u}\cdot\mathbf{u}', \tag{2.1}$$

where $\epsilon$ is defined in Eq. (1.9). Thus we imagine bounded motion ($\epsilon < 0$) to take place in a four-dimensional Euclidean space, and unbounded motion ($\epsilon > 0$) to take place in a four-dimensional Minkowski space. The case $\epsilon = 0$ is exceptional in this framework and must be treated as a common limit of the other two cases. Note that $\epsilon$ can be less than zero only if $k$ is negative.

Before we proceed with the discussion of the projections, let us derive a useful identity relating $M^2$, $L^2$, and $\epsilon$. According to the definition (1.7),

$$\begin{aligned} M^2 &= v^2 L^2 - (\mathbf{v}\cdot\mathbf{L})^2 + \frac{2k}{r}\mathbf{L}\cdot(\mathbf{r}\times\mathbf{v}) + k^2 \\ &= v^2 L^2 - (\mathbf{v}\cdot\mathbf{L})^2 + \frac{2k}{mr}\mathbf{L}\cdot(\mathbf{L} + q\hat{r}) + k^2 \\ &= v^2 L^2 - q^2(\hat{r}\cdot\mathbf{v})^2 + \frac{2k}{mr}L^2 - \frac{2kq^2}{mr} + k^2, \end{aligned} \tag{2.2}$$

where $\mathbf{v}$ denotes velocity ($\dot{\mathbf{r}}$). At the same time, it follows from definition (1.4) that

$$(\hat{r}\cdot\mathbf{v})^2 = v^2 + \frac{q^2}{m^2 r^2} - \frac{L^2}{m^2 r^2}. \tag{2.3}$$

Upon substituting (2.3) into (2.2), one obtains

$$M^2 = (L^2 - q^2)\frac{2\epsilon}{m} + k^2, \tag{2.4}$$

which is the desired identity.

Now let us define

$$u_v \equiv \left(\left(\frac{2|\epsilon|}{m}\right)^{1/2}\left(1 - \frac{2\epsilon r}{k}\right), \frac{2\epsilon r}{k}\mathbf{v}\right) \tag{2.5}$$

$$u_r \equiv \left(-\mathbf{r}\cdot\mathbf{v}\sqrt{2|\epsilon|m}\,, 2\epsilon\mathbf{r} - \mathbf{M}\right). \tag{2.6}$$

These four-vectors satisfy the norm identities

$$u_v^2 \equiv u_v\cdot u_v = 4|\epsilon|Eq^2/m^2k^2, \tag{2.7}$$

$$u_r^2 = (2q^2E/m)(-\text{sgn}\,\epsilon). \tag{2.8}$$

Equation (2.7) is an immediate consequence of the definitions of $\epsilon$ and $E$. To prove (2.8), one first expands the left-hand side:

$$u_r^2 = M^2 - 4\epsilon\mathbf{r}\cdot\mathbf{M} + 4\epsilon^2 r^2 - 2\epsilon m(\mathbf{r}\cdot\mathbf{v})^2, \tag{2.9}$$

and then one substitutes (2.4) into the first term, (2.3) into the last term, and one transforms the second term using the identity

$$\begin{aligned} \mathbf{r}\cdot\mathbf{M} = \mathbf{L}\cdot(\mathbf{r}\times\mathbf{v}) + kr &= \frac{1}{m}\mathbf{L}\cdot(\mathbf{L} + q\hat{r}) + kr \\ &= \frac{1}{m}(L^2 - q^2) + kr. \end{aligned} \tag{2.10}$$

Note that for $\epsilon > 0$, $u_v$ is timelike while $u_v$ is spacelike. For any value of $\epsilon$ it is easy to see that $u_r\cdot u_v = 0$.

$u_r$ and $u_v$ also satisfy

$$0 = u_v\cdot p = u_r\cdot p = u_v\cdot n = u_r\cdot n, \tag{2.11}$$

where

$$p \equiv \left(1, \frac{\mathbf{M}\times\mathbf{L}}{L^2 - q^2}\left(\frac{m}{2|\epsilon|}\right)^{1/2}\right) \tag{2.12}$$

and

$$n \equiv (0, k\,\mathbf{L} + q\mathbf{M}). \tag{2.13}$$

The first equality in (2.11) is proved by expanding

$$u_v \cdot p = \left(\frac{2|\epsilon|}{m}\right)^{1/2}\left[1 - \frac{2\epsilon r}{k} - \frac{mr}{k}\frac{\mathbf{v}\cdot(\mathbf{M}\times\mathbf{L})}{L^2 - q^2}\right], \tag{2.14}$$

and then observing that

$$\begin{aligned}\mathbf{v}\cdot(\mathbf{M}\times\mathbf{L}) &= (\mathbf{v}\cdot\mathbf{L}^2) - v^2L^2 - k\,\mathbf{L}\cdot(\mathbf{r}\times\mathbf{v}) \\ &= q^2(\mathbf{v}\cdot\hat{\mathbf{r}})^2 - v^2L^2 - \frac{k}{mr}(L^2 - q^2) \\ &= \frac{1}{m}\left(2\epsilon - \frac{k}{r}\right)(L^2 - q^2),\end{aligned} \tag{2.15}$$

where the last equality in (2.15) follows from an application of (2.3). The second equality in (2.11) is proved by expanding

$$U_r \cdot p = \sqrt{2m|\epsilon|}\ [-(\mathbf{r}\cdot\mathbf{v}) - \mathbf{r}\cdot(\mathbf{M}\times\mathbf{L})/(L^2 - q^2)], \tag{2.16}$$

and then observing that

$$\begin{aligned}\mathbf{r}\cdot(\mathbf{M}\times\mathbf{L}) = \mathbf{L}\cdot(\mathbf{r}\times\mathbf{M}) &= \mathbf{L}\cdot[\mathbf{r}\times(\mathbf{v}\times\mathbf{L})] \\ &= -\mathbf{L}\cdot(\mathbf{L}(\mathbf{r}\cdot\mathbf{v}) + \mathbf{v}qr) \\ &= -(L^2 - q^2)\mathbf{r}\cdot\mathbf{v}.\end{aligned} \tag{2.17}$$

The third equality in (2.11) follows immediately from the identity

$$k\,\mathbf{L} + q\mathbf{M} = (kmr - q\mathbf{L})\times\mathbf{v}, \tag{2.18}$$

while the last equality in (2.11) reflects the decomposition

$$\begin{aligned}&(k\,\mathbf{L} + q\mathbf{M})\cdot(2\epsilon\mathbf{r} - \mathbf{M}) \\ &\quad = -2\epsilon kqr + 2\epsilon q\mathbf{r}\cdot\mathbf{M} - k\,\mathbf{M}\cdot\mathbf{L} - qM^2,\end{aligned} \tag{2.19}$$

into which one should substitute the right-hand sides of Eq. (2.4) and (2.10), as well as that of the simple relation

$$\mathbf{L}\cdot\mathbf{M} = -kq. \tag{2.20}$$

The 4-vectors $(0,\mathbf{v})$, $u_v$ and $((2|\epsilon|/m)^{1/2},0)$ $(\equiv d)$ are colinear, so that by virtue of (2.7) and (2.11) we can say that, for fixed energy, each orbit in velocity space is a stereographic projection, through a fixed center, of the intersection of a fixed sphere or two-sheeted hyperboloid with some two-dimensional plane through the origin. In addition, by virtue of the structure of $u_r$, for fixed energy the corresponding orbit

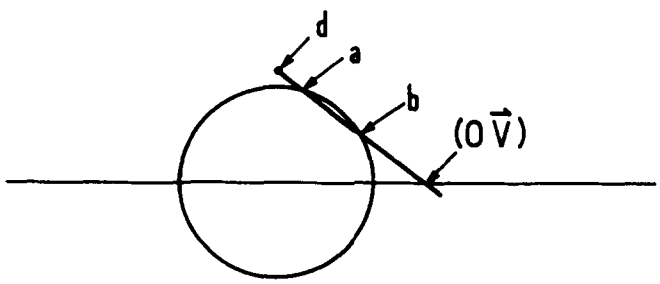



FIG. 1. Stereographic projection (in cross section) of a sphere into a plane through a point ($d$) located above the sphere. Both points $a$ and $b$ project onto the point $(0,\mathbf{v})$.

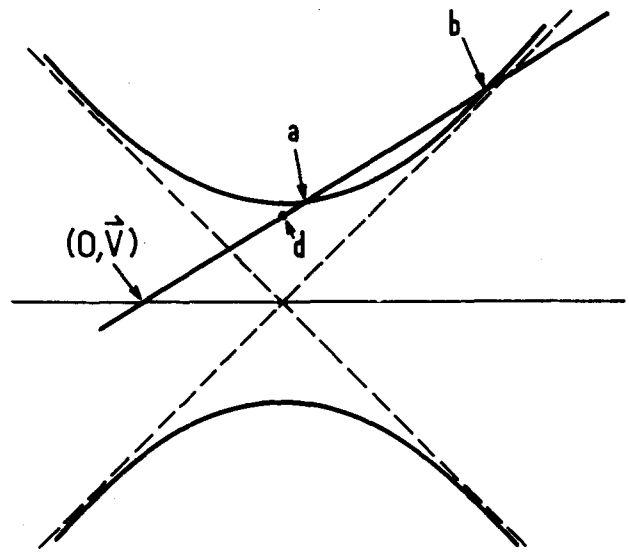



FIG. 2. Stereographic projection of a positive-mass hyperboloid into a plane through a point located below the upper sheet and above the plane. Both points $a$ and $b$ project onto $(0,\mathbf{v})$. The dashed lines indicate asymptotes.

in position space is a vertical projection of the intersection of a fixed sphere or spacelike (one-sheeted) hyperboloid with the same two-dimensional plane through the origin. The sheet of the double-lobed hyperboloid, for $\epsilon > 0$, corresponds to the sign of $u_{v,0}$. This must be a constant of the motion, so that it can be determined from the coordinates of the orbit at large times. Since the motion is unbounded when $\epsilon$ is positive, $u_{v,0}$ at large time is approximately equal to $[-(2|\epsilon|/m)^{1/2}\cdot(2\epsilon/k)r]$, from which one learns that the upper sheet of the hyperboloid is the appropriate one for $k < 0$ (attractive Kepler coupling), while the lower sheet corresponds to $k > 0$ (repulsive coupling).

Note that, for $q \neq 0$, the distance from $d$, the center of projection, to the origin is greater than the radius of the sphere for $\epsilon < 0$ and less than the mass of the hyperboloid for $\epsilon > 0$. This situation is illustrated in Figs. 1 and 2. One sees that there is an outer boundary to the set of velocities that can be reached by such a projection, corresponding to the intersections with the subspace $u_0 = 0$ of lines through $d$ that are tangent to the sphere for $\epsilon < 0$ and tangent to the upper sheet of the hyperboloid for $\epsilon > 0$. This is easy to understand physically; infinitely large velocities are possible only when the potential is unbounded below, and this is not the case here. One also sees from the figures that any velocity inside the boundary corresponds under the projection to two points on the sphere or hyperboloid, although when (for $\epsilon > 0$) $\mathbf{v}$ is within an inner boundary formed by the intersections with $u_0 = 0$ of lines through $d$ asymptotic to the hyperboloid, the two points are on different sheets.

The parameter $\mathbf{p}$ satisfies

$$\begin{aligned}\|\mathbf{p}\|^2 &= \frac{m}{2|\epsilon|}\frac{M^2L^2 - (\mathbf{M}\cdot\mathbf{L})^2}{L^2 - q^2} \\ &= \frac{m}{2|\epsilon|}\left(\frac{2\epsilon L^2/m + k^2}{L^2 - q^2}\right).\end{aligned} \tag{2.21}$$

When $\epsilon < 0$, $L^2$ can be as small as $q^2$ (radial motion) and as

large as $(-mk^2/2\epsilon)$ (circular motion), so that according to (2.21) $0 \leqslant \|\mathbf{p}\| \leqslant \infty$. This means that the intersection of the sphere with each plane through the origin corresponds, under either the stereographic projection (2.15) or the vertical projection (2.6), to a physical orbit and vice-versa. Such curves (the great circles) on the sphere comprise a manifestly SO (4), invariant set. When $\epsilon > 0$, $L^2$ can be as small as $q^2$ (vanishing impact parameter) and as large as infinity (infinite impact parameter), so that in this case $1 < \|\mathbf{p}\| \leqslant \infty$. Despite the gap between $\|\mathbf{p}\| = 0$ and $\|\mathbf{p}\| = 1$, this describes an SO (3,1)-invariant set, all planes that are orthogonal only to spacelike vectors. The intersections of these planes with either a spacelike or a timelike hyperboloid comprise an SO (3,1)-invariant set of curves.

Actually our parametrization of these planes using $p$ and $n$ is singular when $L^2 = q^2$ both because in this limit $\|\mathbf{p}\|$ becomes infinite, as just noted, and because the direction of $\mathbf{n}$ becomes indeterminate. The situation is not helped by replacing $p$ with $p/\|\mathbf{p}\|$ because the direction of $\mathbf{M}\times\mathbf{L}$ is also ill-defined in this limit. The vector with a well-defined direction is $(\mathbf{M}\times\mathbf{L})\times\mathbf{n}$, which satisfies

$$\begin{aligned}(\mathbf{M}\times\mathbf{L})\times\mathbf{n} &= \mathbf{L}(k\,\mathbf{L}\cdot\mathbf{M} + qM^2) - \mathbf{M}(kL^2 + q\mathbf{L}\cdot\mathbf{M}) \\ &= (L^2 - q^2)\left(\frac{2\epsilon q}{m}\mathbf{L} - k\,\mathbf{M}\right) \\ &\equiv (L^2 - q^2)\mathbf{l}. \end{aligned} \tag{2.22}$$

When $L^2 = q^2$, one should replace (2.11) by

$$0 = \mathbf{u}_v \times \mathbf{l} = \mathbf{u}_r \times \mathbf{l}. \tag{2.23}$$

This completes our explanation of the projections. Before turning to the concluding discussion, we want to point out one more amusing consequence of the basic definitions (1.4) and (1.8), of the conserved vectors L and M. As is well know, it follows from (1.4) that any orbit of this system must lie on a fixed cone because

$$\hat{r}\cdot\mathbf{L} = -q. \tag{2.24}$$

On the other hand the orbit must also lie in a plane because according to (2.18) we must have

$$\mathbf{v}\cdot\mathbf{n} = 0. \tag{2.25}$$

Passing to the limit $q = 0$, we have, as far as we know, the fastest and the most direct proof that the orbits of the Kepler system are conic sections.

## 3. CONCLUDING REMARKS—THE MANIFEST SYMMETRY OF THE QUANTIZED SYSTEM

As observed in the Introduction, our stereographic projections of velocity space onto surfaces of manifest SO (4) or SO (3,1) symmetry appear not to have obvious quantum-mechanical counterparts, in part because of limitations built into the definitions of the projections and in part because of the difficulties in adapting velocity space, literally interpreted, to the needs of Fourier superposition in the presence of nonzero magnetic charge. (And this in spite of the fact that degenerate eigenstates of the system are known to fill out complete representations of O (3,1) or O (4), as the case may be.[3-6]) It may thus be productive to contemplate quantum-mechanical alternatives to either "velocity space" or to "functions on a surface of manifest SO (4) or SO (3,1) symmetry." We discuss a number of such possibilities in this section.

1. When $q$ is zero, "Fourier components" means eigenstates of the free wave operator, which is also the Kepler wave operator with vanishing $k$. Thus the natural Fourier basis when $q$ is nonzero might turn out to be the eigenstates of the $q$–Kepler wave operator with vanishing $k$, whose properties are derived in Ref. 3. This is suggested by transformation (1.2) which establishes a classical correspondence between orbits of the free particle and orbits of the $q$–Kepler particle with $k = 0$.

2. According to Hurst,[15] in an appropriate gauge the action of the angular part of the $q$–Kepler wave operator on functions of the two-sphere is equivalent to the action of the group-invariant Laplacian on functions $f_q$ of SO (3) satisfying

$$f_q(ge^{i\xi L_z}) = e^{2i\xi q} f_q(g), \tag{3.1}$$

where $g$ is an arbitrary element of SO (3), $\xi$ is an arbitrary real number and $L_z$ generates rotations about the $z$-axis. Perhaps, by analogy, the correct generalizations of the functions on $S^3$ (or its Minkowski–space analogue) into which the momentum-space Kepler wave functions transform under the stereographic projections of Ref. 7 are functions on SO (4) [or SO (3,1)] satisfying an equivalence relation similar to (3.1).

3. It may turn out that the dynamical symmetry of the quantized $q$–Kepler system is manifest only when the wave equations for several values of $q$ are considered at once. Two facts prepare us for this idea. First, as mentioned in the Introduction, when an array of $q$'s corresponds to the eigenvalues of some generator of a nonabelian Lie group, one can remove the Dirac string singularities from the Schrödinger equations of the array by application of an appropriate local nonabelian gauge transformation. Second, when $q$ is zero and $\epsilon > 0$, one projects the momentum space wavefunction onto a function defined on both sheets of the hyperboloid,[7] even though the classical projection involves only one sheet at a time—upper for $k < 0$, lower for $k > 0$. Thus, in a sense, even for $q = 0$ the quantum-mechanical dynamical symmetry is manifest only when elements of more than one system are considered simultaneously.

4. Finally, there is the possibility that there is some natural way to extend the naive projections to the velocities that are classically excluded and also to resolve the two-to-one ambiguities, but we have no concrete proposals along such lines.


## ACKNOWLEDGMENTS

I am grateful to M. Bander, L. Durand, D. Freedman, C. Goebel, D. Holme and D. Zwanziger for helpful discussions. Parts of this work were carried out at the University of Minnesota under the support of Contract No. E(11-1)-1974 with the U. S. Department of Energy, and at the Aspen Center for Physics.



[1] D. G. Boulware *et al.*, Phys. Rev. D **14**, 2708 (1976).
[2] M. J. Ablowitz, M. Kruskal, and H. Segur, J. Math. Phys. **20**, 999 (1979).

[3]D. Zwanziger, Phys. Rev. **176**, 1480 (1968).
[4]H. V. McIntosh and A. Cisneros, J. Math. Phys. **11**, 896 (1970).
[5]V. A. Andreev, I. A. Malkin, and V. I. Man'ko, "Dynamical symmetries of some models with magnetic charge," P. N. Lebedev Phys. Inst. (Moscow) preprint (1971).
[6]S. Ozaki, "O (4) symmetry of the Hydrogen atom with magnetic monopole and with inverse square distance potential." preprint (1971), as cited in D. M. Stevens, "Magnetic monopoles: An updated bibliography," Virginia Polytechnic Inst. Report VPI-EPP-73-5 (1973).
[7]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[8]H. R. Rogers, J. Math. Phys. **14**, 1125 (1973).
[9]D. Page, Phys. Lett. **85** B, 369 (1979).
[10]J. F. Schonfeld, Nucl. Phys. B **161**, 125 (1979).
[11]A. D'Adda, R. Horsley, and P. Di Vecchia, Phys. Lett. **76** B, 298 (1978).
[12]M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[13]E. E. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976).
[14]F. A. Bais and D. Wilkinson, Phys. Rev. D **19**, 2410 (1979).
[15]C. A. Hurst, Ann. Phys. (New York) **50**, 51 (1968).

# A factorization of $M_4$: Construction of the principal bundle of orthogonal frames over $M_4$ from O(3,3) spinors

Patrick L. Nash
*Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514*

(Received 7 May 1980; accepted for publication 20 June 1980)

The principal bundle of orthogonal frames over $M_4$ is explicitly constructed from certain pairs of O(3,3) spinors that transform as (associated) twistors under the action of the covering group of the Poincaré group. In particular, flat space–time is constructed from these associated twistors, and is thus shown to be an object derivable from geometric structures more basic than the vectors of $M_4$. Associated twistors describe massive elementary particles. The position in $M_4$ of such a particle can be explicitly defined in terms of the components of these twistors. The usual momentum and angular momentum variables which coordinatize the classical phase space of this elementary relativistic system of nonzero mass and arbitrary spin may also be realized in terms of this pair of associated twistors. This realization is not equivalent to descriptions of massive particles using twistors which have previously appeared in the literature.

## 1. INTRODUCTION

It can be argued that the continuum concept of space and time initially arose not so much out of physical considerations as from mathematical convenience. In fact, R. Penrose and M. A. H. MacCallum[1] have suggested that the continuum model of space–time arises solely from its mathematical utility. They go on to say, "... We take the view that to encompass quantum theory and general relativity satisfactorily one needs to do more than simply apply some suitable quantization technique to solutions of Einstein's equations. One should rather be thinking of quantizing spacetime itself. This should not be conceived as simply replacing the continuum by a discrete set of points (though this has been attempted) but rather as seeking a way of treating points as "smeared out" just as quantum theory smears out particles." To this end Penrose introduced twistors into quantum theory, and he and his co-workers developed the twistor formalism into an intriguing, albeit as yet incomplete, physical theory.

An enunciated goal of twistor physics is to derive space–time from combinatorial principles applied to the group SU(2,2). It is widely believed that a necessary step towards the goals of eliminating the continuum from physical theories and the quantization of space–time is some such construction of $M_4$, utilizing concepts more basic than space and time. In this paper we show that, not only $M_4$, but the principal fiber bundle of orthogonal frames over $M_4$ may be constructed from certain geometrical objects that transform as twistors under the action of the universal covering group of the *Poincaré* group. These twistors are intimately related to the real-valued spinors $\psi$ that transform under a real eight-dimensional irreducible representation of O(3,3). [For brevity, we shall simply write "twistor" for the more precise "twistor with respect to the action of the universal covering group of the Poincaré group." Also, throughout this paper "spinor" means a real eight-component O,(3,3) spinor.] This construction is realized as follows. The set of all O,(3,3) spinors can be made into a real eight-dimensional vector space, which we denote as $S_8$. Nondegenerate spinors in $S_8$ (see Ref. 3) characterize some of the properties of massive elementary particles. Every such spinor may be classified according to the type of intrinsic electromagnetic dipole moment possessed (in a rest frame) by the elementary particle that it describes. We shall formulate the construction of the principal orthogonal frame bundle over $M_4$ in terms of the spinors that describe pure magnetic dipole particles (such as electrons), although many other choices are possible. Let $W$ be the seven-dimensional subset of $S_8$ in which these spinors lie (for a characterization of this set see Ref. 3). Consider $S_8 \times S_8$. There exists a submanifold $V$ of $S_8 \times S_8$ which is diffeomorphic to $W \times M_4$. A point in $V$ can be labeled by an ordered pair of spinors $(\psi_1, \psi_2)$ which are subject to four locally independent linear constraints and one other condition. This point determines an element of $W \times M_4$, which in turn defines, by the canonical projection $W \times M_4 \to M_4$: $(\psi, x) \mapsto x$, the coordinates of a point $p \in M_4$ (with respect to some fixed but unspecified coordinate system), and by the canonical projection $W \times M_4 \to W$ a nondegenerate spinor $\psi \in W \subset S_8$. In Ref. 3 it is shown how one constructs an orthogonal tetrad $e^{\alpha}{}_{(\mu)}$ form a nondegenerate spinor $\psi \in S_8$. Hence the mapping $W \times M_4 \to W$: $(\psi, x) \mapsto \psi$ determines a tetrad comprising a basis of the tangent space at $p$, $T_p(M_4)$. The set of points in $V$ which give rise to $x(p)$ generates the set of orthogonal frames over $p \in M_4$. Moreover, the composite map $V \to M_4$ is onto, so that this construction can be carried out for all $p \in M_4$, thereby realizing the principal bundle of orthogonal frames over $M_4$.

The organization of this paper is as follows. In Sec. 2 the submanifold $V \subset S_8 \times S_8$ is implicitly defined, and the diffeomorphism between $V$ and $W \times M_4$ is constructed; explicit formulas for the space–time coordinates $x^{\alpha}$ and the orthogonal tetrad $e^{\alpha}{}_{(\mu)}$ in terms of the pair $(\psi_1, \psi_2) \in V$ are given. In Sec. 3 the interpretation of $(\psi_1, \psi_2) \in V$ as a pair of twistors

 0022-2488/80/102534-04$1.00  

associated to a point in $W \times M_4$ is discussed. It is shown that this pair of twistors contains all of the information needed to specify the position, momentum, and angular momentum (spin + orbital) of a free massive particle moving through $M_4$.

We shall follow the notation of Ref. 3; a short summary of the notation and conventions of Ref. 3 is given in the Appendix.

## 2. $S_8 \times S_8 \supset V \cong W \times M_4$

Let $\psi_1, \psi_2 \in S_8$ and put $\psi_i = \binom{\lambda_i}{\xi_i}$, $i = 1,2$; $\lambda_i$ and $\xi_i$ are real four-component spinors, and $\tilde{\xi}_i$ ($\sim$ denotes transpose) transforms inversely to $\lambda_i$ under SO(3,3) (see Ref. 3). Define $V = \{(\psi_1, \psi_2) \in S_8 \times S_8$: Eqs. (1)–(3) hold$\}$, where Eqs. (1), (2), and (3) are given by

$$A(\lambda_1, \xi_1; \lambda_2, \xi_2)\xi_2 = 0, \tag{1}$$

$$A(\xi_2, \lambda_2; \xi_1, \lambda_1)\lambda_1 = 0, \tag{2}$$

and

$$\tilde{\xi}_2 \lambda_1 = 0; \quad \tilde{\xi}_2 \gamma^5 \lambda_1 \neq 0. \tag{3}$$

Here $A$ is the real $4 \times 4$ matrix defined by

$$A(\lambda_1, \xi_1; \lambda_2, \xi_2) = [\gamma^A \xi_2 \tilde{\xi}_1 \gamma^A, \gamma^5]_+ + [\lambda_2 \tilde{\lambda}_1, \gamma^5]_- + \gamma^5 \lambda_1 \tilde{\lambda}_2 - \lambda_1 \tilde{\lambda}_1 \gamma^5, \tag{4}$$

with $[a,b]_\pm = ab \pm ba$. $\gamma^A$ and $\gamma^5$ are real $4 \times 4$ skew-symmetric matrices (see Appendix) which verify $[\frac{1}{2}\gamma^A, \frac{1}{2}\gamma^5]_- = \frac{1}{2}\epsilon$ and cyclic permutations thereof, where $\epsilon = \gamma^A \gamma^5$. Both Eq. (1) and Eq. (2) are SO(3,1) invariant, and the set of Eqs. (1) and (2) are O(3,1) invariant. Equations (1) and (2) are to be interpreted as follows: One assigns nonzero values to $\lambda_1$ and $\xi_2$ consistent with Eq. (3), and then solves the resulting set of linear equations for $\lambda_2$ and $\xi_1$. Of these eight equations, only four are independent for a particular choice of $\lambda_1$ and $\xi_2$; this is shown below. Hence $V$ is an $8 + 8 - 4 - 1 = 11$ dimensional manifold, which is also the dimension of the principal bundle of orthogonal frames over $M_4$.

The motivation behind Eqs. (1)–(3) is as follows. Define

$$\psi = \begin{pmatrix} \lambda \\ \xi \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \xi_2 \end{pmatrix}. \tag{5}$$

Equation (3) ensures that $\psi \in W$ (see Ref. 3). Equations (1) and (2) are equivalent to

$$\psi_1 = \begin{pmatrix} \lambda \\ \xi - \gamma^A \gamma_\alpha x^\alpha \lambda \end{pmatrix}, \tag{6}$$

and

$$\psi_2 = \begin{pmatrix} \lambda + \gamma_\alpha x^\alpha \gamma^A \xi \\ \xi \end{pmatrix}, \tag{7}$$

where the $x^\alpha$ are four real functions of $(\psi_1, \psi_2) \in V$ defined by

$$x^\alpha = \frac{1}{2\tilde{\xi}_2 \gamma^5 \lambda_1} (\tilde{\xi}_2 \gamma^\alpha \epsilon \xi_1 + \tilde{\lambda}_2 \epsilon \gamma^\alpha \lambda_1). \tag{8}$$

We use the suggestive notation $x^\alpha$ for these functions because there exists a homogeneous transformation of $\psi_1$ and $\psi_2$ which "translates" $x^\alpha$ (see Sec. 3), and hence the $x^\alpha$ may be defined to be the position in $M_4$ of the particle described by $(\psi_1, \psi_2) \in V$. Equations (6) and (7) are straightforward consequences of substituting Eqs. (65)–(68) of the Appendix into Eqs. (1) and (2), and using Eqs. (5) and (8), along with the fact that $\tilde{\xi}_2 \gamma^5 \lambda_2 = \tilde{\xi}_2 \gamma^5 \lambda_1 = \tilde{\xi}_1 \gamma^5 \lambda_1$, a result which follows upon multiplying Eq. (1) with $\tilde{\xi}_2 \gamma^5$ and Eq. (2) by $\tilde{\lambda}_1 \gamma^5$. Equations (5) and (8) define the diffeomorphism $V \to W \times M_4$, while Eqs. (6) and (7) define the inverse mapping $W \times M_4 \to V$. The diffeomorphism defined by Eqs. (6) and (7) is implicit in Eqs. (1) and (2) and the definitions of Eqs. (5) and (8); this is the reason for defining $V$ via Eqs. (1)–(3).

It is obvious that the mapping defined by Eqs. (6) and (7) is a bijection $W \times M_4 \to V$, so that $V$ must have the same dimension as $W \times M_4$, i.e., $V$ is eleven-dimensional. Therefore, as mentioned previously, of the eight constraints imposed by Eqs. (1) and (2), only four are independent for any given values of $\lambda_1$ and $\xi_2$; otherwise $V$ cannot be an eleven-dimensional manifold, which manifestly it is.

The nondegenerate spinor $\psi$ of Eq. (5) determines an orthogonal tetrad $e^\alpha{}_{(\mu)}$, two linearly independent null vectors $n^\alpha$ and $l^\alpha$, a spin tensor $\hbar \Sigma^{\alpha\beta}$, and a scalar $N$. These are defined according to

$$e^\alpha{}_{(1)} = \tfrac{1}{2}\tilde{\xi}\gamma^\alpha \gamma^5 \lambda = -\tfrac{1}{2}\tilde{\psi}\Omega M^{\alpha 5}\psi, \tag{9}$$

$$e^\alpha{}_{(2)} = \tfrac{1}{2}\tilde{\xi}\gamma^\alpha \lambda = -\tfrac{1}{2}\tilde{\psi}\Omega M^{\alpha 6}\psi, \tag{10}$$

$$\Sigma^{\alpha\beta} = -\tilde{\xi} S^{\alpha\beta} \lambda = -\tfrac{1}{2}\tilde{\psi}\Omega M^{\alpha\beta}\psi, \tag{11}$$

$$N = \tfrac{1}{2}\tilde{\xi}\gamma^5 \lambda = -\tfrac{1}{2}\tilde{\psi}\Omega M^{56}\psi, \tag{12}$$

$$n^\alpha = -\tilde{\lambda}\gamma^A \gamma^\alpha \lambda, \tag{13}$$

$$l^\alpha = -\tilde{\xi}\gamma^\alpha \gamma^A \xi, \tag{14}$$

$$e^\alpha{}_{(3)} = \tfrac{1}{4}(n^\alpha - l^\alpha) = -\tfrac{1}{4}\tilde{\psi}\Gamma^4 \Gamma^\alpha \Gamma^7 \psi, \tag{15}$$

and

$$e^\alpha{}_{(4)} = \tfrac{1}{4}(n^\alpha + l^\alpha) = -\tfrac{1}{4}\tilde{\psi}\Gamma^4 \Gamma^\alpha \psi. \tag{16}$$

The $e^\alpha{}_{(\mu)}$ satisfy

$$e^\alpha{}_{(\mu)} e_{(\nu)\alpha} = N^2 \eta_{(\mu)(\nu)}, \tag{17}$$

where

$$\eta_{(\mu)(\nu)} = \mathrm{diag}(1,1,1,-1). \tag{18}$$

These relationships are derived in Ref. 3; the matrices $\Gamma^A$, $M^{AB}$, and $\Omega$ are defined in Ref. 3, and also briefly discussed in the Appendix of this paper.

To summarize: From the pair of spinors $(\psi_1, \psi_2) \in V$ we have extracted the coordinates $x^\alpha(p)$ of a point $p \in M_4$ [defined by Eq. (8)], and erected an orthogonal tetrad $e^\alpha{}_{(\mu)}$ at this point. By varying $\psi$ while keeping $x$ fixed, we generate the set of orthogonal frames over $p \in M_4$. Lastly, $x^\alpha$ may be freely varied to give the orthogonal frame bundle over $M_4$.

## 3. TWISTORS

The transformation properties of $\psi$ under O(3,1) are discussed in Ref. 3. As an example, for a SO(3,1) transformation, $\psi \mapsto \psi' = \exp\{\frac{1}{2}\omega_{\alpha\beta} M^{\alpha\beta}\}\psi$, where $\omega_{\alpha\beta} = -\omega_{\beta\alpha}$ are six real parameters. We now endow $\psi_1$ and $\psi_2$ with another transformation property which enables us to define a left action on $V$ by the universal covering group of the Poin-

caré group. This additional transformation operation is translation. Define a real $4\times4$ matrix $b$ by

$$\tilde{b}\epsilon = \epsilon b, \tag{19}$$

and

$$\tilde{b}\gamma^A = -\gamma^A b. \tag{20}$$

Then $b$ may be uniquely written as

$$b = b_\alpha \gamma^\alpha, \tag{21}$$

where the $b_\alpha$ are four real parameters given by

$$b^\alpha = \tfrac{1}{4}\mathrm{tr}\gamma^\alpha b. \tag{22}$$

Let

$$T(b) = \begin{pmatrix} 1 & 0 \\ \gamma^A b & 1 \end{pmatrix}, \tag{23}$$

and

$$T^*(b) = \begin{pmatrix} \gamma^A & 0 \\ 0 & \gamma^A \end{pmatrix} \tilde{T}^{-1}(b) \begin{pmatrix} -\gamma^A & 0 \\ 0 & -\gamma^A \end{pmatrix} = \begin{pmatrix} 1 & -b\gamma^A \\ 0 & 1 \end{pmatrix}. \tag{24}$$

Under the translation operation on $V$, we define $\psi_1$ and $\psi_2$ to transform according as

$$\psi_1 \mapsto \psi_1' = T(b)\psi_1 = \begin{pmatrix} \lambda_1 \\ \xi_1 + \gamma^A b \lambda_1 \end{pmatrix}, \tag{25}$$

and

$$\psi_2 \mapsto \psi_2' = T^*(b)\psi_2 = \begin{pmatrix} \lambda_2 - b\gamma^A \xi_2 \\ \xi_2 \end{pmatrix}. \tag{26}$$

One easily verifies that translations commute with one another, and that they have the correct commutation relations with the generators $M^{\alpha\beta}$ of $\overline{\mathrm{SO}(3,1)}$. Under this translation operation the $x^\alpha$ transform as

$$\begin{aligned} x^\alpha \mapsto x'^\alpha &= \frac{1}{2\tilde{\xi}_2\gamma^5\lambda_1}\left[\tilde{\xi}_2\gamma^\alpha\epsilon(\xi_1 + \gamma^A b\lambda_1) + (\tilde{\lambda}_2 - \tilde{\xi}_2 b\gamma^A)\epsilon\gamma^\alpha\lambda_1\right] \\ &= x^\alpha - \frac{1}{2\tilde{\xi}_2\gamma^5\lambda_1} b_\beta \tilde{\xi}_2\gamma^5[\gamma^\alpha,\gamma^\beta]_+\lambda_1 \\ &= x^\alpha - b^\alpha. \end{aligned} \tag{27}$$

This corresponds to one's intuitive notion of a translation in space–time, and reinforces the interpretation of the functions defined in Eq. (8) as space–time coordinates.

In order to distinguish the two types of translation transformation laws given by Eqs. (25) and (26), we call $T(b)$ the representation of the translation generated by $b$, $T^*(b)$ the conjugate representation, and say that $\psi_1$ transforms as a twistor under translation, whereas $\psi_2$ transforms as a conjugate twistor under translation; of course both $\psi_1$ and $\psi_2$ transform in the same way under $\overline{\mathrm{O}(3,1)}$. One additional bit of terminology: We say that $(\psi_1,\psi_2)\in V$ are the associated twistors of the element $(\psi,x)\in W\times M_4$ defined by Eqs. (5) and (8).

The associated twistors $(\psi_1,\psi_2)\in V$ provide a classical description of the kinematics of a free massive magnetic dipole particle, such as an electron. The particle's position $r^\alpha$ in space–time, four velocity $v^\alpha$, momentum $p^\alpha$, spin, and angular momentum $J^{\alpha\beta}$ are completely determined by $(\psi_1,\psi_2)\in V$ and a parameter $m$ with dimensions of mass. They are defined by

$$r^\alpha = \hbar m^{-1} x^\alpha; \tag{28}$$

$$v^\alpha = |N|^{-1} e^\alpha{}_{(4)} \quad \text{(see Ref. 3)}, \tag{29}$$

where

$$v_\alpha v^\alpha = -1; \tag{30}$$

$$p^\alpha = m e^\alpha{}_{(4)}, \tag{31}$$

where

$$p_\alpha p^\alpha = -m^2 N^2; \tag{32}$$

$$\text{spin tensor} = \hbar\Sigma^{\alpha\beta} \text{ (see Ref. 3)}, \tag{33}$$

where

$$\Sigma^\alpha{}_\beta v^\beta = 0, \tag{34}$$

and

$$\tfrac{1}{2}\Sigma_{\alpha\beta}\Sigma^{\alpha\beta} = N^2; \tag{35}$$

and

$$\begin{aligned} J^{\alpha\beta} &= \hbar\Sigma^{\alpha\beta} + r^\alpha p^\beta - r^\beta p^\alpha \\ &= (-\hbar/4)(\tilde{\psi}_1 \Omega M^{\alpha\beta}\psi_1 + \tilde{\psi}_2 \Omega M^{\alpha\beta}\psi_2). \end{aligned} \tag{36}$$

$J^{\alpha\beta}$ is invariant under the gauge transformation $x^\alpha \mapsto x'^\alpha = x^\alpha + sv^\alpha$, $s\in\mathbb{R}$. This may also be written as $x'^\alpha = x^\alpha(s) = x^\alpha(0) + sv^\alpha$, which is usually regarded as the equation of the trajectory of the particle moving in $M_4$. With this interpretation of the allowed gauge transformation, we see that $(\psi_1,\psi_2)\in V$ provides a complete classical description of the interaction-free "dynamics" of a massive magnetic dipole particle moving in $M_4$.

An unsolved problem is to define equations of motion for $\psi_1$ and $\psi_2$ which (i) preserve Eqs. (1)–(3), and give $\tilde{\xi}_2\gamma^5\lambda_1 = \text{constant}$ along a trajectory; (ii) yield $\dot{x}^\alpha \propto e^\alpha{}_{(4)}$ as an identity; (iii) admit nontrivial interactions; and (iv) quantize this dynamical scheme. By quantization we mean, loosely speaking, the construction of a representation of the generators of the dynamical symmetry group of the equations of motion in terms of Hermitian operators on an appropriate Hilbert space, assuming that such a symmetry group exists. (For a closed and isolated system, the equations of motion must admit the covering group of the Poincaré group as a dynamical symmetry group.[4]) Once a quantum theory has been formulated, Eq. (8) takes on new meaning, namely, that the *observable* "space–time" is the expectation value of a quantum mechanical operator, which is in accordance with one of the axioms of quantum mechanics.

## APPENDIX

Let $\psi$ denote a real column matrix with eight rows (row indices are suppressed) which coordinatizes the real eight-dimensional vector space $S_8$. $\overline{\mathrm{O}(3,3)}$ acts on $S_8$ on the left as a group of automorphisms that preserves a certain bilinear form; $\psi$ transforms as a spinor under a real $8\times8$ irreducible representation of $\overline{\mathrm{O}(3,3)}$. Write

$$\psi = \begin{pmatrix} \lambda \\ \xi \end{pmatrix}, \tag{37}$$

where $\lambda$ and $\xi$ are real four-component spinors. $\tilde{\xi}$ ($\sim$ denotes transpose) transforms inversely to $\lambda$ under $\overline{\mathrm{SO}(3,3)}$.

Let $\gamma^\alpha$ be a real $4\times4$ irreducible (Majorana) representation of the Dirac matrices, where

$$\gamma^\alpha\gamma^\beta+\gamma^\beta\gamma^\alpha=2g^{\alpha\beta},\quad \alpha,\beta,\cdots=1,2,3,4, \tag{38}$$

and

$$g_{\alpha\beta}=\mathrm{diag}(1,1,1,-1). \tag{39}$$

Define

$$\gamma^5=-\frac{1}{4!}\epsilon_{\alpha\beta\mu\nu}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu =-\gamma^1\gamma^2\gamma^3\gamma^4, \tag{40}$$

$$\epsilon=\gamma^4\gamma^5, \tag{41}$$

and

$$S^{\alpha\beta}=-\tfrac{1}{4}[\gamma^\alpha,\gamma^\beta]_-. \tag{42}$$

Then

$$\tilde{\gamma}^\alpha\gamma^4=-\gamma^4\gamma^\alpha, \tag{43}$$

$$\tilde{\gamma}^\alpha\epsilon=\epsilon\gamma^\alpha, \tag{44}$$

$$\tilde{S}^{\alpha\beta}\gamma^4=-\gamma^4S^{\alpha\beta}, \tag{45}$$

$$\tilde{S}^{\alpha\beta}\epsilon=-\epsilon S^{\alpha\beta}, \tag{46}$$

and

$$[S^{\alpha\beta},\gamma_\mu]_-=\delta^\alpha_\mu\gamma^\beta-\delta^\beta_\mu\gamma^\alpha. \tag{47}$$

Let $\Gamma^A$,$A,B,\cdots=1,\ldots,6$, be six real matrices which generate an irreducible representation of the Clifford algebra $C_6$:

$$\Gamma^A\Gamma^B+\Gamma^B\Gamma^A=2g^{AB}, \tag{48}$$

where

$$g_{AB}=\mathrm{diag}(1,1,1,-1,-1,-1). \tag{49}$$

Define

$$\Gamma^7=\frac{1}{6!}\epsilon_{ABCDEF}\Gamma^A\Gamma^B\Gamma^C\Gamma^D\Gamma^E\Gamma^F, \tag{50}$$

then

$$\Gamma^A\Gamma^7+\Gamma^7\Gamma^A=0, \tag{51}$$

and

$$(\Gamma^7)^2=1. \tag{52}$$

A particular representation of the $\Gamma$ matrices is

$$\Gamma^\alpha=\begin{pmatrix}0 & \gamma^\alpha\epsilon\\ -\epsilon\gamma^\alpha & 0\end{pmatrix}, \tag{53}$$

$$\Gamma^5=\begin{pmatrix}0 & \gamma^\alpha\epsilon\\ -\epsilon\gamma^5 & 0\end{pmatrix}, \tag{54}$$

$$\Gamma^6=\begin{pmatrix}0 & -\epsilon\\ -\epsilon & 0\end{pmatrix}, \tag{55}$$

and

$$\Gamma^7=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{56}$$

The generators of $\overline{\mathrm{SO}(3,3)}$ are

$$M^{AB}=-\tfrac{1}{4}[\Gamma^A,\Gamma^B]_-; \tag{57}$$

in this representation the $M^{\alpha\beta}$ are given by

$$M^{\alpha\beta}=\begin{pmatrix}S^{\alpha\beta} & 0\\ 0 & -\tilde{S}^{\alpha\beta}\end{pmatrix}. \tag{58}$$

The $M^{AB}$ verify

$$[M^{AB},\Gamma_R]_-=\delta^A_R\Gamma^B-\delta^B_R\Gamma^A, \tag{59}$$

$$[M^{AB},M^{RS}]_-=g^{AR}M^{BS}-g^{AS}M^{BR}-g^{BR}M^{AS}+g^{BS}M^{AR}. \tag{60}$$

The skew-symmetric metric spinor of rank two, $\Omega$, may be defined by

$$\tilde{\Gamma}^A\Omega=\Omega\Gamma^A; \tag{61}$$

then

$$\tilde{\Gamma}^7\Omega=-\Omega\Gamma^7, \tag{62}$$

and

$$\tilde{M}^{AB}\Omega=-\Omega M^{AB}. \tag{63}$$

In the above representation $\Omega$ may be chosen to be

$$\Omega=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}. \tag{64}$$

In Ref. 3 the equivalent of

$$-\gamma^A\gamma^\alpha\lambda\tilde{\xi}\gamma_\alpha\gamma^A=\tilde{\xi}\lambda+\xi\tilde{\lambda}-\gamma^5\tilde{\xi}\gamma^5\lambda+\gamma^5\xi\tilde{\lambda}\gamma^5 \tag{65}$$

is proven; see Eq. (42) of Ref. 3. [To derive this identity, contract Eq. (42) with $\lambda_q\xi_r$ and add to the resulting expression $-\gamma^A\gamma^A\lambda\tilde{\xi}\gamma_4\gamma^A=\lambda\tilde{\xi}$.] The transpose of Eq. (65) is

$$-\gamma^\alpha\gamma^A\xi\tilde{\lambda}\gamma^A\gamma_\alpha=\tilde{\xi}\lambda+\lambda\tilde{\xi}+\gamma^5\tilde{\xi}\gamma^5\lambda+\gamma^5\lambda\tilde{\xi}\gamma^5. \tag{66}$$

From Eqs. (65) and (66) trivially follow the identities

$$-\gamma_\alpha\gamma^A\xi\tilde{\xi}\gamma^\alpha=\gamma^A\xi\tilde{\xi}-\epsilon\xi\tilde{\xi}\gamma^5, \tag{67}$$

and

$$-\gamma^\alpha\lambda\tilde{\lambda}\gamma^A\gamma_\alpha=\lambda\tilde{\lambda}\gamma^A+\gamma^5\lambda\tilde{\lambda}\epsilon. \tag{68}$$


[1] R. Penrose and M. A. H. MacCallum, Phys. Rep. **6**, 241 (1973).

[2] $\psi$ denotes a real column matrix with eight rows; row indices are suppressed. Greek indices run from 1 to 4; upper case Latin indices run from 1 to 6. The metric tensor on $M_4$ has components $g_{\alpha\beta}=\mathrm{diag}(1,1,1,-1)$ in a Cartesian coordinate system. $\epsilon_{ABCDEF}$ is the completely antisymmetric Levi-Civita tensor density of weight $-1$ in six dimensions; $\epsilon_{123456}=+1$. $\epsilon_{\alpha\beta\mu\nu}=\epsilon_{\alpha\beta\mu\nu56}$. An index in parenthesis is a vector label. We put $c=1$.

[3] P. L. Nash, J. Math. Phys. **21**, 1024 (1980).

[4] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

# Asymptotic behavior of gravitational fields in a type II coordinate system

Garry Ludwig
*Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

(Received 2 April 1980; accepted for publication 13 June 1980)

With the aid of Penrose's conformal technique the asymptotic behavior of the components of the metric tensor, the Weyl tensor, the Ricci tensor and the spin-coefficients is calculated for a large class of space-times that includes the NUT (Newman–Unti–Tamburino) solution as well as all asymptotically flat space-times. The calculations are done in a coordinate system associated not with null hypersurfaces but with an asymptotically shearfree twisting null congruence. For vacuum the results presented here reduce to those of Aronson and Newman to the order given in their paper.

## 1. INTRODUCTION

Almost a decade ago Aronson and Newman[1] investigated the asymptotic behavior of the Newman–Penrose (NP) vacuum quantities,[2] i.e. of the metric coefficients, the spin-coefficients and the Weyl tensor, for empty "asymptotically flat" space-times in what they called a type II coordinate system. This coordinate system is based on a twisting asymptotically shearfree null geodesic congruence in contrast to the more common type I coordinate system which is based on null hypersurfaces. The asymptotic behavior in type II coordinates was derived, for the vacuum case, from the known behavior[3] in type I coordinates by an asymptotic coordinate transformation.

In this paper we shall generalize their results by removing the restriction that the space-time be vacuum near infinity. Using Penrose's conformal approach[4] we shall derive the asymptotic behavior of the NP quantities for this general case to higher order than obtained in Ref. 1, exhibiting explicitly the behavior of the metric. Penrose's conformal approach has been employed previously to obtain the asymptotic behavior in a type I coordinate system,[5] thus generalizing results of Newman and Unti,[3] and also in obtaining "reduced equations" for exact solutions subject to special assumptions.[6,7]

There are several advantages in using this conformal approach where an unphysical space-time $\hat{M}$ (with boundary $\mathscr{I}$) is introduced whose interior is conformally related to the physical space-time $M$. The notion of the behavior of NP quantities at infinity becomes more meaningful. The coordinate system and the tetrad are chosen at $\mathscr{I}$ right from the outset; they do not have to be changed during the solving process as in the regular Newman–Penrose procedure.[3] All "integration constants" are given directly in terms of quantities defined on $\mathscr{I}$. No integrations are involved, only differentiations, and even these can be avoided to a large extent by making power series expansions and comparing coefficients. The conformal factor gives an extra degree of freedom which can be used to simplify the equations to be solved.

The asymptotic behavior of the Ricci tensor components need not be postulated separately. It follows from other more basic postulates on the space-times under investigation. One of these is that the space-time be asymptotically simple, but not quite in the sense of Penrose.[4] Penrose's requirement that every null geodesic have two end points on $\mathscr{I}$ is removed. Our basic assumptions are the following:

1) The space-time is asymptotically simple in the weakened sense just described;

2) the conformal boundary consists in part of a null hypersurface $\mathscr{I}^+$ with local topology $S^2\times R$, $\hat{\nabla}_a\Omega$ is a (nonzero) null vector on $\mathscr{I}^+$;

3) the unphysical Weyl tensor vanishes on $\mathscr{I}^+$;

4) the transformed Ricci tensor $\hat{R}_{ab}$ is finite and smooth on $\mathscr{I}^+$.

These requirements are weaker than those of asymptotic flatness. They guarantee the proper local asymptotic behavior, i.e., the proper "fall-off" properties, but the global behavior of the space-time in the asymptotic region may be completely different from that of Minkowski space. For instance, the NUT (Newman–Unti–Tamburino) solution[8] of Einstein's equation is included in the space-times under consideration.[6]

The notation used is essentially the same as in Refs. 5, 6, and 7. Careted quantities refer to the unphysical space-time $\hat{M}$ and those without carets refer to the physical space-time $M$. Superscripts on a careted variable denote the appropriate coefficient in the expansion of that variable in powers of the conformal factor $\Omega$; similarly, superscripts on an uncareted variable refer to the expansion of that variable in powers of $r^{-1}$. A dot denotes differentiation with respect to $\hat{u}$ or $u$. Re stands for real part, Im for imaginary part.

## 2. CHOICE OF FRAME AND COORDINATE SYSTEM

In Ref. 5 we showed in detail how to pick the conformal factor $\Omega$ and the tetrad $\{\hat{k}^a,\hat{m}^a,\bar{\hat{m}}^a,\hat{n}^a\}$ on the boundary $\mathscr{I}^+$ in such a way that, on $\mathscr{I}^+$, we have $\hat{n}_a = -\hat{\nabla}_a\Omega$ and

$$\begin{aligned}&\hat{\rho}^0=\hat{\tau}^0=\hat{\lambda}^0=\hat{\nu}^0=0,\\ &\hat{\tilde{\alpha}}^0=-\hat{\beta}^0=\tfrac{1}{2}\hat{\delta}^0\ln P,\\ &2\hat{\gamma}^0=-\hat{\mu}^0=\hat{\Delta}\ln P,\end{aligned}\tag{2.1}$$

where $P$ is an arbitrary real function on $\mathscr{I}^+$. (Actually, this was done only for $\hat{\Delta}P=0$, but the generalization is immediate.) The geodesics of $\hat{M}$ arriving at points on $\mathscr{I}^+$ from the $\hat{k}^a$-directions were shown to be hypersurface orthogonal. Thus $\hat{k}_a=\hat{\nabla}_a\hat{u}$ with the $\hat{u}=$ constant hypersurfaces cutting $\mathscr{I}^+$ in 2-surfaces with coordinates $\hat{x}^3,\hat{x}^4$. Each geodesic could

 0022-2488/80/102538-05$1.00  

therefore be labelled by the $\hat{u},\hat{x}^3$ and $\hat{x}^4$ coordinates of its intersection with $\mathscr{I}^+$. The coordinate system $\{\hat{u},\Omega,\hat{x}^3,\hat{x}^4\}$ is the type I coordinate system (for $\hat{M}$) referred to in the Introduction and its analogue for the physical space-time $M$ is the one most commonly used when discussing asymptotic properties.

We obtain a natural frame in the interior of $\hat{M}$ by propagating parallelly the tetrad chosen on $\mathscr{I}^+$ into the interior along each geodesic. The spin-coefficients $\hat{\kappa},\hat{\pi},\hat{\epsilon}$ then vanish identically. The metric components $\hat{f}$, $\hat{\omega}$, $\hat{\xi}^i$, $\hat{U}$, and $\hat{X}^i$ are defined by

$$\hat{D}=\hat{f}\frac{\partial}{\partial\Omega},\quad \hat{\delta}=\hat{\omega}\frac{\partial}{\partial\Omega}+\hat{\xi}^i\frac{\partial}{\partial\hat{x}^i},$$
$$\hat{\Delta}=\hat{U}\frac{\partial}{\partial\Omega}+\frac{\partial}{\partial\hat{u}}+\hat{X}^i\frac{\partial}{\partial\hat{x}^i}\quad (i=3,4).$$

On $\mathscr{I}^+$ they take on the values

$$\hat{f}^0=-1,\quad \hat{\omega}^0=\hat{U}^0=\hat{X}^{i0}=0,$$
$$\hat{\xi}^{30}=-i\hat{\xi}^{40}=P(\hat{u},\hat{x}^3,\hat{x}^4).$$

The freedom left in the choice of tetrad, conformal factor and coordinate system is that of the Newman–Unti group. Requiring $\partial P/\partial u=0$, as we did in Ref. 5, reduces this freedom to that of the Bondi–Metzner–Sachs group.

A type II coordinate system is introduced in a similar fashion. But first we must make a null rotation about $\hat{n}_a$ at each point of $\mathscr{I}^+$ so that in the new frame the shear $\hat{\sigma}$ vanishes. Since the transformation formulas for tetrad changes are readily available in the literature,[1,5,7] they will not be listed here again. For the shear $\hat{\sigma}$ the required formula is

$$\hat{\sigma}'=\hat{\sigma}+b\hat{\tau}+2b\hat{\beta}+b^2(\hat{\mu}+2\hat{\gamma})+b^3\hat{\nu}-b\hat{\Delta}b-\hat{\delta}b.$$

Let $-L$ be a value of the parameter $b$ for which $\hat{\sigma}'$ vanishes. Using Eq. (2.1) we have

$$0=\hat{\sigma}+\hat{\delta}'L+L\hat{\delta}'\ln P+L^2\hat{\Delta}'\ln P,$$

all quantities being evaluated on $\mathscr{I}^+$. We also have the relations

$$\hat{D}'=\hat{D}-\overline{L\hat{\delta}}-L\hat{\delta}+L\overline{L\hat{\Delta}},\quad \hat{\delta}'=\hat{\delta}-L\hat{\Delta},\quad \hat{\Delta}'=\hat{\Delta}.$$

Since it is the primed frame that is to be used exclusively from now on we shall, for convenience, drop all primes in the following. Some of the remaining spin-coefficients are given (in the new frame) by

$$\hat{\lambda}^0=\hat{\nu}^0=0,\quad \hat{\mu}^0=-2\hat{\gamma}^0=\frac{-\partial}{\partial\hat{u}}\ln P,$$
$$\hat{\tau}^0=P\frac{\partial}{\partial\hat{u}}(LP^{-1}),\quad \hat{\alpha}^0=-\hat{\beta}^0=\tfrac{1}{2}\hat{\delta}^0\ln P \tag{2.2}$$

and by

$$\hat{\rho}^0=\hat{\delta}^0L-L\hat{\delta}^0\ln P. \tag{2.3}$$

To obtain a type II coordinate system $\{\hat{u},\hat{\Omega},\hat{x}^3,\hat{x}^4\}$ we take a point on the geodesic leaving a point $S$ on $\mathscr{I}^+$ for the interior of $\hat{M}$ in the (new) $\hat{k}^a$-direction and label it by the coordinates $(\hat{u},\hat{x}^3,\hat{x}^4)$ of $S$. Again, the conformal factor serves as a fourth coordinate. The geodesic congruence just defined is no longer hypersurface orthogonal in general; instead, it is asymptotically shearfree. The tetrad we have at $S$ is parallelly propagated along the geodesic through $S$, thus providing a tetrad at each point of $\hat{M}$ and making the spin-coefficients $\hat{\kappa},\hat{\pi},\hat{\epsilon}$ vanish identically throughout $\hat{M}$.

In a manner similar to what was done in the type I case the metric components are defined by

$$\hat{D}=\hat{f}\frac{\partial}{\partial\Omega},\quad \hat{\delta}=\hat{\omega}\frac{\partial}{\partial\Omega}+\hat{\xi}^i\frac{\partial}{\partial\hat{x}^i},$$
$$\hat{\Delta}=\hat{U}\frac{\partial}{\partial\Omega}+\hat{X}^i\frac{\partial}{\partial\hat{x}^i}\quad (i=1,3,4), \tag{2.4}$$

where $\hat{x}^1=\hat{u}$. Since $\hat{D}$, $\hat{\delta}$ and $\hat{\Delta}$ are known on $\mathscr{I}^+$ (in terms of type I quantities) it is readily seen that

$$\hat{f}^0=-1,\quad \hat{\omega}^0=\hat{U}^0=\hat{X}^{30}=\hat{X}^{40}=0,$$
$$\hat{X}^{10}=1,\quad \hat{\xi}^{10}=-L,\quad \hat{\xi}^{30}=-i\hat{\xi}^{40}=P. \tag{2.5}$$

Again, $\hat{D}^2\Omega=0$ on $\mathscr{I}^+$ implies that $\hat{f}^{(1)}=0$ and hence

$$\hat{f}=-1+\hat{f}^{(2)}\Omega^2+\hat{f}^{(3)}\Omega^3+O(\Omega^4). \tag{2.6}$$

Furthermore, choosing the conformal factor such that $\mathrm{Re}\hat{\rho}=0$, Eq. (4.2a) of Ref. 2 implies that $\hat{D}\hat{\rho}=0$. Therefore, from Eq. (2.3) we see that the twist $\Sigma(\hat{u},\hat{x}^3,\hat{x}^4)$ is given in terms of $P$ and $L$ by

$$-i\hat{\rho}=\Sigma=\mathrm{Im}(\hat{\delta}^0L-L\hat{\delta}^0\ln P). \tag{2.7}$$

In summary, for a type II coordinate system and associated tetrad and conformal factor the following conditions hold. Equation (2.7) and the equations

$$\hat{\kappa}=\hat{\pi}=\hat{\epsilon}=0$$

hold identically in $\hat{M}$. The remaining spin-coefficients and the metric variables take, on $\mathscr{I}^+$, the values given by Eqs. (2.2) and (2.5) and by $\hat{\sigma}^0=0$. Furthermore, by assumption, the components of the Weyl tensor $\hat{\Psi}_{ABCD}$ vanish on $\mathscr{I}^+$, whereas those of the tracefree Ricci tensor $\hat{\Phi}_{AB}$ and the Ricci scalar $\hat{\Lambda}$ remain finite there.

## 3. BEHAVIOR NEAR $\mathscr{I}^+$

The procedure for finding the (careted) NP quantities for the unphysical space-time $\hat{M}$ is essentially the same as for the type I case.[5] However, instead of evaluating $\Omega$-derivatives of various expressions at $\mathscr{I}^+$ we find it easier to expand each variable in powers of $\Omega$, substitute into the careted NP equations and compare coefficients. The equations we need here are the metric equations, the radial and nonradial "Ricci identities" [Eqs. (4.2) of Ref. 2] and the transformation equations for the Ricci tensor.[5] The Bianchi identities are not required since they are essentially compatibility conditions on the "Ricci identities." Only the metric equations will be exhibited here since they are different for each coordinate system. The other equations will not be listed again due to their length and ready accessibility in the literature.

From the commutator relations,[2] applied to the coordinates, we easily determine the radial metric equations to be

$$\hat{D}\hat{\xi}^i-\hat{\bar{\rho}}\hat{\xi}^i-\hat{\sigma}\hat{\bar{\xi}}^i=0,$$
$$\hat{D}\hat{X}^i-\hat{\bar{\tau}}\hat{\xi}^i-\hat{\tau}\hat{\bar{\xi}}^i=0,$$
$$\hat{D}\hat{\omega}-\hat{\delta}\hat{f}-\hat{\bar{\rho}}\hat{\omega}+(\hat{\bar{\alpha}}+\hat{\beta})\hat{f}-\hat{\sigma}\hat{\bar{\omega}}=0, \tag{3.1}$$
$$\hat{D}\hat{U}-\hat{\Delta}\hat{f}-\hat{\tau}\hat{\omega}-\hat{\bar{\tau}}\hat{\omega}+(\hat{\gamma}+\hat{\bar{\gamma}})\hat{f}=0,$$

and the nonradial ones to be

$$
\begin{aligned}
&\hat{\delta}\hat{\bar{\xi}}^i - \hat{\bar{\delta}}\hat{\xi}^i + (\hat{\bar{\rho}} - \hat{\rho})\hat{X}^i + (\hat{\beta} - \hat{\bar{\alpha}})\hat{\bar{\xi}}^i + (\hat{\alpha} - \hat{\bar{\beta}})\hat{\xi}^i = 0\,,\\
&\hat{\delta}\widehat{\bar{\omega}} - \hat{\bar{\delta}}\widehat{\omega} + (\hat{\bar{\mu}} - \hat{\mu})\hat{f} + (\hat{\bar{\rho}} - \hat{\rho})\widehat{U}\\
&\qquad + (\hat{\beta} - \hat{\bar{\alpha}})\widehat{\bar{\omega}} + (\hat{\alpha} - \hat{\bar{\beta}})\widehat{\omega} = 0\,,\\
&\hat{\Delta}\hat{\xi}^i - \hat{\delta}\hat{X}^i + (\hat{\tau} - \hat{\bar{\alpha}} - \hat{\beta})\hat{X}^i + (\hat{\mu} - \hat{\gamma} + \hat{\bar{\gamma}})\hat{\xi}^i + \hat{\lambda}\hat{\bar{\xi}}^i = 0\,,\\
&\hat{\Delta}\widehat{\omega} - \hat{\delta}\widehat{U} - \hat{\bar{\nu}}\hat{f} + (\hat{\tau} - \hat{\bar{\alpha}} - \hat{\beta})\widehat{U}\\
&\qquad + \hat{\bar{\lambda}}\widehat{\bar{\omega}} + (\hat{\mu} - \hat{\gamma} + \hat{\bar{\gamma}})\widehat{\omega} = 0\,.
\end{aligned}
\tag{3.2}
$$

In lowest order we find, from the nonradial Ricci identities, expressions for the components of the Ricci tensor $\widehat{\Phi}_{AB}$ and $\widehat{\Lambda}$ at $\mathscr{I}^+$ in terms of the spin-coefficients and their (nonradial) derivatives,

$$
\begin{aligned}
&\widehat{\Phi}^0_{00} = \Sigma^2, \quad \widehat{\Phi}^0_{01} = i\hat{\delta}^0\Sigma - 2i\Sigma\hat{\tau}^0,\\
&\widehat{\Phi}^0_{02} = \hat{\delta}^0\hat{\tau}^0 + \hat{\tau}^0(2\hat{\bar{\alpha}}^0 - \hat{\tau}^0)\,,\\
&\widehat{\Phi}^0_{12} = 2\hat{\delta}^0\hat{\gamma}^0, \quad \widehat{\Phi}^0_{22} = 2\dot{\hat{\gamma}}^0,\\
&2\widehat{\Lambda}^0 = \mathrm{Re}[\hat{\bar{\delta}}^0\hat{\tau}^0 - \hat{\tau}^0\hat{\tau}^0 - 2\hat{\alpha}^0\hat{\tau}^0]\,,\\
&\widehat{\Phi}^0_{11} = -\widehat{\Lambda}^0 + \hat{\delta}^0\hat{\alpha}^0 + \hat{\bar{\delta}}^0\hat{\bar{\alpha}}^0 - 4\,\hat{\alpha}^0\hat{\bar{\alpha}}^0\,,
\end{aligned}
\tag{3.3}
$$

as well as the two identities

$$
\begin{aligned}
&\hat{\bar{\delta}}^0\hat{\gamma}^0 - \hat{\alpha}^0 + 2\hat{\alpha}^0\hat{\gamma}^0 - \hat{\gamma}^0\hat{\tau}^0 = 0\,,\\
&\mathrm{Im}[\hat{\bar{\delta}}^0\hat{\tau}^0 - 2\hat{\alpha}^0\hat{\tau}^0] + 4\Sigma\hat{\gamma}^0 = \dot{\Sigma}\,.
\end{aligned}
$$

To first and second order we find

$$
\begin{aligned}
&\hat{\sigma}^{(1)} = 0, \quad \hat{\tau}^{(1)} = -i\Sigma\hat{\tau}^0 - \widehat{\Phi}^0_{01}, \quad \hat{\alpha}^{(1)} = -i\Sigma\hat{\alpha}^0 - \widehat{\Phi}^0_{10}, \quad \hat{\beta}^{(1)} = -i\Sigma\hat{\bar{\alpha}}^0\,,\\
&\hat{\gamma}^{(1)} = -\hat{\tau}^0\hat{\alpha}^0 + \hat{\bar{\tau}}^0\hat{\bar{\alpha}}^0 + \widehat{\Lambda}^0 - \widehat{\Phi}^0_{11}, \quad \hat{\lambda}^{(1)} = -\widehat{\Phi}^0_{20}, \quad \hat{\mu}^{(1)} = -2i\Sigma\hat{\gamma}^0 - 2\,\widehat{\Lambda}^0,\\
&\hat{\nu}^{(1)} = 2\hat{\gamma}^0\hat{\bar{\tau}}^0 - \widehat{\Phi}^0_{21}, \quad \hat{f}^{(1)} = \widehat{\omega}^{(1)} = 0, \quad \hat{\xi}^{i(1)} = i\Sigma\hat{\xi}^{i0}, \quad \widehat{U}^{(1)} = -2\hat{\gamma}^0\,,\\
&\hat{X}^{i(1)} = -\hat{\tau}^0\hat{\bar{\xi}}^{i0} - \hat{\bar{\tau}}^0\hat{\xi}^{i0}, \quad \widehat{\Phi}^{(1)}_{00} = 0\,,\\
&\widehat{\Phi}^{(1)}_{01} = \widehat{\Psi}^{(1)}_1 + 4\,i\Sigma\widehat{\Phi}^0_{01} - 4\hat{\tau}^0\Sigma^2, \quad \widehat{\Phi}^{(1)}_{02} = -\hat{\delta}^0\widehat{\Phi}^0_{01} - 2\,\hat{\bar{\alpha}}^0\widehat{\Phi}^0_{01} + i\Sigma\widehat{\Phi}^0_{02}\,,\\
&\widehat{\Phi}^{(1)}_{11} = \hat{\gamma}^0\Sigma^2 - \Sigma\dot{\Sigma} + \mathrm{Re}\big[\tfrac{3}{2}\widehat{\Psi}^{(1)}_2 - \tfrac{1}{2}\hat{\delta}^0\widehat{\Phi}^0_{10} + (\hat{\alpha}^0 - 2\,\hat{\bar{\tau}}^0)\widehat{\Phi}^0_{01}\big]\,,\\
&\widehat{\Phi}^{(1)}_{12} = \hat{\delta}^0(\widehat{\Lambda}^0 - \widehat{\Phi}^0_{11}) + i\Sigma\widehat{\Phi}^0_{12} - 2\hat{\gamma}^0\widehat{\Phi}^0_{01}\,,\\
&\widehat{\Phi}^{(1)}_{22} = \dot{\Lambda}^0 - \dot{\Phi}^0_{11} - 2\hat{\tau}^0\hat{\bar{\delta}}\hat{\gamma}^0 - 2\,\hat{\bar{\tau}}^0\hat{\delta}^0\hat{\gamma}^0\,,\\
&\widehat{\Lambda}^{(1)} = \Sigma\dot{\Sigma} - 3\,\hat{\gamma}^0\Sigma^2 + \mathrm{Re}[(\hat{\alpha}^0 + 2\,\hat{\bar{\tau}}^0)\widehat{\Phi}^0_{01} - \tfrac{1}{2}\hat{\delta}^0\widehat{\Phi}^0_{10} - \tfrac{1}{2}\widehat{\Psi}^{(1)}_2]\,,\\
&\widehat{\Psi}^{(1)}_2 - \overline{\Psi}^{(1)}_2 = -4\,i\Sigma\widehat{\Phi}^0_{11} + 2i\,\mathrm{Im}[\hat{\delta}^0\widehat{\Phi}^0_{10} + 2\,\hat{\alpha}^0\widehat{\Phi}^0_{01}]\,,\\
&\widehat{\Psi}^{(1)}_3 = -2\,\hat{\delta}^0\hat{\bar{\delta}}^0\hat{\alpha}^0 + \hat{\delta}^0\widehat{\Phi}^0_{20} - 4\,\hat{\bar{\alpha}}^0\widehat{\Phi}^0_{20} + 8\,\hat{\bar{\alpha}}^0\hat{\bar{\delta}}^0\hat{\alpha}^0,\\
&\widehat{\Psi}^{(1)}_4 = \hat{\bar{\delta}}^0\hat{\bar{\tau}}^0 - 2\,\hat{\bar{\delta}}^0\hat{\bar{\delta}}^0\hat{\gamma}^0 + (2\hat{\alpha}^0 - 3\,\hat{\bar{\tau}}^0)(\hat{\bar{\tau}}^0 - 2\,\hat{\bar{\delta}}^0\hat{\gamma}^0)\,,\\
&\hat{\sigma}^{(2)} = -\tfrac{1}{2}\widehat{\Psi}^{(1)}_0, \quad \hat{\tau}^{(2)} = -\widehat{\Psi}^{(1)}_1 - \tfrac{3}{2}i\Sigma\widehat{\Phi}^0_{01} + \tfrac{3}{2}\Sigma^2\hat{\tau}^0\,,\\
&\hat{\alpha}^{(2)} = -\tfrac{1}{2}\overline{\Psi}^{(1)}_1 + \tfrac{5}{2}i\Sigma\widehat{\Phi}^0_{10} + \Sigma^2(2\hat{\bar{\tau}}^0 - \tfrac{1}{2}\hat{\alpha}^0), \quad \hat{\beta}^{(2)} = -\tfrac{1}{2}\widehat{\Psi}^{(1)}_1 + \tfrac{1}{2}\Sigma^2\hat{\bar{\alpha}}^0\,,\\
&\hat{\gamma}^{(2)} = \mathrm{Re}\big[-\tfrac{3}{2}\widehat{\Psi}^{(1)}_2 - 2\,\hat{\gamma}^0\Sigma^2 + \Sigma\dot{\Sigma} + \tfrac{5}{2}\hat{\bar{\tau}}^0\widehat{\Phi}^0_{01}\big] + i\,\mathrm{Im}\big[\tfrac{1}{2}\hat{\bar{\delta}}^0\widehat{\Phi}^0_{01} + i\Sigma\widehat{\Phi}^0_{11} + 2i\Sigma\hat{\alpha}^0\hat{\tau}^0 + \tfrac{1}{2}\hat{\tau}^0\widehat{\Phi}^0_{10}\big]\,,\\
&\hat{\lambda}^{(2)} = i\Sigma\widehat{\Phi}^0_{20} + \tfrac{1}{2}\hat{\bar{\delta}}^0\widehat{\Phi}^0_{10} + \hat{\alpha}^0\widehat{\Phi}^0_{10}\,,\\
&\hat{\mu}^{(2)} = i\Sigma(\widehat{\Phi}^0_{11} - \widehat{\Lambda}^0) + 4\hat{\gamma}^0\Sigma^2 + \tfrac{1}{2}\hat{\bar{\delta}}^0\widehat{\Phi}^0_{01} - (\hat{\alpha}^0 + \hat{\bar{\tau}}^0)\widehat{\Phi}^0_{01} - \hat{\tau}^0\widehat{\Phi}^0_{10} - \Sigma\dot{\Sigma}\,,\\
&\hat{\nu}^{(2)} = \hat{\bar{\delta}}^0(\widehat{\Phi}^0_{11} - \widehat{\Lambda}^0) - \tfrac{1}{2}\dot{\Phi}^0_{10} + \hat{\bar{\tau}}^0(\widehat{\Lambda}^0 - \widehat{\Phi}^0_{11}) + i\Sigma\widehat{\Phi}^0_{21} + 2\,\hat{\gamma}^0\widehat{\Phi}^0_{10} + 2i\Sigma\hat{\gamma}^0\hat{\bar{\tau}}^0\,,\\
&\hat{f}^{(2)} = \tfrac{1}{2}(\Sigma^2 - \Phi^{(4)}_{00}), \quad \widehat{\omega}^{(2)} = \tfrac{1}{2}\widehat{\Phi}^0_{01}, \quad \hat{\xi}^{i(2)} = -\tfrac{1}{2}\Sigma^2\hat{\xi}^{i0}\,,\\
&\widehat{U}^{(2)} = \widehat{\Phi}^0_{11} - \widehat{\Lambda}^0, \quad \hat{X}^{i(2)} = (i/2)\hat{\bar{\xi}}^{i0}\hat{\delta}^0\Sigma - (i/2)\hat{\xi}^{i0}\hat{\bar{\delta}}^0\Sigma\,.
\end{aligned}
\tag{3.4}
$$

Without having to go to second order in the nonradial equations we also obtain

$$
\begin{aligned}
&\hat{\sigma}^{(3)} = -\tfrac{1}{3}\widehat{\Psi}^{(2)}_0, \quad \hat{\sigma}^{(4)} = -\tfrac{1}{4}(\widehat{\Psi}^{(3)}_0 + \tfrac{1}{2}\Sigma^2\widehat{\Psi}^{(1)}_0 - \tfrac{1}{2}\Phi^{(4)}_{00}\widehat{\Psi}^{(1)}_0)\,,\\
&\hat{\sigma}^{(5)} = -\tfrac{1}{5}(\widehat{\Psi}^{(4)}_0 + \hat{f}^{(2}\widehat{\Psi}^{(2)}_0 + \hat{f}^{(3}\widehat{\Psi}^{(1)}_0)\,,\\
&\hat{\xi}^{i(3)} = -\tfrac{1}{6}i\Sigma\Phi^{(4)}_{00}\hat{\xi}^{i0} + \tfrac{1}{6}\widehat{\Psi}^{(1)}_0\hat{\bar{\xi}}^{i0}\,,\\
&\hat{\xi}^{i(4)} = \tfrac{1}{4}i\Sigma\hat{\xi}^{i0}(-\tfrac{1}{3}\Phi^{(5)}_{00} + \tfrac{1}{2}i\Sigma^3 - \tfrac{2}{3}i\Sigma\Phi^{(4)}_{00}) + \tfrac{1}{12}\hat{\bar{\xi}}^{i0}(\widehat{\Psi}^{(2)}_0 - i\Sigma\widehat{\Psi}^{(1)}_0)\,,\\
&\hat{f}^{(3)} = -\tfrac{1}{3}\Phi^{(5)}_{00}\,,\\
&\widehat{\omega}^{(3)} = \tfrac{1}{3}\widehat{\Psi}^{(1)}_1 + \tfrac{1}{6}\hat{\delta}^0\Phi^{(4)}_{00} + \tfrac{4}{3}i\Sigma\widehat{\Phi}^0_{01} - \tfrac{4}{3}\hat{\tau}^0\Sigma^2\,,\\
&\widehat{U}^{(3)} = 2\Lambda^{(3)} - \Sigma\dot{\Sigma} + 2\,\hat{\gamma}^0\Sigma^2 + \mathrm{Re}[\widehat{\Psi}^{(1)}_2 - 2\,\hat{\tau}^0\widehat{\Phi}^0_{10}]\,,\\
&\hat{X}^{i(3)} = -\tfrac{1}{3}\mathrm{Re}[\hat{\bar{\xi}}^{i0}(\hat{\tau}^0\Sigma^2 - \hat{\tau}^0\Phi^{(4)}_{00} - i\Sigma\widehat{\Phi}^0_{01} - 2\,\widehat{\Psi}^{(1)}_1)]\,,\\
&\widehat{\Phi}^{(2)}_{00} = \widehat{\Phi}^{(3)}_{00} = 0, \quad \widehat{\Phi}^{(4)}_{00} = -\tfrac{1}{4}|\widehat{\Psi}^{(1)}_0|^2\,.
\end{aligned}
\tag{3.5}
$$

The transformation equations for the Ricci tensor determine the following asymptotic behavior for the components of the physical Ricci tensor,

$$\Phi_{00} = \Phi_{00}^{(4)}\Omega^4 + \Phi_{00}^{(5)}\Omega^5 + O(\Omega^6),$$
$$\Phi_{01} = \Phi_{01}^{(4)}\Omega^4 + O(\Omega^5), \quad \Phi_{02} = \Phi_{02}^{(4)}\Omega^4 + O(\Omega^5), \tag{3.6}$$
$$\Phi_{11} = \Phi_{11}^{(3)}\Omega^3 + O(\Omega^4), \quad \Phi_{12} = \Phi_{12}^{(3)}\Omega^3 + O(\Omega^4),$$
$$\Phi_{22} = \Phi_{22}^{(2)}\Omega^2 + O(\Omega^3), \quad \Lambda = \Lambda^{(3)}\Omega^3 + O(\Omega^4).$$

In addition, from lowest order, we find that

$$\dot{\Phi}_{00}^{(4)} = 4\hat{\gamma}^0\Phi_{00}^{(4)} + 12\Lambda^{(3)}, \tag{3.7}$$
$$\Phi_{11}^{(3)} = -3\Lambda^{(3)}, \quad \Phi_{01}^{(4)} = -\tfrac{1}{2}\hat{\delta}^0\Phi_{00}^{(4)}.$$

In the next highest order these transformation equations give expressions for $\hat{\Phi}_{AB}^{(2)}$ and $\hat{\Lambda}^{(2)}$ which, when substituted into the second-order nonradial Ricci identities, yield the following relations between the "initial data":

$$\begin{aligned}
\hat{\dot{\Psi}}_0^{(1)} &= \hat{\delta}^0\hat{\Psi}_1^{(1)} + (2\hat{\bar{\alpha}}^0 - 4\hat{\tau}^0)\hat{\Psi}_1^{(1)} + 6\hat{\gamma}^0\hat{\Psi}_0^{(1)} + 3\Phi_{02}^{(4)} \\
&\quad - \tfrac{1}{2}\hat{\delta}^0\hat{\delta}^0\Phi_{00}^{(4)} - \hat{\bar{\alpha}}^0\hat{\delta}^0\Phi_{00}^{(4)} + \Phi_{00}^{(4)}\hat{\Phi}_{02}^0, \\
\hat{\dot{\Psi}}_1^{(1)} &= \hat{\delta}^0\hat{\Psi}_2^{(1)} - 3\hat{\tau}^0\hat{\Psi}_2^{(1)} + 6\hat{\gamma}^0\hat{\Psi}_1^{(1)} + 2\Phi_{12}^{(3)} - 4\hat{\delta}^0\Lambda^{(3)}, \\
\hat{\dot{\Psi}}_2^{(1)} &= \hat{\delta}^0\hat{\Psi}_3^{(1)} - (2\hat{\tau}^0 + 2\hat{\bar{\alpha}}^0)\hat{\Psi}_3^{(1)} + 6\hat{\gamma}^0\hat{\Psi}_2^{(1)} - 2\dot{\Lambda}^{(3)} + 12\hat{\gamma}^0\Lambda^{(3)} + \Phi_{22}^{(2)},
\end{aligned} \tag{3.8}$$
$$\begin{aligned}
\dot{\Phi}_{00}^{(5)} &= \mathrm{Re}[-\hat{\delta}^0\hat{\bar{\delta}}^0\Phi_{00}^{(4)} + (2\hat{\alpha}^0 + 4\hat{\bar{\tau}}^0)\hat{\delta}^0\Phi_{00}^{(4)} + 12\Lambda^{(4)} \\
&\quad + 6\hat{\gamma}^0\Phi_{00}^{(5)} + \Phi_{00}^{(4)}(6\hat{\Lambda}^0 - 2\hat{\Phi}_{11}^0)].
\end{aligned}$$

## 4. ASYMPTOTIC BEHAVIOR IN PHYSICAL SPACE-TIME

Before we can discuss the asymptotic behavior of the NP variables in the physical space-time $M$ we must introduce a suitable coordinate system in $M$. This is done as for the type I case. The conformal factor $\Omega$ is replaced as a coordinate by an affine parameter $r$ along each null geodesic. The remaining coordinates $x^1(=u)$, $x^3$ and $x^4$ remain the same. Thus,

$$\hat{x}^i = x^i, \quad \Omega = \Omega(r,x^i), \quad (i = 1,3,4),$$

where the function $\Omega(r,x^i)$ is yet to be found. The spin-frame $\{o^A,\iota^A\}$ associated with $\{\hat{o}^A,\hat{\iota}^A\}$ as in Ref. 5 does not correspond to a convenient tetrad since the spin-coefficient $\pi$ does not vanish. Therefore, we rotate the tetrad at each point about $k^a$ with parameter $c$ given by

$$c = c^0 - \hat{\bar{\omega}}^{(2)}\Omega - \tfrac{1}{2}\hat{\bar{\omega}}^{(3)}\Omega^2 + O(\Omega^3).$$

Although $c^0$ is arbitrary a convenient choice is $c^0 = \hat{\bar{\tau}}^0$ since this will simplify several of the spin-coefficients. As in the type I case we find quite readily that

$$\Omega = r^{-1} - \hat{f}^{(2)}r^{-3} - \tfrac{1}{2}\hat{f}^{(3)}r^{-4} + O(r^{-5}), \tag{4.1}$$

and with the aid of the transformation formulas in the Appendix of Ref. 7 the asymptotic expansions of all variables are now quite easily worked out in powers of $r^{-1}$. The final results are

$$\begin{aligned}
\Psi_0 &= r^{-5}\hat{\Psi}_0^{(1)} + r^{-6}\hat{\Psi}_0^{(2)} + r^{-7}[\hat{\Psi}_0^{(3)} - \tfrac{5}{2}\Sigma^2\hat{\Psi}_0^{(1)} + \tfrac{5}{2}\Phi_{00}^{(4)}\hat{\Psi}_0^{(1)}] + O(r^{-8}), \\
\Psi_1 &= r^{-4}\hat{\Psi}_1^{(1)} + O(r^{-5}), \quad \Psi_2 = r^{-3}\hat{\Psi}_2^{(1)} + O(r^{-4}), \\
\Psi_3 &= r^{-2}\hat{\Psi}_3^{(1)} + O(r^{-3}), \quad \Psi_4 = r^{-1}\hat{\Psi}_4^{(1)} + O(r^{-2}), \\
\xi^i &= \hat{\xi}^{i0}[r^{-1} + i\Sigma r^{-2} + r^{-3}(\tfrac{1}{2}\Phi_{00}^{(4)} - \Sigma^2) + r^{-4}(\tfrac{5}{6}i\Sigma\Phi_{00}^{(4)} + \tfrac{1}{6}\Phi_{00}^{(5)} - i\Sigma^3)] \\
&\quad + \tfrac{1}{6}\hat{\Psi}_0^{(1)}\hat{\bar{\xi}}^{i0}r^{-4} + O(r^{-5}), \\
\omega &= \hat{\tau}^0 - \hat{\Phi}_{01}^0 r^{-1} + r^{-2}(-\tfrac{1}{2}\hat{\Psi}_1^{(1)} + \tfrac{1}{4}\hat{\delta}^0\Phi_{00}^{(4)} - 2\hat{\tau}^0\Sigma^2 + \Sigma\hat{\delta}^0\Sigma) + O(r^{-3}), \\
X^i &= \hat{X}^{i0} + r^{-3}\mathrm{Re}[\tfrac{1}{3}\hat{\bar{\xi}}^{i0}(\hat{\Psi}_1^{(1)} - \tfrac{1}{2}\hat{\delta}^0\Phi_{00}^{(4)} + \hat{\tau}^0\Phi_{00}^{(4)})] + O(r^{-4}), \\
U &= 2\hat{\gamma}^0 r + (\hat{\Lambda}^0 - \hat{\Phi}_{11}^0 + \hat{\tau}^0\hat{\bar{\tau}}^0) - r^{-1}(\tfrac{1}{2}\hat{\Psi}_2^{(1)} + \tfrac{1}{2}\hat{\bar{\Psi}}_2^{(1)} - 2\Lambda^{(3)}) + O(r^{-2}), \\
\kappa &= \epsilon = \pi = 0, \\
\rho &= -r^{-1} + i\Sigma r^{-2} + (\Sigma^2 - \Phi_{00}^{(4)})r^{-3} + r^{-4}(i\Sigma\Phi_{00}^{(4)} - i\Sigma^3 - \tfrac{1}{2}\Phi_{00}^{(5)}) + O(r^{-5}), \\
\sigma &= -\tfrac{1}{2}\hat{\Psi}_0^{(1)}r^{-4} - \tfrac{1}{3}\hat{\Psi}_0^{(2)}r^{-5} + r^{-6}(-\tfrac{1}{4}\hat{\Psi}_0^{(3)} + \tfrac{7}{8}\Sigma^2\hat{\Psi}_0^{(1)} - \tfrac{7}{8}\Phi_{00}^{(4)}\hat{\Psi}_0^{(1)}) + O(r^{-7}), \\
\tau &= r^{-3}(-\tfrac{1}{2}\hat{\Psi}_1^{(1)} + \tfrac{1}{4}\hat{\delta}^0\Phi_{00}^{(4)} - \tfrac{1}{2}\hat{\tau}^0\Phi_{00}^{(4)}) + O(r^{-4}), \\
\alpha &= (\hat{\alpha}^0 - \hat{\bar{\tau}}^0)(r^{-1} - i\Sigma r^{-2}) \\
&\quad + r^{-3}[\tfrac{1}{4}\hat{\bar{\delta}}^0\Phi_{00}^{(4)} + (\tfrac{1}{2}\hat{\alpha}^0 - \hat{\bar{\tau}}^0)\Phi_{00}^{(4)} + \Sigma^2(\hat{\bar{\tau}}^0 - \hat{\alpha}^0)] + O(r^{-4}),
\end{aligned} \tag{4.2}$$

$$\beta = -\hat{\bar{\alpha}}^0(r^{-1} + i\Sigma r^{-2}) + r^{-3}(-\tfrac{1}{2}\hat{\Psi}_1^{(1)} + \hat{\bar{\alpha}}^0\Sigma^2 - \tfrac{1}{2}\hat{\bar{\alpha}}^0\Phi_{00}^{(4)}) + O(r^{-4}),$$

$$\gamma = -\hat{\gamma}^0 + r^{-2}(2\Lambda^{(3)} - \tfrac{1}{2}\hat{\Psi}_2^{(1)}) + O(r^{-3}),$$

$$\mu = (r^{-1} + i\Sigma r^{-2})(\hat{\Lambda}^0 - \hat{\Phi}_{11}^0 + \hat{\tau}^0\hat{\bar{\tau}}^0 + 2i\Sigma\hat{\gamma}^0 - i\dot{\Sigma}) - r^{-2}(\hat{\Psi}_2^{(1)} + 2\Lambda^{(3)}) + O(r^{-3}),$$

$$\lambda = -\tfrac{1}{2}\Phi_{20}^{(4)}r^{-3} + O(r^{-4}),$$

$$\nu = -2\hat{\bar{\delta}}^0\hat{\gamma}^0 + \hat{\bar{\tau}}^0 - r^{-1}\hat{\Psi}_3^{(1)} + O(r^{-2}),$$

$$\Phi_{00} = \Phi_{00}^{(4)}r^{-4} + \Phi_{00}^{(5)}r^{-5} + O(r^{-6}),$$

$$\Phi_{01} = (-\tfrac{1}{2}\hat{\delta}^0\Phi_{00}^{(4)} + \hat{\tau}^0\Phi_{00}^{(4)})r^{-4} + O(r^{-5}),$$

$$\Phi_{02} = \Phi_{02}^{(4)}r^{-4} + O(r^{-5}), \quad \Phi_{11} = -3\Lambda^{(3)}r^{-3} + O(r^{-4}),$$

$$\Phi_{12} = \Phi_{12}^{(3)}r^{-3} + O(r^{-4}), \quad \Phi_{22} = \Phi_{22}^{(2)}r^{-2} + O(r^{-3}), \quad \Lambda = \Lambda^{(3)}r^{-3} + O(r^{-4}),$$

where $\hat{\delta}^0 = -L(\partial/\partial\hat{u}) + P[(\partial/\partial\hat{x}^3) + i(\partial/\partial\hat{x}^4)]$ and $\Sigma, \hat{\alpha}^0, \hat{\tau}^0, \hat{\gamma}^0, \hat{\xi}^0, \hat{X}^0, \hat{\Lambda}^0, \hat{\Phi}_{01}^0, \hat{\Phi}_{11}^0, \mathrm{Im}\hat{\Psi}_2^{(1)}$, and $\hat{\Psi}_3^{(1)}$ are given in Eqs. (2.2), (2.5), (2.7), (3.3) and (3.5) in terms of functions defined on $\mathscr{I}^+$. The functions (on $\mathscr{I}^+$) $P, L, \hat{\Psi}_0^{(1)}, \hat{\Psi}_0^{(2)}, \hat{\Psi}_0^{(3)}, \hat{\Psi}_1^{(1)}, \mathrm{Re}\hat{\Psi}_2^{(1)}, \Phi_{00}^{(4)}, \Phi_{00}^{(5)}, \Lambda^{(3)}$ are part of the initial data and are subject only to equations obtained from Eqs. (3.7) and (3.8) by replacing there $\Phi_{01}^{(4)}$ by $\Phi_{01}^{(4)} - \hat{\tau}^0\Phi_{00}^{(4)}, \Phi_{02}^{(4)}$ by $\Phi_{02}^{(4)} - (\hat{\tau}^0)^2\Phi_{00}^{(4)} + \hat{\tau}^0\hat{\delta}^0\Phi_{00}^{(4)}$, and $\Phi_{12}^{(3)}$ by $\Phi_{12}^{(3)} - 2\hat{\tau}^0\Phi_{11}^{(3)}$. These changes are necessitated by the null rotation about $k^a$ made at the beginning of this section.

In the vacuum case these expansions agree with those of Aronson and Newman[1] to the orders given there, up to changes in notation and a few minor misprints. If we expand the special solution derived in Ref. 6 in powers of $r^{-1}$ we again get agreement with the results of this paper. Moreover, these results can easily be, and have been, checked by direct substitution into the NP equations.

The contravariant components of the metric can now be obtained from

$$\tfrac{1}{2}g^{ab} = k^{(a}n^{b)} - m^{(a}\bar{m}^{b)},$$

where

$$k^a = (0,1,0,0), \quad n^a = (X^1, U, X^3, X^4), \quad m^a = (\xi^1, \omega, \xi^3, \xi^4),$$

whence the covariant components are found to be

$$g_{12} = 1 + O(r^{-4}), \quad g_{22} = O(r^{-6}), \quad g_{23} + ig_{24} = L/P + O(r^{-4}),$$

$$g_{11} = -4\hat{\gamma}^0 r + 2(\hat{\Phi}_{11}^0 - \hat{\Lambda}^0 - \hat{\tau}^0\hat{\bar{\tau}}^0) + r^{-1}(\hat{\Psi}_2^{(1)} + \hat{\Psi}_2^{(1)} - 4\Lambda^{(3)}) + O(r^{-2}),$$

$$g_{13} + ig_{14} = -rP^{-2}\frac{\partial}{\partial u}(LP) + 2LP^{-1}(\hat{\Phi}_{11}^0 - \hat{\Lambda}^0 - \hat{\tau}^0\hat{\bar{\tau}}^0) + iP^{-1}(\hat{\delta}^0\Sigma - \Sigma\hat{\tau}^0)$$
$$+ r^{-1}P^{-1}[L(\hat{\Psi}_2^{(1)} + \hat{\Psi}_2^{(1)} - 4\Lambda^{(3)}) + \tfrac{2}{3}\hat{\Psi}_1^{(1)} - \tfrac{1}{3}\hat{\delta}^0\Phi_{00}^{(4)} + \tfrac{2}{3}\hat{\tau}^0\Phi_{00}^{(4)}] + O(r^{-2}), \tag{4.3}$$

$$g_{33} + g_{44} = P^{-2}\mathrm{Re}\{-r^2 - r(2L\hat{\bar{\tau}}^0 + 4L\bar{L}\hat{\gamma}^0) + \Phi_{00}^{(4)} - \Sigma^2 + 2L\bar{L}(\hat{\Phi}_{11}^0 - \hat{\Lambda}^0 - \hat{\tau}^0\hat{\bar{\tau}}^0)$$
$$+ 2i\bar{L}(\hat{\delta}^0\Sigma - \Sigma\hat{\tau}^0) + r^{-1}[\tfrac{1}{3}\Phi_{00}^{(5)} + \tfrac{2}{3}\bar{L}(2\hat{\Psi}_1^{(1)} + 2\hat{\tau}^0\Phi_{00}^{(4)} - \hat{\delta}^0\Phi_{00}^{(4)}) + L\bar{L}(2\hat{\Psi}_2^{(1)} - 4\Lambda^{(3)})]\} + O(r^{-2}),$$

$$g_{33} - g_{44} + 2ig_{34} = P^{-2}\{-2L\dot{L}r + 2L^2(\hat{\Phi}_{11}^0 - \hat{\Lambda}^0 - \hat{\tau}^0\hat{\bar{\tau}}^0) + 2iL(\hat{\delta}^0\Sigma - \Sigma\hat{\tau}^0)$$
$$+ r^{-1}[\tfrac{1}{3}\hat{\Psi}_0^{(1)} + L^2(\hat{\Psi}_2^{(1)} + \hat{\Psi}_2^{(1)} - 4\Lambda^{(3)}) + L(\tfrac{4}{3}\hat{\Psi}_1^{(1)} - \tfrac{2}{3}\hat{\delta}^0\Phi_{00}^{(4)} + \tfrac{4}{3}\hat{\tau}^0\Phi_{00}^{(4)})]\} + O(r^{-2}).$$


## ACKNOWLEDGMENT

This work was supported in part by a grant from the Natural Sciences/Engineering Research Council of Canada.



[1] B. Aronson and E. T. Newman, J. Math. Phys. **13**, 1847 (1972).
[2] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[3] E. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).
[4] R. Penrose in *Battelle Rencontres* (Benjamin, New York, 1968).
[5] G. Ludwig, Gen. Rel. Grav. **7**, 293 (1976).
[6] G. Ludwig, Gen. Rel. Grav. **9**, 1009 (1978).
[7] G. Ludwig, J. Math. Phys. **21**, 154 (1980).
[8] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).

# Space-times with geodesic, shearfree, twistfree, nonexpanding rays

Garry Ludwig
*Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

(Received 21 May 1980; accepted for publication 13 June 1980)

The Kundt class of metrics containing geodesic rays with vanishing divergence, shear and curl is obtained for more general Ricci tensors using the standard Newman–Penrose formalism. These solutions are then rederived using Penrose's conformal technique, thereby clarifying their asymptotic behavior.

## 1. INTRODUCTION

Two decades ago Kundt[1] obtained a class of space-times containing geodesic rays whose shear, twist and divergence all vanish. The solutions were restricted by the condition that the Ricci tensor be proportional to $l_a l_b$, where the vector $l_a$ is tangent to the ray congruence.

In this paper we generalize these metrics. We require that the Weyl tensor be algebraically special and that the Ricci scalar vanish, but do not, a priori, impose any further conditions on the Ricci tensor. Using the well-known Newman–Penrose[2] (NP) formalism we find, in Sec. 2, the spin-coefficients, the metric coefficients, the metric, and the tetrad components (relative to a suitable tetrad) of the Weyl and Ricci tensors for such space-times for which the repeated principal null vector of the Weyl tensor is, at each point, tangential to a congruence of shearfree, twistfree, nonexpanding null geodesics subject to one additional requirement ($\tau = 0$) to be discussed later. The Kundt solutions, which include the plane wave vacuum solutions, are a special case, as shown in Sec. 3.

The solutions found in this paper are, as usual, implicit. We derive the reduced gravitational field equations which together with the equations governing the source must still be solved before an explicit solution is obtained. In general, such equations, although few in number, are difficult to solve. Only in rare cases has this been done (cf. the explicit Robinson–Trautman[3] type solution found by Foster and Newman).[4]

In Sec. 4 we rederive our solutions with the aid of Penrose's conformal technique.[5,6] This technique has been used successfully on previous occasions for solving (both approximately near infinity and exactly) for space-times with expanding rays.[7-9] To find solutions we transform to an unphysical space-time $\hat{M}$ by rescaling the metric, solve the NP equations there and then transform back to the physical space-time $M$. The advantages of this method are as follows. The asymptotic behavior of the solution is built in from the start. At the very outset the coordinate system and the tetrad are chosen at "infinity" and need not be changed during the solving process as in the regular NP procedure. All "integration constants" have an obvious meaning, usually at "infinity." Further, the conformal factor $\Omega$ gives an extra degree of freedom which can be used to simplify the equations to be solved.

Our notation is basically the same as in Ref. 7. Careted quantities refer to the rescaled space-time $\hat{M}$; those without carets refer to the physical space-time $M$. A zero superscript on a quantity means that this quantity is independent of the radial coordinate ($r$ or $\Omega$). Ordinary differentiation is denoted by a comma, but a dot is also used to denote differentiation with respect to the coordinate $u$. Complex conjugation is denoted by a bar, symmetrization by round brackets around indices. Tensor indices are denoted by small Latin letters and usually run form 1 to 4 (except for $i$ and $j$ which run from 3 to 4 only). The usual symbols[2,10] are used for the NP quantities. Parameters $a$, $\phi$, $c$, and $\theta$ refer, respectively, to boosts (referred to in the following as "rescalings") in the $k$-$n$ plane of a standard null tetrad $\{k^a, m^a, \bar{m}^a, n^a\}$, to spatial rotations in the $m$-$\bar{m}$ plane, to null rotations about $k^a$, and to changes of the conformal factor. The transformation laws of the NP variables for tetrad and conformal changes are found in Refs. 7 and 9, the NP equations are found in Refs. 2 and 10. Because of their length these formulas and equations will not be repeated here.

## 2. NP PROCEDURE

In this section we derive the desired metric by employing the regular NP procedure,[2,11] using a step-by-step approach to construct a special coordinate system with an associated tetrad in which to present this solution. At each step we reconsider the freedom we have in the choice of such a frame and use it to simplify some quantity, thereby further restricting this freedom.

As in Ref. 11, we begin by introducing a family of null hypersurfaces $u$ = constant, and by choosing coordinates ($u$, $r$, $x^3$, $x^4$) as well as a null tetrad $\{k^a, m^a, \bar{m}^a, n^a\}$, such that

$$g^{12} = 1, \quad g^{11} = g^{1i} = 0, \quad (i = 3,4) \tag{2.1}$$

$$k_a = u_{,a}, \quad k^a = \frac{dx^a}{dr}, \tag{2.2}$$

where $r$ is an affine parameter associated with the null geodesics of the hypersurfaces. To satisfy the orthonormality conditions, $k_a n^a = -m_a \bar{m}^a = 1$ with all other inner products vanishing, the vectors $m$ and $n$ (i.e. $\delta$ and $\Delta$) must have the form

$$\delta = \omega \frac{\partial}{\partial r} + \xi^i \frac{\partial}{\partial x^i} \quad (i = 3,4)$$

$$\Delta = \frac{\partial}{\partial u} + U \frac{\partial}{\partial r} + X^i \frac{\partial}{\partial x^i}, \tag{2.3}$$

 0022-2488/80/102543-06$1.00  

where the "metric variables" $\omega, \xi^i, U, X^i$ are arbitrary functions of the coordinates.

With the aid of the transformation laws mentioned at the end of Sec. 1 we find that conditions (2.1)–(2.3) are left invariant by a combined coordinate and tetrad change given by

$$u' = \gamma(u), \quad r' = \dot{\gamma}^{-1} r + R(u, x^i), \quad x^{i'} = x^{i'}(u, x^j) \quad (i,j = 3,4)$$
$$a^2 = \dot{\gamma}, \quad \phi \text{ and } c \text{ arbitrary},$$

where $\gamma$, $R$, and $x^{i'}$ are arbitrary functions of their arguments (with $\dot{\gamma} > 0$) and the parameters $a$, $\phi$, and $c$ were defined at the end of Sec. 1. For such a transformation the tetrad vectors and the metric variables change as follows:

$$k' = \dot{\gamma} k, \quad m' = \bar{c}\dot{\gamma} k + e^{2i\phi} m,$$
$$n' = c\bar{c}\dot{\gamma} k + c e^{2i\phi} m + \bar{c} e^{-2i\phi} \bar{m} + \dot{\gamma}^{-1} n,$$
$$\omega' = \bar{c}\dot{\gamma} + e^{2i\phi}(\dot{\gamma}^{-1}\omega + R_{,i}\,\xi^i), \quad \xi^{i'} = e^{2i\phi}\frac{\partial x^{i'}}{\partial x^i}\xi^i, \tag{2.4}$$
$$U' = c\bar{c}\dot{\gamma} + c e^{2i\phi}(\dot{\gamma}^{-1}\omega + R_{,i}\xi^i) + \bar{c} e^{-2i\phi}(\dot{\gamma}^{-1}\bar{\omega} + R_{,i}\bar{\xi}^i) + \dot{\gamma}^{-1}(\dot{R} + X^i R_{,i}) - r\dot{\gamma}^{-3}\ddot{\gamma} + \dot{\gamma}^{-2} U,$$
$$X^{i'} = c e^{2i\phi}\xi^i\frac{\partial x^{i'}}{\partial x^i} + \bar{c} e^{-2i\phi}\bar{\xi}^i\frac{\partial x^{i'}}{\partial x^i} + \dot{\gamma}^{-1}\left(X^i\frac{\partial x^{i'}}{\partial x^i} + \frac{\partial x^{i'}}{\partial u}\right).$$

From Eq. (2.2) and the fact that the vector $k^a$ is tangent to a null geodesic we find that the spin-coefficients must obey

$$\kappa = \epsilon + \bar{\epsilon} = 0, \quad \rho = \bar{\rho}, \quad \tau = \bar{\alpha} + \beta.$$

Parallel propagation of the tetrad along the null geodesics leads to the further restrictions

$$\pi = \epsilon = 0$$

on the spin-coefficients and

$$Dc = D\phi = 0$$

on the tetrad freedom. Thus, the parameters $\phi$ and $c$ for spatial rotations in the $m$-$\bar{m}$ plane and null rotations about $k$ must be independent of the radial coordinate $r$.

So far the proceedings have been quite general except that we have arranged for the rays under discussion to be hypersurface orthogonal. Now we make the following assumptions:

(i) the rays are shearfree ($\sigma = 0$, and hence $\Psi_0 = 0$),

(ii) the rays are non-expanding ($\rho + \bar{\rho} = 0$, and hence $\rho = 0$),

(iii) the space-time is algebraically special ($\Psi_1 = 0$),

(iv) the Ricci scalar vanishes ($\Lambda = 0$).

(There will be one additional assumption ($\tau = 0$) made below.) Due to the generality of the Ricci tensor the Goldberg–Sachs theorem[10] is not applicable and conditions (i) and (iii) are independent. All four conditions are invariant under the coordinate-tetrad freedom described above. In summary, we have, so far, the following simplifications:

$$\kappa = \sigma = \rho = \epsilon = \pi = \Lambda = \Psi_0 = \Psi_1 = 0, \quad \tau = \bar{\alpha} + \beta.$$

The equations we need here are the radial and nonradial "Ricci identities" (Eqs. 4.2 of Ref. 2) and the metric equations (Eqs. 6.10 of Ref. 11). The Bianchi identities are not required since they are essentially compatibility conditions on the "Ricci identities". It is now a straightforward matter to solve these equations. We find, for instance that

$$\alpha = \alpha^0, \quad \beta = \beta^0, \quad \tau = \tau^0, \quad \lambda = \lambda^0 + r\Phi^0_{20},$$
$$\mu = \mu^0 + r(\bar{\xi}^{i0}\tau^0_{,i} - 2\alpha^0\tau^0), \tag{2.5}$$
$$\omega = \omega^0 - \tau^0 r, \quad \xi^i = \xi^{i0}, \quad \Phi_{00} = 0.$$

Before exhibiting the complete solution let us simplify further with the aid of our coordinate-tetrad freedom and one additional assumption. By means of the coordinate change

$$u' = u, \quad r' = r, \quad \zeta' = \zeta'(\zeta, \bar{\zeta}, u),$$

where $\zeta = -x^3 + ix^4$, we make $\xi^3 + i\xi^4$ vanish at $r = 0$ and hence, according to Eq. (2.5), everywhere. The coordinate freedom now has the additional restriction that

$$\zeta' = \zeta'(u, \zeta).$$

In the new coordinate system we define

$$P(u, x^3, x^4) = \xi^3 (= -i\xi^4)$$

so that

$$\xi^i\frac{\partial}{\partial x^i} = P\left(\frac{\partial}{\partial x^3} + i\frac{\partial}{\partial x^4}\right) = -2P\frac{\partial}{\partial\zeta} \equiv P\nabla.$$

By means of a spatial rotation the variable $P$ can be made real, restricting our coordinate-tetrad freedom further by the requirement that

$$e^{2i\phi}\frac{\partial\zeta'}{\partial\zeta} = \text{real}.$$

Under a coordinate change

$$u' = u, \quad r' = r + R(u, x^i), \quad \zeta' = \zeta$$

accompanied by a null rotation about $k$, the variables $\omega$ and $\mu$ transform as

$$\omega' = \bar{c} + \omega + P\nabla R,$$
$$\mu' = 2c\beta + \mu + P\nabla c,$$

where $c = c(u, x^3, x^4)$, $R = R(u, x^3, x^4)$. Since $\omega$ and $\mu$ are given by Eq. (2.5) we see quite readily that $c$ and $R$ may be chosen such that $\mu^0$ and $\omega^0$ vanish. Our coordinate-tetrad freedom is now given by

$$u' = \gamma(u), \quad r' = \dot{\gamma}^{-1} r + R(u, \zeta, \bar{\zeta}), \quad \zeta' = \zeta'(u, \zeta),$$

accompanied by tetrad changes with parameters restricted by

$$a^2 = \gamma(u), \quad Dc = D\phi = 0, \quad e^{2i\phi}\frac{\partial\zeta'}{\partial\zeta} = \text{real},$$
$$\bar{c}\dot{\gamma} + e^{2i\phi}(P\nabla R + \dot{\gamma}^{-1}\tau^0 R) = 0,$$
$$e^{2i\phi}(2c\beta^0 + 2icP\nabla\phi + P\nabla c) - R(\bar{\xi}^{i0}\tau^0_{,i} - 2\alpha^0\tau^0) = 0.$$

Our final assumption, $\tau^0 = 0$, is invariant under this freedom and is, therefore, a geometrical condition. The solution of the NP equations now becomes

$$\kappa = \sigma = \rho = \tau = \mu = \epsilon = \pi = 0,$$
$$\lambda = \lambda^0, \quad \alpha = \alpha^0, \quad \beta = -\bar{\alpha}^0, \quad \gamma = \gamma^0 + K^0 r,$$

$$\nu = -P\bar{\nabla}U^0 + rP\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) + r^2P\bar{\nabla}K^0,$$
$$\omega = 0, \quad \xi^i = \xi^{i0}, \quad X^i = X^{i0}, \tag{2.6}$$
$$U = U^0 - r(\gamma^0 + \bar{\gamma}^0) - r^2K^0, \Psi_0 = \Psi_1 = \Psi_2 = 0,$$
$$\Psi_3 = \tfrac{1}{2}P\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) - \tfrac{1}{2}P\nabla\lambda^0 + 2\lambda^0\bar{\alpha}^0 + rP\bar{\nabla}K^0,$$
$$\Psi_4 = [-\bar{\nabla}(P^2\bar{\nabla}U^0) - \dot{\lambda}^0 + \lambda^0(\bar{\gamma}^0 - 3\gamma^0) + \tfrac{1}{2}X^0\nabla\lambda^0 + \tfrac{1}{2}\bar{X}^0\bar{\nabla}\lambda^0]r[\bar{\nabla}(P^2\bar{\nabla}(\gamma^0 + \bar{\gamma}^0)) + K^0\bar{\nabla}X^0] + r^2\bar{\nabla}(P^2\bar{\nabla}K^0),$$
$$\Phi_{00} = \Phi_{01} = \Phi_{02} = \Lambda = 0,$$
and
$$\Phi_{11} = K^0, \Phi_{12} = \tfrac{1}{2}P\nabla(\gamma^0 + \bar{\gamma}^0) + \tfrac{1}{2}P\bar{\nabla}\bar{\lambda}^0 - 2\alpha^0\bar{\lambda}^0 + rP\nabla K^0, \tag{2.7}$$
$$\Phi_{22} = -P^2\nabla\bar{\nabla}U^0 - \lambda^0\bar{\lambda}^0 + rP^2\nabla\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) + r^2P^2\bar{\nabla}K^0,$$
where
$$\alpha^0 = \tfrac{1}{2}\bar{\nabla}P, \quad \xi^{30} = P = -i\xi^{40}, \quad K^0 = P\nabla\bar{\nabla}P - |\nabla P|^2,$$
$$\lambda^0 = -\tfrac{1}{2}\bar{\nabla}X^0, \quad \gamma^0 - \bar{\gamma}^0 = \tfrac{1}{4}(\nabla X^0 - \bar{\nabla}\bar{X}^0), \tag{2.8}$$
$$X^0 = -X^{30} + iX^{40}.$$

The variables $P$, $X^0$, $\gamma^0 + \bar{\gamma}^0$, and $U^0$ are independent of the radial coordinate $r$ and subject only to

$$\dot{P} = -\tfrac{1}{4}P(\nabla X^0 + \bar{\nabla}\bar{X}^0) + \tfrac{1}{2}\bar{X}^0\bar{\nabla}P + \tfrac{1}{2}X^0\nabla P. \tag{2.9}$$

Equations (2.7) and (2.9) are the reduced equations which must be solved for $P$, $X^0$, $\gamma^0 + \bar{\gamma}^0$, and $U^0$ before Eq. (2.6) represents an explicit solution.

The contravariant components of the metric tensor are obtained, as usual from the relation

$$g^{ab} = k^{(a}n^{b)} - m^{(a}\bar{m}^{b)};$$

here $k^a = (0,1,0,0)$, $n^a = (1, U, X^3, X^4)$, $m^a = (0,0,P,iP)$. By inverting the matrix $(g^{ab})$ we find the metric to be

$$ds^2 = \{2U + \tfrac{1}{2}P^{-2}[(X^{30})^2 + (X^{40})^2]\}du^2 - 2\,du\,dr - P^{-2}X^{30}\,dx^3\,du - P^{-2}X^{40}\,dx^4\,du + \tfrac{1}{2}P^{-2}[(dx^3)^2 + (dx^4)^2], \tag{2.10}$$

or, equivalently,

$$ds^2 = 2U\,du^2 - 2\,du\,dr + \tfrac{1}{2}P^{-2}|d\zeta - X^0\,du|^2,$$

where $U = U^0 - r(\gamma^0 + \bar{\gamma}^0) - r^2K^0$. It is now apparent the variable $P$ defines the metric of the 2-surface $u$ = constant, $r$ = const. and that $K^0$ is proportional to the Gaussian curvature of this 2-surface.

## 3. KUNDT'S METRIC AS A SPECIAL CASE

It is not difficult to see that the vanishing of the tetrad component $\Phi_{11}$ of the Ricci tensor is the necessary and sufficient condition to enable us to make the variable $P$ a constant (say $2^{-1/2}$) by means of the remaining freedom. If, in addition $\Phi_{12}$ vanishes then the reduced equations become

$$\nabla X^0 + \bar{\nabla}\bar{X}^0 = 0, \tag{3.1}$$
$$\nabla(\gamma^0 + \bar{\gamma}^0 - \tfrac{1}{2}\bar{\nabla}\bar{X}^0) = 0, \tag{3.2}$$
$$\Phi_{22} = -\tfrac{1}{2}\nabla\bar{\nabla}U^0 - \tfrac{1}{4}|\bar{\nabla}X^0|^2. \tag{3.3}$$

The reason why $\Phi_{22}$ contains no longer the term linear in $r$, namely $\frac{1}{2}r\nabla\bar{\nabla}(\gamma^0 + \bar{\gamma}^0)$, is that by Eq. (3.2) we have

$$\nabla\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) = \tfrac{1}{2}\nabla\bar{\nabla}\bar{\nabla}\bar{X},$$

where the left side is obviously real and the right side pure imaginary according to Eq. (3.1).

We intend to show that with the above specialization our solution reduces to that given by Eq. (3.9) of Kundt's paper.[1] If we let $E = \gamma^0 + \bar{\gamma}^0$ and $F = \frac{1}{2}i\bar{\nabla}\bar{X}^0$ then, by Eqs. (3.1) and (3.2), $F$ is real and $\nabla(E + iF) = 0$. Thus $E + iF$ is independent of $-\zeta$, i.e. of $x^3 - ix^4$, and is, therefore, analytic in $x^3 + ix^4$. The Cauchy–Riemann equations now show that $E$ and $F$ are harmonic in $x^3$ and $x^4$ and that there exists a real function $b^0(u, x^3, x^4)$ such that

$$E = \tfrac{1}{2}b^0_{,3} \quad \text{and} \quad F = -\tfrac{1}{2}b^0_{,4}.$$

Note that $X^0$ is also determined by $b^0$ according to

$$\frac{\partial X^0}{\partial\zeta} = \tfrac{1}{2}ib^0_{,4}.$$

Therefore, we take the real variables $b^0(u, x^3, x^4)$ and $U^0(u, x^3, x^4)$ as our undetermined variables, subject, however, to Eq. (3.3) and to $b^0_{,33} + b^0_{,44} = 0$.

To get agreement with Kundt's solution we must first perform the following change of coordinates (which will, of course, violate some of our conditions):

$$u = u', \quad r = r' + R(u', \zeta', \bar{\zeta}'), \quad \zeta = \zeta'.$$

With $R$ chosen so that $\nabla R = \bar{X}^0 + b^0$, the metric and Eq. (3.3) become, respectively,

$$ds^2 = H\,du'^2 - 2\,du'\,dr' + |d\zeta' + b^0\,du'|^2,$$
$$-4\Phi_{22} = \nabla\bar{\nabla}A^0 + \left(2b^0\frac{\partial}{\partial x^3} + b^0_{,3} + 2\frac{\partial}{\partial u}\right)b^0_{,3} - (b^0_{,4})^2,$$

where

$$H = A^0 - b^0_{,3}r',$$

and the variable $A^0$, defined by

$$A^0 = 2U^0 - \frac{\partial R}{\partial u'} + X^0\bar{X}^0 - b^0_{,3}R,$$

is independent of $r$ and replaces $U^0$ as undetermined variable.

A further change given by $r' \to -r'$ and $x^3 \to -x^3$ now yields Kundt's solution except for an unexplained difference in the numerical coefficients of $(b_{,3})^2$ and $(b_{,4})^2$.

## 4. CONFORMAL APPROACH

In this section we employ Penrose's conformal method[5,6] to rederive the metric found in Sec. 2. The advantages of this technique have been discussed in Sec. 1 and in previous papers.[7-9]

It is well known[12] that in Minkowski space null geodesics lying in parallel null hyperplanes all end up on one and the same generator $N$ of (conformal future null infinity) $\mathscr{I}^+$, with all the geodesics of a particular null hyperplane reaching the same point $S$ of $N$. These null hyperplanes are actually null cones whose vertices lie on $N$ and for which $N$ is, in fact, one of the generators.

With this in mind consider a space-time which has as (part of) its conformal boundary a line $N$ on which the conformal factor $\Omega$ vanishes and on which $\hat{\nabla}_a\Omega \neq 0$. At a point

$S$ on $N$ consider the null "cone" generated by all the possible null geodesics arriving at $S$ from the interior of the rescaled space-time $\hat{M}$. Each such null geodesic has a tangent vector $\hat{k}^a$ defined up to a proportionality factor which depends on the geodesic. Corresponding to each geodesic choose at $S$ another null vector $\hat{n}^a$ satisfying the conditions $\hat{n}_a \hat{k}^a = 1$ and

$$\hat{\nabla}_a \Omega \,|_{\Omega=0} = K^0 \hat{k}_a - \hat{n}_a \tag{4.1}$$

for some function $K^0$ also depending on the geodesic. (In general, the null vector $\hat{n}^a$ depends on the geodesic. Only when $N$ is null will $\hat{n}^a$ be the same for all null geodesics arriving at $S$; it will then equal $-\hat{\nabla}_a \Omega$ and point along $N$ just as in the Minkowski case.) Arbitrarily choose a complex null vector $\hat{m}^a$ such that $\{\hat{k}^a, \hat{m}^a, \bar{\hat{m}}^a, \hat{n}^a\}$ form a null tetrad. This gives us many null tetrads at $S$, one for each geodesic arriving there. Propagate these tetrads parallelly into the interior of $\hat{M}$ along their respective geodesics. For each point in the interior (at least near $N$) we now have precisely one tetrad defined there. The spin-coefficients $\kappa, \epsilon, \pi$ vanish identically.

Let us now consider the freedom in the choice of tetrad and conformal factor. We can, at a point $P$,

(i) rescale the tetrad with parameter $a$ depending on the geodesic on which $P$ lies, provided this is accompanied by a conformal rescaling $\Omega' = \theta\Omega$ with $\theta = a^2$;

(ii) make spatial rotations with parameter $\phi$ depending on the geodesic;

(iii) conformally rescale with parameter $\theta$, where $\theta \to 1$ as $\Omega \to 0$, provided this is accompanied by a null rotation about $\hat{k}^a$ with parameter $c$ satisfying the restrictions

$$\hat{D}c + \hat{\delta}\theta = 0,$$
$$c \to 0 \quad \text{as} \quad \Omega \to 0.$$

Since $\hat{k}_a$ is clearly hypersurface orthogonal it must be proportional to $\hat{\nabla}_a u$ for some function $u$ which labels the hypersurfaces. We can use freedom (i) above to make $\hat{k}_a$ equal to this gradient so that

$$\hat{k}_a = \hat{\nabla}_a u, \qquad \hat{\rho} = \bar{\hat{\rho}}, \quad \hat{\tau} = \bar{\hat{\alpha}} + \hat{\beta}. \tag{4.2}$$

Freedom (i) is now restricted by

$$u' = \gamma(u), \quad \theta = \gamma(u),$$

but freedom (ii) and (iii) remain unaffected.

A hypersurface $\Omega = \text{const} \neq 0$ cuts a $u = \text{const}$ hypersurface in a 2-surface on which we select two coordinates $x$ and $y$ subject to $\hat{\delta}\zeta = 0$, where $\zeta = -x + iy$. Propagate these along the geodesics to obtain a coordinate system

$$(\hat{x}^a) = (u, \Omega, x, y).$$

The freedom in the choice of the coordinate $\zeta$ is

(iv) $\zeta' = \zeta'(\zeta, \bar{\zeta}, u)$ restricted by $\hat{\delta}\zeta' = 0$.

From Eq. (4.2) it follows that

$$\hat{D}u = \hat{k}^a \hat{\nabla}_a u = 0,$$
$$\hat{\delta}u = \hat{m}^a \hat{\nabla}_a u = 0,$$
$$\hat{\Delta}u = \hat{n}^a \hat{\nabla}_a u = 1,$$

and from Eq. (4.1) that for any of the tetrads at $S$,

$$\hat{D}\Omega = \hat{k}^a \hat{\nabla}_a \Omega = -1,$$
$$\hat{\delta}\Omega = \hat{m}^a \hat{\nabla}_a \Omega = 0,$$
$$\hat{\Delta}\Omega = \hat{n}^a \hat{\nabla}_a \Omega = K^0.$$

Also clearly,

$$\hat{D}\hat{x}^i = 0.$$

Hence we have

$$\hat{D} = f \frac{\partial}{\partial \Omega}, \quad \hat{\delta} = \hat{\omega} \frac{\partial}{\partial \Omega} + \hat{\xi}^i \frac{\partial}{\partial \hat{x}^i},$$
$$\hat{\Delta} = \frac{\partial}{\partial u} + \hat{U} \frac{\partial}{\partial \Omega} + \hat{X}^i \frac{\partial}{\partial \hat{x}^i},$$

with $\hat{f} \to -1$, $\hat{\omega} \to 0$, $\hat{U} \to K^0$ as $\Omega \to 0$, thus defining the metric variables $\hat{f}$, $\hat{\omega}$, $\hat{\xi}^i$, $\hat{U}$, and $\hat{X}^i$.

Note that on the 2-surface used to define the coordinates $x$ and $y$, the relation

$$\hat{\xi}^3 = -i\hat{\xi}^4$$

holds since $\hat{\delta}\zeta = 0$. Further, the coordinate freedom (iv) may now be written

(iv) $\zeta' = \zeta'(u, \zeta)$.

Having set up our coordinate system and tetrad in a fairly general manner we now state our major assumptions. We assume that the null geodesics considered above are not only hypersurface orthogonal but also shearfree and expansionfree and correspond to a repeated principal null direction of the Weyl tensor, i.e.

$$\hat{\sigma} = \hat{\Psi}_0 = \hat{\Psi}_1 = 0,$$
$$\hat{\rho} = -\hat{f}\Omega^{-1}.$$

(Two more assumptions will be stated shortly.) The last equation follows from the assumption made that $\rho$ vanish and the transformation law[7] for $\rho$ under a conformal rescaling. The above conditions are invariant under freedom (i)–(iv) in the choice of frame (i.e. of coordinate system, tetrad and conformal factor). With the aid of freedom (iii) we can arrange that

$$\hat{\rho} = \Omega^{-1}, \quad \hat{f} = -1,$$

thereby restricting freedom (iii) by

$$\theta = (1 + R\Omega)^{-1},$$

where $R = R(u, x, y)$ is arbitrary.

The condition

$$\hat{\tau} = 0$$

is invariant under the remaining freedom. We shall assume it to hold as well as the vanishing of the Ricci scalar, i.e.,

$$\Lambda = 0.$$

Applying the commutator[2] relations to the coordinates the metric equations result. Two of these, namely

$$\hat{D}\hat{\omega} = \Omega^{-1}\hat{\omega}, \qquad \hat{D}\hat{\xi}^i = \Omega^{-1}\hat{\xi}^i \tag{4.3}$$

are easily solved. Taking into account that $\hat{\omega}$ must approach zero as $\Omega$ does, we find

$$\hat{\omega} = 0,$$

$$\hat{\xi}^i = \Omega^{-1}\xi^{i0}.$$

Let us define a function

$$P(u,x,y) = \xi^{30}(= -i\xi^{40}),$$

so that

$$\hat{\xi}^i \frac{\partial}{\partial \hat{x}^i} = \Omega^{-1} P\nabla,$$

where

$$\nabla = -2\frac{\partial}{\partial \zeta} = \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}.$$

We arrange for $P$ to be real by means of freedom (ii).

To summarize, with respect to the frame we have constructed, the following hold:

$$\hat{\kappa} = \hat{\epsilon} = \hat{\pi} = \hat{\tau} = \hat{\sigma} = \hat{\omega} = \hat{\Psi}_0 = \hat{\Psi}_1,$$

$$\hat{\rho} = \Omega^{-1}, \quad \hat{\alpha} = -\hat{\beta}, \quad \hat{f} = -1, \quad \hat{\xi}^i \frac{\partial}{\partial \hat{x}^i} = \Omega^{-1} P\nabla,$$

$$U \to K^0 \quad \text{as} \quad \Omega \to 0.$$

Let us list once more the remaining freedom in the choice of coordinate system, tetrad, and conformal factor. We can

(i) change the $u$-coordinate by $u' = \gamma(u)$ provided we also change the conformal factor by $\Omega' = \theta(u)\Omega$ with $\theta = \dot{\gamma}$ and rescale the tetrad with $a^2 = \theta$;

(ii) change the $x$ and $y$ coordinates by $\zeta' = \zeta'(u,\zeta)$ provided we also make a spatial rotation with parameter $\phi(u,x,y)$ subject to

$$e^{2i\phi}\frac{\partial \zeta'}{\partial \zeta} = \text{real};$$

(iii) make a conformal change $\Omega' = \theta\Omega$ accompanied by a null rotation about $\hat{k}^a$ with

$$\theta = (1 + R\Omega)^{-1}, \quad c = -\Omega(1 + R\Omega)^{-1} P\bar{\nabla}R,$$

where $R(u,\hat{x}^i)$ is arbitrary.

The remaining metric equations

$$\begin{aligned}
&\hat{D}\hat{U} = \hat{\gamma} + \bar{\gamma},\\
&\hat{D}\hat{X}^i = 0,\\
&\hat{\delta}\hat{U} = \hat{\bar{\nu}},\\
&\hat{\mu} = \hat{\bar{\mu}},\\
&\hat{\delta}\hat{X}^i - \hat{\Delta}\hat{\xi}^i = \hat{\bar{\lambda}}\hat{\bar{\xi}}^i + (\hat{\mu} - \hat{\gamma} + \hat{\bar{\gamma}})\hat{\xi}^i,\\
&\hat{\bar{\delta}}\hat{\xi}^i - \hat{\delta}\hat{\bar{\xi}}^i = (\hat{\beta} - \hat{\bar{\alpha}})\hat{\bar{\xi}}^i + (\hat{\alpha} - \hat{\bar{\beta}})\hat{\xi}^i
\end{aligned} \tag{4.4}$$

together with the Ricci identities [Eqs. (4.2) of Ref. 2] and the transformation equation for the Ricci scalar [the last of Eqs. (1.5) of Ref. 7] can now be solved in a straightforward manner. It turns out that $\hat{\mu} - \Omega^{-1}\hat{U}$ is independent of $\Omega$ and can be made to vanish with the aid of freedom (iii), which now becomes further constrained by

$$\nabla\bar{\nabla}R = 0.$$

The results of this calculation are

$$\begin{aligned}
&\hat{\kappa} = \hat{\sigma} = \hat{\tau} = \hat{\epsilon} = \hat{\pi} = \hat{\omega} = \hat{\Psi}_0\\
&\quad = \hat{\Psi}_1 = \hat{\Psi}_2 = \hat{\Phi}_{00} = \hat{\Phi}_{01} = 0,\\
&\hat{\rho} = \Omega^{-1}, \quad \hat{\alpha} = -\hat{\bar{\beta}} = \Omega^{-1}\alpha^0,\\
&\hat{\lambda} = \lambda^0, \quad \hat{\gamma} = U^0\Omega - \bar{\gamma}^0,\\
&\hat{\mu} = -U^0\Omega + \gamma^0 + \bar{\gamma}^0 + \Omega^{-1}K^0,\\
&\hat{\nu} = \Omega^{-1}P\bar{\nabla}K^0 + P\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) - \Omega P\bar{\nabla}U^0,\\
&\hat{f} = -1, \quad \hat{\xi}^i = \Omega^{-1}\xi^{i0}, \quad \hat{X}^i = \hat{X}^{i0},\\
&\hat{U} = -U^0\Omega^2 + (\gamma^0 + \bar{\gamma}^0)\Omega + K^0,\\
&\hat{\Psi}_3 = \Omega^{-2}P\bar{\nabla}K^0 + \Omega^{-1}[\tfrac{1}{2}P\bar{\nabla}(\gamma^0 + \bar{\gamma}^0)\\
&\qquad - \tfrac{1}{2}P\nabla\lambda^0 + 2\lambda^0\bar{\alpha}^0],\\
&\hat{\Psi}_4^0\\
&= \Omega^{-2}\bar{\nabla}(P^2\bar{\nabla}K^0) + \Omega^{-1}[\bar{\nabla}(P^2\bar{\nabla}(\gamma^0 + \bar{\gamma}^0))] - 2\lambda^0K^0\\
&\quad - \bar{\nabla}(P^2\bar{\nabla}U^0) + \lambda^0(\bar{\gamma}^0 - 3\gamma^0) - \dot{\lambda}^0 + \tfrac{1}{2}X^0\nabla\lambda^0\\
&\quad + \tfrac{1}{2}\bar{X}^0\bar{\nabla}\lambda^0,\\
&\hat{\Lambda} = -\tfrac{1}{2}\Omega^{-1}(\gamma^0 + \bar{\gamma}^0) + U^0,\\
&\hat{\Phi}_{02} = -\Omega^{-1}\bar{\lambda}^0, \quad \hat{\Phi}_{11} = -\tfrac{1}{2}\Omega^{-1}(\gamma^0 + \bar{\gamma}^0),\\
&\hat{\Phi}_{12} = P\nabla U^0 + \Omega^{-1}[\tfrac{1}{2}P\bar{\nabla}\bar{\lambda}^0 - \tfrac{1}{2}P\nabla(\gamma^0 + \bar{\gamma}^0) - 2\alpha^0\bar{\lambda}^0],\\
&\hat{\Phi}_{22} = \Omega^{-2}P^2\nabla\bar{\nabla}K^0 + \Omega^{-1}[P^2\nabla\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) - \dot{K}^0\\
&\qquad + \tfrac{1}{2}X^0\nabla K^0\\
&\qquad + \tfrac{1}{2}\bar{X}^0\bar{\nabla}K^0]\\
&\qquad + [-P^2\nabla\bar{\nabla}U^0 - \dot{\gamma}^0 - \bar{\gamma}^0\\
&\qquad + \tfrac{1}{2}X^0\nabla(\gamma^0 + \bar{\gamma}^0)\\
&\qquad + \tfrac{1}{2}\bar{X}^0\bar{\nabla}(\gamma^0 + \bar{\gamma}^0) - \lambda^0\bar{\lambda}^0]\\
&\qquad + \Omega(\dot{U}^0 - \tfrac{1}{2}X^0\nabla U^0\\
&\qquad - \tfrac{1}{2}\bar{X}^0\bar{\nabla}U^0),
\end{aligned} \tag{4.5}$$

where the variables $\alpha^0, \xi^{i0}, K^0, \lambda^0, \gamma^0 - \bar{\gamma}^0, X^0$ are defined as in Eq. (2.8). The variables $P, X^0, \gamma^0 + \bar{\gamma}^0$ and $U^0$ again have to satisfy Eq. (2.9) but are otherwise left arbitrary.

The metric is easily obtained in standard fashion from the metric coefficients. The result is

$$\begin{aligned}
d\hat{s}^2 = {}&[2\hat{U} - \tfrac{1}{2}P^{-2}\Omega^2((X^{30})^2 + (X^{40})^2)]\,du^2 - 2\,du\,d\Omega\\
&+ P^{-2}X^{30}\Omega^2\,du\,dx^3 + P^{-2}X^{40}\Omega^2\,du\,dx^4\\
&- \tfrac{1}{2}P^{-2}\Omega^2[(dx^3)^2 + (dx^4)^2].
\end{aligned} \tag{4.6}$$

The meaning of the function $P$ now becomes clear. It is the metric of a $u$ = constant, $\Omega$ = constant 2-surface. If $K^0$ vanishes we can, as in Sec. 2, use the remaining freedom to set $P$ equal to a constant.

Let us now convert back to the physical space-time $M$. We first introduce coordinates

$$(x^a) = (u, r, x, y),$$

where $r$ is an affine parameter for the geodesics, determined as follows:

$$\frac{\partial \Omega}{\partial r} = D\Omega = \Omega^2\hat{D}\Omega = -\Omega^2;$$

hence $\Omega = r^{-1}$ by a suitable choice of origin.

The transformation formulas for the spin-coefficients and the tetrad components of the Ricci and Weyl tensors are given in Ref. 7 by, respectively, Eqs. (A5), (1.5), and (1.6). For the metric coefficients the required transformation laws are

$$\xi^i = \Omega\hat{\xi}^i, \quad X^i = \hat{X}^i, \quad \omega = \Omega^{-1}\hat{f}^{-1}\left(\hat{\omega} - \hat{\xi}^i\frac{\partial \Omega}{\partial x^i}\right),$$

$$U = \Omega^{-2} \hat{f}^{-1} \left( \hat{U} - \frac{\partial \Omega}{\partial u} - \hat{X}^i \frac{\partial \Omega}{\partial x^i} \right).$$

After some calculation we find the solution given by Eqs. (2.6)–(2.10) of Sec. 2. Since $\hat{g}_{ab} = \Omega^2 g_{ab}$ the metric (2.10) can be obtained more directly from Eq. (4.6).

## 5. CONCLUSION

To summarize, we have generalized a class of metrics first discovered by Kundt many years ago. These metrics describe a class of space-times containing geodesic rays whose divergence, shear, and curl all vanish. By assumption, the Ricci scalar vanishes, but the Ricci tensor is otherwise as arbitrary as the field equations allow. Since the Goldberg–Sachs theorem is no longer applicable we made the further simplifying assumption that the rays in question correspond to repeated principal null directions of the Weyl tensor. Therefore, our solutions are all algebraically special.

These solutions were derived in two ways, first by using the standard Newman–Penrose formalism, second by using a conformal approach. The advantages of the conformal method were discussed in Sec. 1. Both methods have the advantage that they yield not only the metric but also, at the same time, the components of the Weyl and Ricci tensors (relative to a specially constructed tetrad).


## ACKNOWLEDGMENTS

This work was supported in part by a grant from the Natural Sciences/Engineering Research Council of Canada. I am also indebted to Dr. H. Künzle, Dr. P. Sommers, and Dr. M. Walker for some useful comments.



[1] W. Kundt, Z. Physik **163**, 77 (1961).
[2] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[3] I. Robinson and A. Trautman, Proc. Roy. Soc. A **265**, 463 (1962).
[4] J. Foster and E. T. Newman, J. Math. Phys. **8**, 189 (1967).
[5] R. Penrose, Proc. Roy Soc. A **284**, 159 (1965).
[6] R. Penrose, in *Battelle Rencontres* (Benjamin, New York, 1968).
[7] G. Ludwig, Gen. Rel. Grav. **7**, 293 (1976).
[8] G. Ludwig, Gen. Rel. Grav. **9**, 1009 (1978).
[9] G. Ludwig, J. Math. Phys. **21**, 154 (1980).
[10] F. A. E. Pirani, in *Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, N.J., 1965).
[11] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).
[12] R. Penrose, in *Group Theory in Non-Linear Problems* (Reidel, Dordrecht, Holland, 1974).

# Lattice Green's functions for cubic lattices

M. A. Rashid
*Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria*

(Received 17 July 1979; accepted for publication 2 November 1979)

The Green's functions for a cubic lattice given by
$G(E) = 1/\pi^3 \int_0^\pi \int_0^\pi \int_0^\pi (dxdydz)/[E - \omega(x,y,z)]$,
where
(i) $\omega(x,y,z) = (a_1\cos x + a_2\cos y)(1 + \cos z) + a_3\cos z$,
(ii) $\omega(x,y,z) = a_1\cos x(1 + \cos y + \cos z + \cos y\cos z) + a_2\cos y + a_3\cos z + a_{23}\cos y\cos z$
are evaluated exactly and expressed as products of two ${}_2F_1$'s each of which represents a complete elliptic integral of the first kind. The expressions for the Green's functions manifest the expected symmetries.

## I. INTRODUCTION

On account of their application in several physical problems, many authors have considered the analytic evaluation of the Lattice Green's functions for different types of cubic lattices.[1-6] Hioe, in a recent paper,[7] has summarized the various cases that have been considered in the past. He has also evaluated the special Green's function

$$G(E;a_1,a_2,a_3) = \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dx\,dy\,dz}{E - \omega(x,y,z)}, \tag{1}$$

where

$$\omega(x,y,z) = (a_1\cos x + a_2\cos y)(1 + \cos z) + a_3\cos z \tag{2}$$

and

$$E \geqslant 2a_1 + 2a_2 + a_3.$$

Hioe has expressed his answer as a product of two complete elliptic integrals. However, the obvious invariance of the Green's function under the transformations involving change of sign of $a_1$ and/or $a_2$ and interchange of $a_1$ and $a_2$ is not manifest in this expression. In the present paper, we attempt this problem in a very simple and straightforward manner which immediately results in an expression for the Green's function in terms of an $F_4$ function of Appell. The parameters of this Appell function are such that it factorizes into two ${}_2F_1$'s each of which can also be written as a complete elliptic integral, though, the first factorization looks better. In addition, since our procedure retains the above-mentioned symmetries at every stage, our final answer manifestly exhibits the desired symmetries.

We have also given a simple analytic expression for the Lattice Green's function corresponding to

$$\omega(x,y,z) = a_1\cos x(1 + \cos y + \cos z + \cos y\cos z) + a_2\cos y + a_3\cos z + a_{23}\cos y\cos z, \tag{3}$$

which is a generalization of the case previously considered by Glasser[5] (His case is obtained by taking $a_1 = a_2 = a_3 = a_{23} = 1$).

## II. COMPUTATION OF THE LATTICE GREEN'S FUNCTION CONSIDERED BY HIOE

We perform the $z$-integration in Eq. (1) using[8]

$$\int_0^\pi \frac{dz}{a + b\cos z} = \frac{\pi}{(a^2 - b^2)^{1/2}}, \quad |a| > |b|, \tag{4}$$

which results in

$$G(E;a_1,a_2,a_3) = \frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{dx\,dy}{[(E + a_3)(E - a_3 - 2a_1\cos x - 2a_2\cos y)]^{1/2}}$$

$$= \frac{1}{(E^2 - a_3^2)^{1/2}} \sum_{n,m=0}^{\infty} \frac{(1/2)n + m}{n!m!}\left(\frac{2a_1}{E - a_3}\right)^n\left(\frac{2a_2}{E - a_3}\right)^m \tag{5}$$

$$\times \frac{1}{\pi^2}\int_0^\pi\int_0^\pi \cos^n x\cos^m y\,dx\,dy. \tag{6}$$

Since

$$\frac{1}{\pi}\int_0^\pi \cos^{2n+1}x\,dx = 0,$$

and

$$\frac{1}{\pi}\int_0^\pi \cos^{2n}x\,dx = \frac{(1/2)_n}{n!}, \tag{7}$$

where

 0022-2488/80/102549-04$1.00  

(i) $(\alpha)_n = \Gamma(n+\alpha)/\Gamma(n)$, (8)

and

(ii) $n$ is a nonnegative integer, Eq. (6) becomes

$$G(E;a_1,a_2,a_3) = \frac{1}{(E^2-a_3^2)^n} \sum_{n,m=0}^{\infty} \frac{(1/2)2n+2m(1/2)_n(1/2)_m}{(2n)!(2m)!n!m!} \left(\frac{2a_1}{E-a_3}\right)^{2n} \left(\frac{2a_2}{E-a_3}\right)^{2m}. \tag{9}$$

Applying the duplication formula[9]

$$\Gamma(2z) = 2^{2z-1} \frac{\Gamma(z)\Gamma[z+(1/2)]}{\pi^{1/2}}, \tag{10}$$

to the above equation, we arrive at[10]

$$G(E;a_1,a_2,a_3)$$

$$= \frac{1}{(E^2-a_3^2)^{1/2}} \sum_{n,m=0}^{\infty} \frac{(1/4)n+m(3/4)n+m}{n!n!m!m!} = \left(\frac{2a_1}{E-a_3}\right)^{2n} \left(\frac{2a_2}{E-a_3}\right)^{2m} \tag{11}$$

$$= \frac{1}{(E^2-a_3^2)^{1/2}} F_4\left(\tfrac{1}{4},\tfrac{3}{4};1,1;\left(\frac{2a_1}{E-a_3}\right)^2,\left(\frac{2a_2}{E-a_3}\right)^2\right). \tag{12}$$

Now[11]

$$F_4(a,\, b+c-a-1;\, b,c;\, u(1-v),\, v(1-u)) = {}_2F_1(a,b+c-a-1;\, c;u)\; {}_2F_1(a,\, b+c-a-1;\, c;\, v). \tag{13}$$

The use of the above result transforms Eq. (12) to

$$G(E;a_1,a_2,a_3) = \frac{1}{(E^2-a_3^2)^{1/2}} {}_2F_1(\tfrac{1}{4},\tfrac{3}{4},1,u)\, {}_2F_1(\tfrac{1}{4},\tfrac{3}{4};1,v), \tag{14}$$

where

$$u = \left[\frac{[(E-a_3+2a_1+2a_2)(E-a_3+2a_1-2a_2)]^{1/2} - [(E-a_3-2a_1+2a_2)(E-a_3-2a_1-2a_2)]^{1/2}}{2(E-a_3)}\right]^2,$$

$$v = \left[\frac{[(E-a_3+2a_1+2a_2)(E-a_3+2a_1-2a_2)]^{1/2} - [(E-a_3-2a_1+2a_2)(E-a_3-2a_1-2a_2)]^{1/2}}{2(E-a_3)}\right]^2. \tag{15}$$

Since

(i) $u$ and $v$ are invariant under change of sign of $a_1$ and/or $a_2$,

(ii) $a_1 \leftrightarrow a_2 \Rightarrow u \leftrightarrow v$, and

(iii) the two ${}_2F_1$'s in Eq. (14) above have the same parameters, the analytic expression for the Green's function $G(E;a_1,a_2,a_3)$ in Eq. (14) above obviously contains the symmetries mentioned in the introduction.

Noting that[12]

$${}_2F_1(\tfrac{1}{4},\tfrac{3}{4};1;z) = (1+\sqrt{z})^{-1/2}\, {}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;\frac{2\sqrt{z}}{1+\sqrt{z}}\right), \tag{16}$$

we observe that each of the two ${}_2F_1$'s appearing in Eq. (14) can be expressed as a complete elliptic integral, though its argument is not any simpler.

Our expression for the Lattice Green's function and its derivation may be compared with the ones in Hioe's work[7].

*Special cases*:

(i) $a_1 = a_2 = a_3 = 1$, then

$$u = v = \left(\frac{\sqrt{(E+3)} - \sqrt{(E-5)}}{2\sqrt{(E-1)}}\right)^2,$$

and

$$G(E;1,1,1) = \frac{1}{\sqrt{(E^2-1)}} \left[{}_2F_1\left(\tfrac{1}{4},\tfrac{3}{4};1;\left(\frac{\sqrt{(E+3)} - \sqrt{(E-5)}}{2\sqrt{(E-1)}}\right)^2\right)\right]^2. \tag{17}$$

In particular

$$G(5;1,1,1) = \frac{1}{2\sqrt{6}}\left[{}_2F_1\left(\tfrac{1}{4},\tfrac{3}{4};1;\tfrac{1}{2}\right)\right]^2$$
$$= \frac{1}{2\sqrt{6}}\left[\frac{\Gamma(1/2)}{\Gamma(5/8)\Gamma(7/8)}\right]^2$$
$$= \frac{1}{16\sqrt{6}\pi^3}\left[\Gamma(1/8)\Gamma(3/8)\right]^2. \quad \text{(ii)} \tag{18}$$

$$G(2a_1+2a_2+a_3;a_1,a_2,a_3) = \frac{1}{2[(a_1+a_2)(a_1+a_2+a_3)]^{1/2}}\,{}_2F_1\left(\tfrac{1}{4},\tfrac{3}{4};1;\frac{a_1}{a_1+a_2}\right){}_2F_1\left(\tfrac{1}{4},\tfrac{3}{4};1;\frac{a_2}{a_1+a_2}\right). \tag{19}$$

When $a_1 \to 0$, the first ${}_2F_1 \to 1$, whereas the second one diverges. Thus $G(2a_1+2a_2+a_3;a_1,a_2a_3) \to \infty$ as $a_1 \to 0$. The above result and the results in Eqs. (17–19) are the same as previously obtained by Hioe.

## III. LATTICE GREEN'S FUNCTION FOR GENERALIZATION OF A CASE CONSIDERED BY GLASSER

In this section, we wish to compute

$$G(E;a_1,a_2,a_3,a_{23}) = \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dx\,dy\,dz}{E-\omega(x,y,z)}, \tag{20}$$

where

$$\omega(x,y,z) = a_1\cos x(1+\cos y+\cos z+\cos y\cos z) + a_2\cos y + a_3\cos z + a_{23}\cos y\cos z, \tag{21}$$

and

$$E \geqslant 4a_1 + a_2 + a_3 + a_{23}. \tag{22}$$

Performing the $z$-integration, we obtain

$$G(E;a_1,a_2,a_3,a_{23}) = \frac{1}{\pi^2}\int_0^\pi\int_0^\pi dx\,dy\left[(E+a_3-(a_2-a_{23})\cos y) \times (E-a_3-2a_1\cos x-\cos y(a_2+a_{23}+a_1\cos x))\right]^{-1/2} \tag{23}$$

To do the $y$-integration, we take $\cos y$ as the new variable and use[13]

$$\int_u^c \frac{dx}{\sqrt{(a-x)(b-x)(c-x)(x-d)}} = \frac{2}{\sqrt{(a-c)(b-d)}}F(\gamma,k), \tag{24}$$

where $a>b>c>u\geqslant d$ and

$$\text{(i)}\ \gamma = \sin^{-1}\left[\frac{(b-d)(c-u)}{(c-d)(b-u)}\right]^{1/2}, \tag{25a}$$

$$\text{(ii)}\ k = \left[\frac{(a-b)(c-d)}{(a-c)(b-d)}\right]^{1/2}, \tag{25b}$$

and

(iii) $F(\gamma,k)$ is in general an *incomplete* elliptic integral. In particular, the complete elliptic integral $F(\pi/2,k)$ is expressed as[14]

$$F(\pi/2,k) = \tfrac{1}{2}\pi\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;k^2\right). \tag{26}$$

Since in our problem

$$a = \frac{E+a_3}{a_2-a_{23}}, \quad b = \frac{E-a_3-2a_1\cos x}{a_2+a_{23}+2a_1\cos x}, \tag{27}$$

and $c=-u=-d=1$, we find $\gamma=\pi/2$ and

$$G(E;a_1,a_2,a_3,a_{23}) = \frac{1}{\pi[(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})]^{1/2}} \times \int_0^\pi dx\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;k^2\right), \tag{28}$$

where

$$k^2 = A + B\cos x, \tag{29}$$

and

$$\text{(i)}\ A = \frac{4(Ea_{23}+a_2a_3)}{(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})}, \tag{30a}$$

$$\text{(ii)}\ B = \frac{4a_1(E+a_2+a_3-a_{23})}{(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})}. \tag{30b}$$

The integral in Eq. (28) can now be evaluated by integrating term by term the series expansion of the hypergeometric function, using contour integration techniques. This results in

$$G(E;a_1,a_2,a_3,a_{23}) = \frac{1}{[(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})]^{1/2}} \times \sum_{n,m=0}^{\infty}\frac{(\frac{1}{2})n+m(\frac{1}{2})n+m}{n!n!m!m!}(A_+)^n(A_-)^m, \tag{31}$$

where

$$A_\pm = \tfrac{1}{2}(A \pm (A^2-B^2)^{1/2}). \tag{32}$$

On using the definition of Appell's $F_4$ function[10] and Bailey's theorem,[11] we arrive at

$$G(E;a_1,a_2,a_3,a_{23}) = \frac{1}{[(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})]^{1/2}} \times F_4\left(\tfrac{1}{2},\tfrac{1}{2};1,1;A_+,A_-\right)$$
$$= \frac{1}{[(E-a_2+a_3+a_{23})(E+a_2-a_3+a_{23})]^{1/2}} \times {}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1,1;u\right){}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;v\right), \tag{33}$$

where

$$u(1-v) = A_+, \quad v(1-u) = A_-, \tag{34}$$

or

$$u = \tfrac{1}{2}\{1 + (A^2 - B^2)^{1/2} - [(1-A)^2 - B^2]^{1/2}\}, \quad (35a)$$

and

$$v = \tfrac{1}{2}\{1 - (A^2 - B^2)^{1/2} - [(1-A)]^2 - B^2\}^{1/2}. \quad (35b)$$

Note that the expression for the Green's function in Eq. (33) above obviously possesses symmetry under (i) change of sign of $a_1$, (ii) interchange of $a_2, a_3$.

*Special cases*:

(i) When $a_1 = a_2 = a_{23} = 1$, we find

$$A = \frac{4}{E+2-a_3}, \quad B = \frac{4}{E+2-a_3}, \quad A_{\pm} = \frac{2}{E+2-a_3},$$

$$u = v = \tfrac{1}{2}[1-(1-2A)^{1/2}] = \frac{(E+2-a_3)^{1/2} - (E-6-a_3)^{1/2}}{2(E+2-a_3)^{1/2)}},$$

and

$$G(E;1,1,a_3,1) = \frac{1}{[(E+a_3)(E+2-a_3)]^{1/2}} \times \left[{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;\frac{(E+2-a_3)^{1/2} - (E-6-a_3)^{1/2}}{2(E+2-a_3)^{1/2}}\right)\right]^2. \quad (36)$$

In particular

$$G(E;1,1,1,1) = \frac{1}{(E+1)}\left[{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;\frac{(E+1)^{1/2} - (E-7)^{1/2}}{2(E+1)^{1/2}}\right)\right]^2, \quad (37)$$

which is exactly the expression obtained by Glasser.[5]

(ii) When $E \to 4a_1 + a_2 + a_3 + a_{23}$,

$$A \to \frac{(4a_1 + a_2 + a_3 + a_{23})a_{23} + a_2 a_3}{(2a_1 + a_2 + a_{23})(2a_1 + a_3 + a_{23})},$$

$$B \to \frac{2a_1(2a_1 + a_2 + a_3)}{(2a_1 + a_2 + a_{23})(2a_1 + a_3 + a_{23})}.$$

Thus

$$A + B \to 1, \quad u \to \tfrac{1}{2}[1 + (1-2B)^{1/2}], \quad v \to \tfrac{1}{2}[1 - (1-2B)^{1/2}],$$

with

$$B = \frac{2a_1(2a_1 + a_2 + a_3)}{(2a_1 + a_2 + a_{23})(2a_1 + a_3 + a_{23})}.$$

In particular, when $a_1 = a_2 = a_{23} = 1$ [case (i) above], and $E \to 6 + a_3$, $A = B \to 1/2$, $u = v \to 1/2$, and

$$G(6 + a_3;1,1,a_3,1) \to \frac{1}{4(3+a_3)^{1/2}} \times \left|{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;\left(\tfrac{1}{2}\right)\right|^2 = \frac{\pi}{4(3+a_3)^{1/2}[\Gamma(3/4)]^4},$$

using Eq. (2.8(50)), p. 104, in Ref. (12).


[1]E.W. Montroll, Proc. 3rd Berkeley Symp. **3**, 209 (1956).
[2]A.A. Maradudin, E.W. Montroll, G.H. Weiss, R. Herman, and W.H. Miles, Acad. R. Belg., Bruxelles (1960).
[3]G. Iwata, Natural Science Report (Ochanomizu Univ.) **20**, 13 (1969), No. 2.
[4]G.S. Joyce, J. Phys. C **4**, L 53 (1971).
[5]M.L. Glasser, J. Math. Phys. **13**, 1145 (1972).
[6]G.S. Joyce, Trans. R. Soc. A **273**, 46 (1973).
[7]F.T. Hioe, J. Math. Phys. **19**, 1064 (1978).
[8]I.S. Gradshteyn and I.M. Rhysik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Eq. (3.665 (1)), p. 384.
[9]Ref. 8, Eq. (8.335 (1)), p. 936.
[10]For a definition of Appell's $F_4$ function, see Ref. 8, Eq. (9.180 (4)), p. 1053.
[11]Ref. 8, Eq. (9.182 (5)), p. 1055.
[12]Bateman Manuscript Project, *Higher Transcendental Functions*, Vol. I, edited by A. Erdelyi *et al.* (McGraw-Hill, New York, 1953), Eq. (2.11 (17)), p. 112, with $a = 1/4$, $c = 1$.
[13]Ref. 8, Eq. (3.147(2)), p. 242.
[14]Ref. 8, Eqs. (8.112(1), 8.113(1)), p. 905.

# Covariant objects and invariant equations on fiber bundles

Richard Kerner
*Département de Mécanique, Université Pierre et Marie Curie, 4, Place Jussieu, 75005, Paris, France*

(Received 13 April 1979; accepted for publication 16 October 1979)

Let $P(M_4,G)$ be a principal fiber bundle over the Minkowskian space–time $M_4$ with the structural group $G$. The group $G$ is supposed to be a compact and semisimple Lie group. Let $A$ be a connection form on $P(M_4,G)$ and $F = DA$ its curvature form. Let $g_G$ be the Cartan–Killing metric on $G$, and $g_{M_4}$ the Minkowskian metric on $M_4$. Let us define $d\pi : TP \to TM_4$, the differential of the canonical projection from $P$ onto $M_4$. Then we can define a scalar product for any two vectors from $P(M_4,G)$:

$$g_P(X,Y) = g_G(A(X),A(Y)) + g_{M_4}(d\pi(X),d\pi(Y)).$$

In this metric the horizontal and vertical subspaces of the connection $A$ are orthogonal to each other. Next, we construct the Clifford algebra corresponding to the metric $g_P$. The metric $g_P$ can be always diagonalized locally to give diag$((3+N)+,1-)$, where $N$ is the dimension of $G$. The lowest faithful representation of this algebra, which we call $C(3+N,1)$ is of the dimension $K = 2^{[(N+5)/2]}$. This $K$-dimensional vector space is called the space of spinors over $P(M_4,G)$. We study the decomposition of these spinors into multiplets of Lorentz spinors. We also define the generalized Dirac equation for such a spinor, construct an explicit representation in the case of $G = \mathrm{SU}(2)$, and give the formulae for the mass splitting. Finally, the invariant interaction with vector fields over $P(M_4,G)$ and scalar multiplets is discussed, together with the physical implications of the coupled equations.

## 1. INTRODUCTION AND NOTATIONS

Since the appearance of the first papers[1-3] introducing fiber bundles as the most appropriate framework for describing the Yang–Mills field theory, this mathematical language has been adopted by most physicists. The Yang–Mills field is interpreted as the curvature form on the principal fiber bundle over the space–time, the structural group being identified with the gauge group. The gauge potential is the connection form on this fiber bundle. Still, when it comes to define an interaction between the gauge field and a spin-$\frac{1}{2}$ field, or a scalar multiplet, the fiber bundle structure somehow gets out of view and the interaction is defined *ad hoc*, by analogy with electrodynamics, the Klein–Gordon equation, minimal coupling, and so on. In this paper we want to show that however these analogies could be justified, it seems to be more logical to push such an analogy to the end and work all the time through with covariant objects over the fiber bundle, and the equations invariant under the isomorphisms of the fiber bundle manifold itself.

Let us first recall the usual notation. Let $P(M_4,G)$ be a principal fiber bundle over the basis manifold $M_4$ (the Minkowskian space–time) and with the structural group $G$ which we suppose to be a compact, semisimple Lie group. We design the structure constants of the group $G$ by

$$C^a_{bc} = -C^a_{cb}, \quad \text{with } a,b,\cdots = 1,2,\cdots,N = \dim G.$$

Let us also define the invariant Cartan–Killing metric form on $G$,

$$\mathring{g}_{ab} = C^c_{ad}C^d_{cb}\,. \tag{1}$$

In the basis $M_4$ the metric is diagonal,

$$\mathring{g}_{ij} = \mathrm{diag}(+\ +\ +\ -). \tag{2}$$

Let the Greek indices $\alpha,\beta,\cdots$ run from 1 to $N+4$; symbolically we can write $\{\alpha\} = \{i \text{ or } a\}$. Remembering that $P(M_4,G)$ is locally trivial, which means that sufficiently small open sets of $P(M_4,G)$ are isomorphic to the Cartesian products of open sets of $M_4$ and $G$, by choosing local maps in $P(M_4,G)$, in $M_4$ and in $G$ we can describe any point in $P(M_4,G)$ as follows: Let $p \in P(M_4,G)$; then

$$p \underset{\text{local map}}{\longrightarrow} \{p^\alpha\} \underset{\substack{\text{isomorphism or}\\ \circ \text{ local maps}}}{\longrightarrow} \{x^i,\xi^a\}, \tag{3}$$

where $\{x^i\}$ are the local coordinates of the point in $M_4$ onto which $p$ projects; this point fixes a whole fiber $\pi^{-1}(x)$ in $P(M_4,G)$. This fiber is isomorphic to $G$; $\{\xi^a\}$ gives a point in $G$ in some local map. The projection $\pi$ acts on $p$ in an obvious way:

$$\pi(p) = x \quad \text{if } p = (x,\xi)$$

or in local coordinates

$$\pi(\{p^\alpha\}) = \{x^i\} \quad \text{if } \{p^\alpha\} = \{x^i,\xi^a\}. \tag{4}$$

A tangent vector on $P(M_4,G)$ can be written as

$$X = X^\alpha\partial_\alpha = X^i\partial_i + X^a\partial_a \tag{5}$$

with $X^\alpha$ being smooth functions of $p$, i.e., of both $x^i$ and $\xi^a$. Similarly, any 1-form on $P(M_4,G)$ can be written as

$$\omega = \omega_\alpha\, dp^\alpha = \omega_i\, dx^i + \omega_a\, d\xi^a \tag{6}$$

with $\omega_\alpha$ being smooth functions of $p$. The differential of the projection operator, $d\pi$, acts locally as follows:

$$T_P \ni X, \quad X^\alpha \to d\pi^i_\alpha X^\alpha = \{d\pi(X)^i\}, \quad d\pi(X) \in T_{M_4}, \tag{7}$$

where $T_P$ means the tangent space to $P(M_4,G)$, and $T_{M_4}$ is

 0022-2488/80/102553-07$1.00  

the tangent space to $M_4$. The tangent space to $G$ is the Lie algebra of group $G$ and will be denoted by $\mathscr{A}_G$.

A homomorphism of $\mathscr{A}_G$ into $T_P$ can also be defined. It acts as follows: If $B$ is a vector in $\mathscr{A}_G$, then

$$B^a \to \sigma^\alpha_a B^a = [\sigma(B)]^\alpha, \quad \sigma(B) \in T_P \tag{8}$$

is a vertical vector in $T_P$. The vertical vector fields on $P(M_4,G)$ are defined by the action of $G$ on $P(M_4,G)$, they are just tangent to the orbits of $G$. The connection $A$ on $P(M_4,G)$ is a left-invariant Lie-algebra-valued 1-form on the fiber bundle. In local coordinates it defines a mapping from $T_P$ into $\mathscr{A}_G$,

$$T_P \ni X, \quad X^\alpha \to A^a_\alpha X^\alpha = [A(X)]^a, \quad A(X) \in \mathscr{A}_G . \tag{9}$$

Any vector $X$ for which $A(X) = 0$ will be called a *horizontal vector*. At any point of $P(M_4,G)$ the 1-form $A$ defines a horizontal supspace $T_P^{\text{hor}} \subset T_P$. Therefore we can define a unique lift from $T_{M_4}$ into $T_P$,

$$T_{M_4} \ni V, \quad V^i \to \tau^\alpha_i V^i = [\tau(V)]^\alpha, \quad \tau(V) \in T_p . \tag{10}$$

The obvious relations and definitions are satisfied then:

$$\begin{aligned} &X \in T_P : (\tau \circ d\pi)(X) = \text{hor}\, X, \quad \text{the horizontal part of } X, \\ &V \in T_{M_4} : (d\pi \circ \tau)(V) = V, \quad \text{i.e., } d\pi \circ \tau = \text{Id}_{T_{M_4}}, \\ &X \in T_P : (\sigma \circ A)(X) = \text{ver}\, X, \quad \text{the vertical part of } X, \\ &B \in \mathscr{A}_G : (A \circ \sigma)(B) = B, \quad \text{i.e., } A \circ \sigma = \text{Id}_{\mathscr{A}_G}. \end{aligned} \tag{11}$$

Moreover,

$$\tau \circ d\pi + \sigma \circ A = \text{Id}_{T_p} \tag{12}$$

because obviously

$$X = \text{hor}\, X + \text{ver}\, X. \tag{13}$$

Because $P(M_4,G)$ is locally trivial, we can always choose the local coordinates in which, with no loss in generality, we shall have

$$\begin{aligned} &\sigma^a_b = \delta^a_b, \quad \sigma^i_b = 0, \\ &A^q_b = \delta^a_b, \quad A^a_i = A^a_i(x), \\ &d\pi^i_a = 0, \quad d\pi^i_j = \delta^i_j , \end{aligned} \tag{14}$$

and, because of $\text{hor}\, X = X - \text{ver}\, X$,

$$\tau^a_i = -A^a_i, \quad \tau^i_j = \delta^i_j . \tag{15}$$

We can finally introduce a unique metric on $P(M_4,G)$ induced by the metrics on $M_4$ and $G$: Namely, we define for any $X, Y \in T_p$ a scalar product

$$g(X,Y) = g_G(A(X),A(Y)) + g_{M_4}(d\pi(X),d\pi(Y)). \tag{16}$$

In other words, the scalar product of any two vectors tangent to $T_p$ is the sum of the Lie algebra scalar product of their vertical parts and of the Minkowskian scalar product of their horizontal parts. By definition

$$g(\text{hor}\, X, \text{ver}\, Y) = g(\text{ver}\, X, \text{hor}\, Y) = 0 \tag{17}$$

so that

$$g(X,Y) = g(\text{ver}\, X, \text{ver}\, Y) + g(\text{hor}\, X, \text{hor}\, Y). \tag{18}$$

The horizontal and vertical subspaces of $T_P$ are orthogonal to each other. Keeping in mind (14) and (15) we see that in local coordinates the unique form of our metric is

$$g_{\alpha\beta} = \left(\begin{array}{c|c} \mathring{g}_{ij} + \mathring{g}_{ab} A^a_i A^b_j & \mathring{g}_{ab} A^a_i \\ \hline \mathring{g}_{ab} A^b_j & \mathring{g}_{ab} \end{array}\right), \tag{19}$$

$$g^{\alpha\beta} = \left(\begin{array}{c|c} \mathring{g}^{ij} & -\mathring{g}^{ij} A^b_j \\ \hline -\mathring{g}^{ij} A^a_i & \mathring{g}^{ab} + \mathring{g}^{ij} A^a_i A^b_j \end{array}\right). \tag{20}$$

Now we have at our disposal the basic notions which will serve to develop the geometry of our manifold.

## 2. GEOMETRICAL OBJECTS ON A FIBER BUNDLE

The fiber bundle $P(M_4,G)$ endowed with the connection form $A^a_\alpha$ and the metric tensor (19)–(20) is now a Riemannian manifold. We define the Christoffel symbols as usual,

$$\left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} = \tfrac{1}{2} g^{\alpha\delta}(\partial_\beta g_{\gamma\delta} + \partial_\gamma g_{\beta\delta} - \partial_\delta g_{\beta\gamma}) . \tag{21}$$

In order to compute them explicitly we have to define $\partial_a A^b_j$. We recall that $A^b_j$ is the 1-form of type *ad*, i.e., the group elements act on $A$ by left translation by means of the adjoint representation. This means in turn that

$$\partial_a A^b_j + C^b_{ac} A^c_j = 0, \tag{22}$$

which can be also interpreted as the vanishing of the covariant derivative of $A$ in the group space—here $C^a_{bc}$ plays the role of the affine connection coefficients. Being antisymmetric, this connection is just pure *torsion tensor* on the group.

Having (22) in mind, we easily compute the Christoffel symbols

$$\begin{aligned} &\left\{ \begin{matrix} a \\ bc \end{matrix} \right\} = 0, \quad \left\{ \begin{matrix} k \\ bc \end{matrix} \right\} = 0, \quad \left\{ \begin{matrix} k \\ bj \end{matrix} \right\} = \left\{ \begin{matrix} k \\ jb \end{matrix} \right\} = \mathring{g}^{kl} \mathring{g}_{bc} F^c_{jl}, \\ &\left\{ \begin{matrix} a \\ bj \end{matrix} \right\} = \left\{ \begin{matrix} a \\ jb \end{matrix} \right\} = \mathring{g}^{kl} \mathring{g}_{bc} A^a_k F^c_{lj} + C^a_{cb} A^c_j, \\ &\left\{ \begin{matrix} a \\ ij \end{matrix} \right\} = \left\{ \begin{matrix} a \\ ji \end{matrix} \right\} = \tfrac{1}{2}(\partial_i A^a_j + \partial_j A^a_i) \\ &\qquad\qquad + \mathring{g}^{kl} \mathring{g}_{bc} A^a_k (A^b_i f^c_{lj} + A^b_j f^c_{li}), \\ &\left\{ \begin{matrix} k \\ ij \end{matrix} \right\} = \left\{ \begin{matrix} k \\ ji \end{matrix} \right\} = \mathring{g}^{kl} \mathring{g}_{bc} (A^b_i f^c_{jl} + A^b_j f^c_{il}). \end{aligned} \tag{23}$$

Here

$$F^a_{ij} = \tfrac{1}{2}(\partial_i A^a_j - \partial_j A^a_i) + C^a_{bc} A^b_i A^c_j \tag{24}$$

and

$$f^a_{ij} = \tfrac{1}{2}(\partial_i A^a_j - \partial_j A^a_i). \tag{25}$$

$F^a_{ij}$ is the curvature 2-form of the connection form $A$, identified with the Yang–Mills field tensor. Still, the connection coefficients (23) are not fully gauge invariant. The reason is that in the fibers—and therefore in the vertical subspaces—we had the nonvanishing torsion tensor

$$\mathring{S}^a_{bc} = C^a_{bc}. \tag{26}$$

This tensor should be also imbedded into $T_p$ to give

$$S^\alpha_{\beta\gamma} = \mathring{S}^a_{bc} A^b_\beta A^c_\gamma \sigma^\alpha_a . \tag{27}$$

If the torsion tensor in the basis space were non null, we should also add the term

$$S^i_{jk}\, d\pi^j_\beta\, d\pi^k_\gamma \tau^\alpha_i . \tag{28}$$

From now on we put $S^i_{jk} = 0$. The nonvanishing coefficients of (27) are therefore

$$S^a_{bc} = C^a_{bc}, \quad S^a_{bj} = -S^a_{jb} = C^a_{bc}A^c_j,$$
$$S^a_{ij} = -S^a_{ij} = C^a_{bc}A^b_i A^c_j . \tag{29}$$

The fully covariant connection coefficients are therefore

$$\Gamma^\alpha_{\beta\gamma} = \begin{Bmatrix} \alpha \\ \beta\gamma \end{Bmatrix} + S^\alpha_{\beta\gamma} . \tag{30}$$

The curvature can now be computed, as well as the Ricci tensor and the scalar curvature. The scalar curvature of the connection (30) is equal (up to a constant) to

$$R = -\tfrac{1}{4}\mathring{g}^{ik}\mathring{g}^{jl}\mathring{g}_{ab}F^a_{ij}F^b_{kl} \tag{31}$$

which is exactly the gauge-invariant Lagrangian of the Yang–Mills field. The invariant principle analogous to Einstein's one can be generalized onto the $P(M_4,G)$,

$$\delta\int_P \sqrt{-g}R\, d^{N+4}p = 0 \tag{32}$$

and it will reduce itself to the usual variational principle on $M_4$ because $g = \det(g_{\alpha\beta}) = (\det\mathring{g}_{ij})(\det\mathring{g}_{ab})$ so that (32) is equivalent to

$$V\delta\int\sqrt{-g}R\, d^4x = 0, \tag{33}$$

$V$ being the volume of $G$.

Now we want to widen the theory in order to include spinors. We propose to construct the spinors right on the fiber bundle $P(M_4,G)$ and then try to reduce them into the Lorentz spinors (see, e.g., Refs. 4 and 5).

First let us construct the Clifford algebra corresponding to the metric (19). It is enough to define the generators corresponding to the locally diagonalized metric,

$$\mathring{g}_{\alpha\beta} = \left(\begin{array}{c|c} \mathring{g}_{ij} & 0 \\ \hline 0 & \mathring{g}_{ab} \end{array}\right). \tag{34}$$

These generators have to satisfy the following autocommutation relations:

$$\mathring{\gamma}_\alpha\mathring{\gamma}_\beta + \mathring{\gamma}_\beta\mathring{\gamma}_\alpha = 2\mathring{g}_{ab} , \tag{35}$$

or, more explicitly

$$\begin{aligned} &\mathring{\gamma}_a\mathring{\gamma}_b + \mathring{\gamma}_b\mathring{\gamma}_a = 2\mathring{g}_{ab} , \\ &\mathring{\gamma}_i\mathring{\gamma}_j + \mathring{\gamma}_j\mathring{\gamma}_i = 2\mathring{g}_{ij} , \\ &\mathring{\gamma}_a\mathring{\gamma}_j + \mathring{\gamma}_j\mathring{\gamma}_a = 0. \end{aligned} \tag{36}$$

Defining

$$\mathring{\gamma}^\alpha = \mathring{\gamma}_\beta\mathring{g}^{\alpha\beta} \tag{37}$$

we immediately get

$$\mathring{\gamma}^\alpha\mathring{\gamma}^\beta + \mathring{\gamma}^\beta\mathring{\gamma}^\alpha = 2\mathring{g}^{\alpha\beta}. \tag{38}$$

In order to now construct the Clifford algebra corresponding to the metric (21) it is enough to define

$$\gamma_\alpha = \begin{cases} \gamma_i = \mathring{\gamma}_i + \mathring{\gamma}_a A^a_i , \\ \gamma_a = \mathring{\gamma}_a , \end{cases} \tag{39}$$

$$\gamma^\beta = \begin{cases} \gamma^j = \mathring{\gamma}^j, \\ \gamma^a = \mathring{\gamma}^a - A^a_j\mathring{\gamma}^j. \end{cases} \tag{40}$$

Then obviously

$$\gamma_\alpha\gamma_\beta + \gamma_\beta\gamma_\alpha = 2g_{\alpha\beta} , \quad \gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2g^{\alpha\beta}. \tag{41}$$

The lowest-dimensional faithful representation of this Clifford algebra is given by the

$$2^{[(n+1)/2]}\times 2^{[(n+1)/2]} \text{ matrices.} \tag{42}$$

Here $n = N+4$, and $[m]$ means the integer part of $m$. The representation space of this algebra, which is a complex $2^{[(n+1)/2]}$-dimensional space, is the space of spinors on $P(M_4,G)$.

## 3. INVARIANT EQUATIONS FOR SPINORS ON THE BUNDLE

The fully invariant equation for the thus defined spinor defined on our fiber bundle with connection is of course the generalized Dirac equation

$$(\gamma^\alpha\nabla_\alpha + m)\psi = 0. \tag{43}$$

(Note that if $\gamma^\alpha$ are real, in order to make this operator self-adjoint we have to take $i\gamma^\alpha\nabla_\alpha$.) The generalized Klein–Gordon equation is then obtained by

$$(\gamma^\alpha\nabla_\alpha - m)(\gamma^\beta\nabla_\beta + m)\psi = 0 \tag{44}$$

giving

$$(\gamma^\alpha\gamma^\beta\nabla_\alpha\nabla_\beta - m^2)\psi = 0. \tag{45}$$

$\nabla_\alpha$ means the covariant derivative with respect to the Christoffel connection (23) or the full connection (30), if specified. The equations (43)–(45) are not well defined unless we define the action of $\nabla_\alpha$ on a spinor. We propose the following choice,

$$\nabla_\alpha\psi = \partial_\alpha\psi + \Gamma^\delta_{\alpha\beta}\gamma^\beta\gamma_\delta\psi. \tag{46}$$

The Eq. (45) can be written as

$$[g^{\alpha\beta}\nabla_\alpha\nabla_\beta + i\sigma^{\alpha\beta}R_{\alpha\beta}{}^\gamma{}_\delta\gamma^\delta\gamma_\gamma - m^2]\psi = 0, \tag{47}$$

where

$$\sigma^{\alpha\beta} = \frac{1}{2i}(\gamma^\alpha\gamma^\beta - \gamma^\beta\gamma^\alpha), \tag{48}$$

and $R_{\alpha\beta}\gamma^\delta$ is the Riemann tensor. An analogous generalization of Dirac's equation has been proposed by Drechsler (see Refs. 6 and 7). However in Drechsler's papers the fiber is isomorphic to a homogeneous space with constant curvature, with the de Sitter symmetry group operating on it. The noncompactness of the group is a cause of negative energy, which must be ruled out somehow. In our approach we discuss only the fibers which are isomorphic to some *compact* groups; moreover, we think that the generalized spinors have to be defined on a *principal fiber bundle*, and not on an *associate fiber bundle*, which would be equivalent to the choice of some special representation. In other words, we think that the *gauge group* is more fundamental than the representations we choose afterwards.

For a scalar field on $P(M_4,G)$ the Klein–Gordon equation (47) reduces to

$$[g^{\alpha\beta}\nabla_\alpha\nabla_\beta - m^2]\varphi = 0 \tag{49}$$

which, when calculated, gives just

$$(\mathring{g}^{ij}\partial_i\partial_j - m^2)\varphi = (\Box - m^2)\varphi = 0 \tag{50}$$

if we assume $\partial_a\varphi = 0$, which seems to be natural.

The analog of the Dirac equation (43), when made ex-

plicit, gives different results depending on whether we take the covariant derivative with respect to the Christoffel connection (23) or the full connection $\Gamma^{\alpha}_{\beta\gamma}$ including the torsion tensor (30).

In the first case (Christoffel connection) we get

$$[\mathring{\gamma}^{i}\partial_i + \mathring{\gamma}^{a}\partial_a - \mathring{\gamma}^{i}A_i^{\ a}\partial_a + \mathring{\gamma}^{i}A_i^{\ a}\mathring{g}^{cd}C_{ca}^{\ b}\mathring{\sigma}_{db} - m]\psi = 0. \tag{51}$$

Here

$$\mathring{\sigma}_{db} = (1/2i)(\mathring{\gamma}_d\mathring{\gamma}_b - \mathring{\gamma}_b\mathring{\gamma}_d). \tag{52}$$

The second case (Christoffel symbols + torsion tensor) gives

$$[\mathring{\gamma}^{i}\partial_i + \mathring{\gamma}^{a}\partial_a - \mathring{\gamma}^{i}A_i^{\ a}\partial_a + \mathring{\gamma}^{i}A_i^{\ a}\mathring{g}^{cd}C_{ca}^{\ b}\mathring{\sigma}_{db} + \mathring{\gamma}^{a}C_{acb}\mathring{\sigma}^{cb} - m]\psi = 0 \tag{53}$$

(we put $C_{abc} = \mathring{g}_{ad}C_{bc}^{\ d}$).

The final result depends on how we define the derivative of our spinor $\psi$ along the fiber direction, $\partial_a\psi$. We face two more or less natural choices: Either

$$\partial_a\psi = 0 \tag{54}$$

or

$$\partial_a\psi + C_{acb}\mathring{\sigma}^{cb}\psi = 0 \tag{55}$$

(the last choice being equivalent to $\nabla_a\psi = 0$). Combining (54)–(55) with (51) and (53) gives four different equations if $\partial_a\psi = 0$, then (51) reduces to

$$[\mathring{\gamma}^{i}(\partial_i - A_i^{\ a}\mathring{g}^{cd}C_{ac}^{\ b}\sigma_{db}) - m]\psi = 0 \tag{56}$$

and (53) gives

$$[\mathring{\gamma}^{i}(\partial_i - A_i^{\ a}\mathring{g}^{cd}C_{ac}^{\ b}\mathring{\sigma}_{db}) + \mathring{\gamma}^{a}C_{abc}\mathring{\sigma}^{bc} - m]\psi = 0, \tag{57}$$

whereas if $\partial_a\psi + C_{abc}\mathring{\sigma}^{bc}\psi = 0$, we get

$$[\mathring{\gamma}^{i}\partial_i - \mathring{\gamma}^{a}C_{abc}\mathring{\sigma}^{bc} - m]\psi = 0 \tag{58}$$

instead of (51), and

$$[\mathring{\gamma}^{i}\partial_i - m]\psi = 0 \tag{59}$$

instead of (53).

Thus the different choices of the geometric behavior of $\psi$ acquire a clear physical meaning. For example, (59) is just a free equation: All the interaction has been factorized out by covariance. (58) gives the mass splitting (breaking of the group symmetry), but no interaction with the gauge field. Equation (56) gives the invariant interaction with the gauge field without any mass-splitting term. In order to make Eqs. (56)–(59) "operative" we have to find out the eigenvalues of $\mathring{\sigma}_{ab}$, which will depend on the choice of representation.

## 4. THE CHOICE OF REPRESENTATION OF THE CLIFFORD ALGEBRA

Let us remember that the spinors on the fiber bundle defined above cannot have any direct physical meaning, because it is not an irreducible representation of the Lorentz group. We can hope, however, that in some cases our Clifford algebra can be decomposed into a product (or even a direct sum) of some lower-dimensional Clifford algebras, and the spinor can be decomposed into several Lorentz spinors. Let us first give a particular example of such a situation. Take the case when $G = \mathrm{SU}(2)$. The dimension of SU(2) is equal to 3; the Clifford algebra of the SU(2) group space endowed with the Cartan–Killing metric is given by three Pauli matrices $\sigma_a$ $(a = 1,2,3)$,

$$\sigma_a\sigma_b + \sigma_b\sigma_a = 2\mathring{g}_{ab}\,\mathrm{Id}_{(2\times 2)}\,. \tag{60}$$

$\mathrm{Id}_{(n\times n)}$ means the unit $(n\times n)$ matrix, $\mathring{g}_{ab} = \delta_{ab}$. The Clifford algebra of $M_4$ with Minkowskian metric is given by the usual Dirac matrices

$$\mathring{\gamma}_i\mathring{\gamma}_j + \mathring{\gamma}_j\mathring{\gamma}_i = 2\mathring{g}_{ij}\,\mathrm{Id}_{(4\times 4)}\,. \tag{61}$$

So, from now on

$$\mathring{\gamma}_a = \{\mathring{\gamma}_i, \mathring{\gamma}_a\} = \{\mathring{\gamma}_i, \sigma_a\}\,. \tag{62}$$

We want our representation to have the dimension

$$2^{[(n+1)/2]} = 2^{[(N+4+1)/2]} = 2^{[(3+4+1)/2]} = 2^4 = 16 \tag{63}$$

in which

$$\begin{aligned}
&\mathring{\gamma}_i\mathring{\gamma}_j + \mathring{\gamma}_j\mathring{\gamma}_i = 2\mathring{g}_{ij}\,\mathrm{Id}_{(16\times 16)}\,,\\
&\mathring{\gamma}_a\mathring{\gamma}_b + \mathring{\gamma}_b\mathring{\gamma}_a = 2\mathring{g}_{ab}\,\mathrm{Id}_{(16\times 16)}\,,\\
&\mathring{\gamma}_a\mathring{\gamma}_j + \mathring{\gamma}_j\mathring{\gamma}_a = 0.
\end{aligned} \tag{64}$$

Such a representation is easy to obtain in terms of $(4\times 4)$ blocks. Our building blocks will be either $(4\times 4)$ Dirac matrices or $(4\times 4)$ identity matrices.

The $16\times 16$ representation satisfying (64) is then given by the following matrices:

$$\mathring{\gamma}^{i}_{(16\times 16)} = \left(\begin{array}{cc|cc} & & \mathring{\gamma}^{i} & 0 \\ & 0 & & \\ \hline & & & \mathring{\gamma}^{i} \\ \mathring{\gamma}^{i} & 0 & 0 & \\ & & & \\ 0 & \mathring{\gamma}^{i} & & \end{array}\right) \tag{65}$$

and

$$\mathring{\gamma}^{a}_{(16\times 16)} = \left(\begin{array}{c|c} \sigma^{a} & 0 \\ \hline 0 & -\sigma^{a} \end{array}\right). \tag{66}$$

Here $\sigma_a$ mean the $2\times 2$ Pauli matrices in which the element-numbers are replaced by the $4\times 4$ identity matrices. We then easily get

$$C_{abc}\mathring{\gamma}^{b}\mathring{\gamma}^{c} = \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_a \end{pmatrix} \tag{67}$$

and

$$\mathring{\gamma}^{a}C_{abc}\mathring{\gamma}^{b}\mathring{\gamma}^{c} = \begin{pmatrix} \vec{\sigma}^2 & 0 \\ 0 & -\vec{\sigma}^2 \end{pmatrix}, \tag{68}$$

where

$$\vec{\sigma}^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2\,. \tag{69}$$

We see, therefore, that our 16-dimensional column (spinor) $\psi$ can be regarded upon as two couples of Dirac spinors:

$$\psi = \begin{pmatrix} p_1 \\ n_1 \\ p_2 \\ n_2 \end{pmatrix} \tag{70}$$

on which the Lorentz transformation acts in a reducible way, i.e., the same Lorentz transformation applies to all the

four Dirac spinors $p_1$, $n_1$, $p_2$, and $n_2$. That is so because the generators of the Lorentz transformations,

$$\mathring{\sigma}_{ij} = (1/2i)(\mathring{\gamma}_i\mathring{\gamma}_j - \mathring{\gamma}_j\mathring{\gamma}_i), \tag{71}$$

are block-diagonal, as it is easy to see. The group action is contragradient on $p_1$, $n_1$, and $p_2$, $n_2$—but the couples never mix up under the group action.

Due to the term $\mathring{\gamma}^b\mathring{\gamma}^c$ there is a mass-shifting, because the Dirac equation takes on the following form:

$$\left[\mathring{\gamma}^i(\partial_i - A_i^a C_{abc}\mathring{\sigma}^{bc}) + m + \begin{pmatrix} \vec{\sigma}^2 & 0 \\ 0 & -\vec{\sigma}^2 \end{pmatrix}\right]\psi. \tag{72}$$

Even in the absence of the Yang–Mills potential the masses will be shifted: If we act on (71) from the left with the same operator in which $m$ is replaced by $-m$, we get

$$[\Box + m^2 + (\vec{\sigma}^2)^2]\psi = 0. \tag{73}$$

In this formula the operator $(\vec{\sigma}^2)^2$ is diagonal, and its eigenvalues depend on the representation. In our case it is the lowest-dimensional one, and we can put $\vec{\sigma}^2\psi = s(s+1)\psi$, with $s = \frac{1}{2}$. In general, when the structural group is other than SU(2), the eigenvalues of the operator $C_{abc}\gamma^a\gamma^b\gamma^c$ can be quite difficult to compute. Anyway, we cannot compare this contribution to the mass term with any masses of known physical objects, because the fiber-bundle spinor $\psi$ is not an irreducible representation of the Lorentz group. The mass-term correction can be rather thought of as a mean value term for all the irreducible Lorentz components contained in the fiber-bundle spinor $\psi$. In order to obtain some meaningful mass terms in the generalized Dirac equation (57) we have to perform the following program: (a) to factorize out the group dependence, (b) to decompose the fiber-bundle spinor into a superposition of the Lorentz spinors, (c) to integrate the Lagrangian $\bar{\psi}(\gamma^\alpha\overleftrightarrow{\nabla}_\alpha + m)\psi$ over the group space.

By (a) we mean the following. Locally, any point in the fiber bundle $P(M_4,G)$ can be represented by the coordinates $(x,g)$, where $x$ belongs to $M_4$ and $g$ symbolizes the coordinates of the group element $g$, so that we can explicitly write $\psi$ as $\psi^A(x,g)$, with $A = 1,2,\ldots,2^{[(N+5)/2]}$. Having chosen the action of the Lie algebra generators on $\psi$ we can in principle perform a finite left translation by $g^{-1}$ and write $\psi(x,g)$ in terms of the components of $\psi$ at the point $(x,e)$: $\psi^A(x,g) = D^A{}_B(g)\psi^B(x)$, where we write for simplicity $\psi(x,e) = \psi(x)$. $D^A{}_B(g)$ is a reducible representation of $G$ in the Spin $(N+3,1)$, and as such can be decomposed into the sum of irreducible representations of $G$.

By (b) we mean the decomposition of $\psi^A(x)$ into the ordinary Lorentz spinors. This can be done, because we know how the group SO(3,1) acts on $\psi^A(x)$ explicitly. This will give us some decomposition in terms of the 6-$j$ symbols.

Finally, by (c) we mean integrating with the invariant Haar measure over the group space, the Lagrangian density corresponding to the generalized Dirac equation (43), i.e., $\bar{\psi}(\gamma^\alpha\overleftrightarrow{\nabla}_\alpha + m)\psi$.

Such a program has been performed by Domokos and Kövesi-Domokos for the SU(2) group, in their paper.[5] In the case of the SU(3) group or some higher-dimensional Lie group, this program becomes a formidable task. It seems to be worth trying, because it could provide us with a rich mass spectrum of the fermions. At this stage it is too early to formulate any guesses as to its fitting with the experiment. All we can say is that in our theory there are no massless fermions in the absence of an Abelian subgroup, as it clearly follows from the term, $\mathring{\gamma}^a C_{abc}\mathring{\sigma}^{bc}$.

The last remark concerning the masses is that in order to compare them with the physical data our structure constants have to be rescaled, i.e., they have to contain the coupling parameter $\lambda$: $C^c_{ab} \to \lambda C^c_{ab}$. The nature of this coupling parameter depends on the kind of interactions we want to describe by the gauge field; e.g., for the Abelian (electromagnetic) case $\lambda$ is the electric charge $e$.

Now, the general situation can be described as follows.

Let $C(p,q)$ be the Clifford algebra corresponding to the metric diag$(p^+,q^-)$. We have the following reduction formulae[8]:

$$\begin{aligned} &C(p+1,q+1) = C(p,q) \otimes C(1,1), \\ &C(q+2,p) = C(p,q) \otimes C(2,0), \\ &C(q,p+2) = C(p,q) \otimes C(0,2). \end{aligned} \tag{74}$$

Let us also remember that

$$\begin{aligned} &C(2,0) = C(1,1) = \mathrm{Mat}_2(R) = 2\times 2 \text{ real matrices}, \\ &C(2,0) = \mathbb{H} = \text{quaternions}, \\ &C(0,1) = \mathbb{C} = \text{complex numbers}, \\ &C(1,0) = \mathbb{R} + \mathbb{R} = \text{sum of two real lines}. \end{aligned}$$

The splitting of the representation into a sum of lower representations can occur if the $C(1,0)$ appears in the decomposition of $C(p,q)$. For example, in the case when $G = \mathrm{SU}(2)$, dim SU(2) = 3, the metric $\mathring{g}_{\alpha\beta}$ becomes diag$(6+,1-)$ and we get

$$\begin{aligned} C(6,1) &= C(1,1)\otimes C(5,0) = C(1,1)\otimes C(2,0)\otimes C(0,3) \\ &= C(3,1)\otimes C(0,3) = C(3,1)\otimes C(0,2)\otimes C(1,0) \\ &= C(3,1)\otimes \mathbb{H}\otimes[\mathbb{R}\oplus\mathbb{R}] \\ &= [C(3,1)\otimes\mathbb{H}]\oplus[C(3,1)\otimes\mathbb{H}]. \end{aligned} \tag{75}$$

The decomposition

$$C(6,1) = [C(3,1)\otimes\mathbb{H}]\oplus[C(3,1)\otimes\mathbb{H}] \tag{76}$$

describes exactly the representation of $\mathring{\gamma}_\alpha$'s we have chosen in (65) and (66)—$C(3,1)$ is the space–time Clifford algebra, and the quaternions are represented by the $\sigma_a$ matrices ($+$ the unit matrix).

The important observation here is that such a decomposition works for $C(2k,1)$ and does not work for $C(2k+1,1)$—unless we take an *unfaithful* representation of $C(2k+1,1)$ embedded into $C(2k,1)$. For example, for the SU(3) group the same kind of calculus yields dim SU(3) = 8, therefore

$$\begin{aligned} C(8+3,1) &= C(11,1) = C(1,1)\otimes C(10,0) \\ &= C(1,1)\otimes C(2,0)\otimes C(0,8) = C(3,1)\otimes C(0,8) \\ &= C(3,1)\otimes C(0,2)\otimes C(6,0) = C(3,1) \\ &\quad \otimes\mathbb{H}\otimes C(2,0)\otimes C(0,4), \end{aligned}$$

but $C(2,0) = C(1,1)$, so

$$C(3,1)\otimes\mathbb{H}\otimes C(2,0)\otimes C(0,4)$$

$$= C(3,1)\otimes \mathbb{H}\otimes C(1,1)\otimes C(0,4)$$
$$= C(3,1)\otimes \mathbb{H}\otimes C(1,1)\otimes C(0,2)\otimes C(2,0)$$
$$= [C(3,1)\otimes \mathbb{H}]\otimes[C(3,1)\otimes \mathbb{H}]. \quad (77)$$

The quaternions being represented by $2\times 2$ matrices, the representation of $C(11,1)$ acts in a 64-dimensional space. Instead of the direct sum of two $C(3,1)\otimes \mathbb{H}$, as in (75), we get the tensor product. This, in turn, can be decomposed into a sum of irreducible representations with the use of Clebsch–Gordan coefficients.

## 5. VECTORS AND SCALAR MULTIPLETS

In order to complete the theory, we have to include in it not only *scalar* and *spinor* fields on the bundle, but also *vector fields* and *scalar multiplets*. We don't mention the *gauge* field, because it is now part of kinetics, being included in the metric geometry of the fiber bundle space. It is easy to see that both the vector field and the scalar multiplet are included in a single *vector field* over the bundle:

$$X^\alpha = \{X^i, X^a\}. \quad (78)$$

The problem now is to find a fully invariant equation for $X^\alpha$ respecting the geometrical structure of the bundle. Continuing the analogy, we would like to obtain a generalization of the Klein–Gordon equation for a massive vector field. The simplest Lagrangian of such a field, leading to the generalized Klein–Gordon equation, is the following:

$$\mathcal{L}(X) = \tfrac{1}{2} g_{\alpha\beta} g^{\rho\delta} \nabla_\rho X^\alpha \nabla_\delta X^\beta + (\mu^2/2) g_{\alpha\beta} X^\alpha X^\beta. \quad (79)$$

There is also another invariant quantity we can add to this Lagrangian, namely

$$R_{\alpha\beta} X^\alpha X^\beta. \quad (80)$$

The full Lagrangian is then

$$\mathcal{L}(X) = \tfrac{1}{2} g_{\alpha\beta} g^{\rho\delta} \nabla_\rho X^\alpha \nabla_\delta X^\beta + (\mu^2/2) g_{\alpha\beta} X^\alpha X^\beta + b R_{\alpha\beta} X^\alpha X^\beta, \quad (81)$$

$b$ being some arbitrary constant. One can also introduce the Higgs–Kibble mechanism by modifying the "potential energy" part, i.e., by adding

$$V(X) = (\lambda/4)[g_{\alpha\beta} X^\alpha X^\beta - C]^2. \quad (82)$$

The invariant equation derived from the Lagrangian (81) by the variational principle is formally very simple,

$$\nabla_\rho g^{\rho\delta} \nabla_\delta X^\alpha = 2b R^\alpha{}_\beta X^\beta + \mu^2 X^\alpha. \quad (83)$$

The direct calculus of (83) is extremely boring and apparently does not lead to any meaningful result, because the field $X^\alpha$ has no direct physical meaning. Just as in the case with spinors we have to split $X^\alpha$ into parts having well-defined properties with respect to the Lorentz transformations. The splitting is very simple and amounts to defining ver $X$ and hor $X$; in local coordinates we write

$$X^a = \varphi^a - A^a_k \varphi^k, \quad X^k = \varphi^k. \quad (84)$$

Here $\varphi^a$ is a scalar multiplet (i.e., a scalar with respect to the Lorentz group, and an $N$-tuplet in the space of the adjoint representation of the gauge group), and $\varphi^i$ is a Lorentz vector. It is much more convenient to write (81)–(83) in terms of $\varphi^a$ and $\varphi^k$. It is easy to see that now

$$g_{\alpha\beta} X^\alpha X^\beta = \mathring{g}_{ab} \varphi^a \varphi^b + \mathring{g}_{ij} \varphi^i \varphi^j \quad (85)$$

and

$$g_{\alpha\beta} g^{\rho\delta} \nabla_\rho X^\alpha \nabla_\delta X^\beta = \mathring{g}_{ab} \mathring{g}^{ij} D_i \varphi^a D_j \varphi^b + \mathring{g}_{ij} \mathring{g}^{kl} \begin{Bmatrix} i \\ kc \end{Bmatrix} \begin{Bmatrix} j \\ ld \end{Bmatrix} \varphi^c \varphi^d + \text{terms quadratic in } (\varphi^a, \varphi^k) \text{ and } (\varphi^k, \varphi^l), \quad (86)$$

where

$$D_i \varphi^a = \partial_i \varphi^a + C^a_{bc} A^b_i \varphi^c. \quad (87)$$

We choose the simplest case in which $\varphi^k = 0$, i.e., when the field $X^a$ is a *left-invariant vertical* vector field on the bundle. We don't really loose any generality, because it can be shown that $\varphi^k$ does not interact with the gauge field $A^a_k$, its contribution to the Lagrangian being of the form of free field energy. Therefore, we identify

$$g_{\alpha\beta} g^{\rho\delta} \nabla_\rho X^\alpha \nabla_\delta X^\beta = \mathring{g}_{ab} \mathring{g}^{ij} D_i \varphi^a D_j \varphi^b + R_{ab} \varphi^a \varphi^b,$$
($R_{\alpha\beta}$ is the Ricci tensor of the metric $g_{\alpha\beta}$). (88)

Now we see that in order to get a *renormalizable* theory we ought to get rid of the last term in (88). This is possible if the constant $b$ in the full Lagrangian (81) is equal to $-\frac{1}{2}$; so that

$$\mathcal{L}(X) = \tfrac{1}{2} g_{\alpha\beta} g^{\rho\delta} \nabla_\rho X^\alpha \nabla_\delta X^\beta - \tfrac{1}{2} R_{\alpha\beta} X^\alpha X^\beta + (\mu^2/2) g_{\alpha\beta} X^\alpha X^\beta + V(X). \quad (89)$$

The full Lagrangian, containing both the scalar multiplet $\varphi^a$ and the spinor field $\psi$, as well as the gauge field, is then equal to

$$\mathcal{L} = -\tfrac{1}{4} \mathring{g}_{ab} \mathring{g}^{ik} \mathring{g}^{jl} F^a_{ij} F^b_{kl} + \tfrac{1}{2} \mathring{g}_{ab} \mathring{g}^{ij} D_i \varphi^a D_j \varphi^b + V(\varphi) + \bar{\psi}[\mathring{\gamma}^i (\overleftrightarrow{\partial}_i - A^a_i C_{abc} \mathring{\sigma}^{bc}) + \mathring{\gamma}^a C_{abc} \mathring{\sigma}^{bc} - m]\psi. \quad (90)$$

Here $\bar{\psi}$ means $\psi^\dagger \gamma^{N+5}$, where

$$\gamma^{N+5} = \mathring{\gamma}^0 \mathring{\gamma}^1 \mathring{\gamma}^2 \cdots \mathring{\gamma}^{N-1} \mathring{\gamma}^N. \quad (91)$$

The full set of resulting field equations is the following:

$$\partial^i F^a_{ij} + C^a_{bc} A^{bi} F^c_{ij} = C^a_{bc} D_j \varphi^b \varphi^c - \bar{\psi}[\mathring{\gamma}_j C^a_{bc} \mathring{\sigma}^{bc}]\psi,$$
$$D^i D_j \varphi^a - (\delta V/\delta \varphi^b) \mathring{g}^{ab} = 0, \quad (92)$$
$$[\mathring{\gamma}^i (\partial_i - A^a_i C_{abc} \mathring{\sigma}^{bc}) + \mathring{\gamma}^a C_{abc} \mathring{\sigma}^{bc} - m]\psi = 0.$$

## 6. CONCLUDING REMARKS

Our main goal here is including the spinors into a geometrical unification scheme. That is why we start with a principal fiber bundle, and all our geometrical objects can *a*

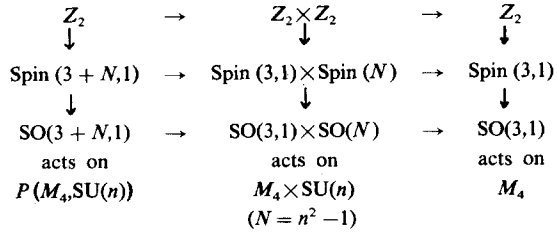



FIG. 1.

*priori* depend on the space–time point and the point on the fiber as well. Only afterwards the covariance or invariance with respect to the group action on the fiber bundle (therefore, on all the geometrical objects defined on it) restrict the way in which these objects may depend on the group parameters. We did not make any topological assumptions on the bundle $P(M_4,G)$. This may be important for the realization of global spinor structures. As a matter of fact, our problem may be illustrated by Fig. 1.

Although it is always possible to make such a diagram exact, the question remains open if the same will be true for the corresponding spinor structures over $M_4$, over $M_4\times G$ and over $P(M_4,G)$ which may be globally nontrivial. These problems have been studied deeply by Kosmann[9] and we refer to her papers for more information. For the infinitesimal study we are performing here, the local triviality of the bundle is sufficient.

The renormalizability of the set of Eq. (92) is due to the absence of terms of Pauli type, i.e., $\overset{\circ}{\gamma}_a F^a_{ij}\overset{\circ}{\sigma}^{ij}$. This, in turn, comes from the particular choice of the covariant derivative of a spinor (46) as well as from the choice of the connection itself. As a matter of fact, the connection (30) is not unique. One may argue that it is not metric, i.e., that for

$$\Gamma^{\alpha}_{\beta\gamma} = \begin{Bmatrix} \alpha \\ \beta\gamma \end{Bmatrix} + S^{\alpha}_{\beta\gamma} \tag{30}$$

the covariant derivative of the metric tensor does not vanish

$$\partial_\alpha g_{\beta\gamma} - \Gamma^{\delta}_{\alpha\beta} g_{\delta\gamma} - \Gamma^{\delta}_{\alpha\gamma} g_{\beta\delta} = Q_{\alpha\beta\gamma} = -(S^{\delta}_{\alpha\beta} g_{\delta\gamma} + S^{\delta}_{\alpha\gamma} g_{\beta\delta}) \neq 0. \tag{93}$$

There exists a unique affine connection with torsion $S$ which is metric, i.e., giving the vanishing covariant derivative of the metric tensor,

$$\tilde{\Gamma}^{\alpha}_{\beta\gamma} = \begin{Bmatrix} \alpha \\ \beta\gamma \end{Bmatrix} + S^{\alpha}_{\beta\gamma} + Q^{\alpha}_{\beta\gamma}\,, \tag{94}$$

where

$$Q^{\alpha}_{\beta\gamma} = \tfrac{1}{2} g^{\alpha\delta} Q_{\beta\delta\gamma} \tag{95}$$

and is called the nonmetric object.

Using the covariant derivative $\tilde{\nabla}_\alpha$ with respect to the connection $\tilde{\Gamma}^{\alpha}_{\beta\gamma}$ instead of the Christoffel connection or Christoffel + torsion would, of course, modify the result.

The action of the covariant derivative on spinors is by no means unique; one can argue that the proper definition should be, as proposed by Domokos and Kövesi-Domokos[5]

$$\nabla_\alpha \psi = \partial_\alpha \psi + \Gamma^{\beta}_{\alpha\gamma} g_{\beta\delta} \sigma^{\gamma\delta} \psi\,, \tag{96}$$

where

$$\sigma^{\gamma\delta} = (1/2i)(\gamma^\gamma \gamma^\delta - \gamma^\delta \gamma^\gamma). \tag{97}$$

This choice gives rise to the Pauli term in Eq. (92), while including the nonmetricity into the connection changes just the ordering in the possibilities (51), (53), (56), and (57).

The reduction of the spinors over $P(M_4,G)$ into the Lorentz spinors can already be carried out in the Lagrangian (90). Remember that the variational principle is defined over the fiber bundle; therefore we have to perform the integration in the group space (along the fiber) too. In the expression

$$\int_G \bar{\psi}[\overset{\circ}{\gamma}{}^i(\overset{\circ}{\partial}_i - A^{\ a}_i C_{abc}\overset{\circ}{\sigma}{}^{bc}) + \overset{\circ}{\gamma}{}^a C_{abc}\overset{\circ}{\sigma}{}^{bc} - m]\psi\, dV, \tag{98}$$

where $dV$ is the volume element of the group, we shall get the decomposition (see Ref. 5) into the Lorentz spinors

$$\int_G \bar{\psi}\psi\, dV = \sum_{j,\mu} (j + \tfrac{1}{2} - \tfrac{1}{4}\mu)\bar{u}_{j\mu} u_{j\mu}\,. \tag{99}$$

The decomposition (99) is derived in Ref. 5 for the SU(2) gauge group. The spinor $u_{j\mu}, \mu = 1,2,3,4$, belong to the irreducible space $(j)$ of the gauge group and can be thought of as $4(2j+1)$-column matrices.

To close these remarks, we should point out that the most useful application of the formalism exposed above would be to find some nontrivial solutions of the system (92).

We don't know if any exact solutions can be found easily, nevertheless we can expect some general features by investigating the approximate solutions with a given gauge potential $A^{\ a}_i$ (the exterior field approximation). The approximate solutions in this case can be found by means of a generalized Foldy–Wuthoysen transformation (see Ref. 10). The essential result in this case, if we take the SU(3) gauge group, is the Gell–Mann–Okubo mass formula, in which the coefficients depend on the spin configurations of the constituting quarks.


[1] A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[2] R. Kerner, Ann. Inst. H. Poincaré **9**, 147 (1968).
[3] Y.M. Cho, J. Math. Phys. **16**, 2029 (1975).
[4] R. Kerner, in *Group Theoretical Methods in Physics*, edited by Janner and Boon (Springer-Verlag, New York, 1976).
[5] G. Domokos and S. Kövesi-Domokos, Phys. Rev. D **16**, 3060 (1977).
[6] W. Drechsler and R. Sasaki, Nuovo Cimento **46**, 527 (1978).
[7] W. Drechsler, Nuovo Cimento **41**, 597 (1977).
[8] A.A. Kirillov, *Elements de la theorie des representations* (Mir, Moscow, 1975).
[9] Y. Kosmann, "Dérivées de Lie des spineurs," Ph.D. thesis, Paris (1972).
[10] R. Kerner, Ann. Inst. H. Poincaré **23**, 379 (1975).

# $S$-matrix for interacting $A$-fields

Tchavdar D. Palev[a)]
*Institut für Theoretische Physik der Technischen Universität Clausthal, Claushal-Zellerfeld, West Germany*

(Received 2 January 1980; accepted for publication 30 May 1980)

In the framework of the Lagrangian field theory a new statistics for charged tensor fields is considered. An interaction Lagrangian is constructed such that the $S$-matrix is unitary, covariant and causal.

Creation and annihilation operators (CAO's) of Bose type appear simultaneously as representatives of two different algebraical structures. Under commutation they yield the nilpotent Heisenberg or Bose Lie algebra (LA) and, on the other hand, they span a basis in the odd part of the simple orthosymplectic Lie superalgebra (LS) or Bose LS, and generate it.[1] Similarly, the Fermi operators can be viewed either as operators generating the simple Lie algebra of the orthogonal group,[2] the Fermi LA, or as odd generators of a nilpotent LS, the Fermi LS. In both cases the nilpotent algebras have a single Fock representation, whereas the Bose LS and the Fermi LA possess infinite sequences of such representations with CAO's satisfying the three-linear paraBose and paraFermi relations, respectively.[3]

Among the known simple LS's of particular interest are four infinite series, denoted by Kac[4] as $A$, $B$, $C$, and $D$. Each one resolves into an infinite sequence of Lie algebras and an infinite sequence of Lie superalgebras, that are not LA's. Any $n$ pairs of paraFermi (paraBose) CAO's generate the LA $B_n$ [the LS $B(0,n)$] from the series $B$ and therefore the parafields (and hence Bose and Fermi fields) can be called $B$-fields.

Recently,[5,6] within the framework of the Lagrangian field theory a quantization of so-called $A$-fields was given, which corresponds to the series $A$ in the same sense as parafields correspond to $B$-superalgebras. More precisely, the CAO's of the tensor fields generate proper $A$-LS's, whereas the CAO's of the spinor fields close $A$-LA's.

In the present letter we show how to build an interaction Lagrangian from tensor fields, so that the ordinary perturbation approach for constructing of an $S$-matrix remains unaltered. Such $A$-fields have some new physical properties and in particular lead to additional selection rules.

Let $\phi_1(x,\eta),\ldots,\phi_n(x,\eta)$ be a family of charged tensor fields with charge $\eta = \pm$ and positive ($\xi = +$) and negative ($\xi = -$) frequency parts represented via CAO's $a_i^{\xi}(\mathbf{p},\eta)$ as

$$\phi_i^{\xi}(x,\eta) = (2\pi)^{-3/2}\int \mathbf{dp}\,(2p^0)^{-1/2}e^{i\xi px}a_i(\mathbf{p},\eta). \tag{1}$$

The $A$-fields are defined by the three-linear relations of their creation ($\xi = +$) and annihilation ($\xi = -$) operators $a_i^{\xi}(\mathbf{p},\eta)$ given as

$$\begin{aligned}[\{a_i^{\xi}(\mathbf{p},-\epsilon\xi),a_j^{\eta}(\mathbf{q},\epsilon\eta)\},a_k^{\nu}(\mathbf{k},\mu\nu)] &= \tfrac{1}{2}(\nu-\eta)\delta_{-\epsilon,\mu}\delta_{j,k}\delta(\mathbf{q}-\mathbf{k})a_i^{\xi}(\mathbf{p},-\epsilon\xi)\\ &+\tfrac{1}{2}(\nu-\xi)\delta_{\epsilon,\mu}\delta_{i,k}\delta(\mathbf{p}-\mathbf{k})a_j^{\eta}(\mathbf{q},\epsilon\eta)\\ &+\tfrac{1}{2}(\xi-\eta)\mu\epsilon\delta(\mathbf{p}-\mathbf{q})a_k^{\nu}(\mathbf{k},\mu\nu),\end{aligned} \tag{2}$$

$$\{a_i^{\xi}(\mathbf{p},\epsilon\xi),a_j^{\eta}(\mathbf{q},\epsilon\eta)\} = 0, \tag{3}$$

with $\xi,\eta,\delta,\epsilon,\mu,\nu = \pm$ or $\pm 1$, $[x,y] = xy - yx$, $\{x,y\} = xy + yx$. From (2) we have

$$\{\phi(x,\eta),\phi(y,\eta)\} = 0, \tag{4}$$

whereas for $(x-y)^2 < 0$

$$\{\phi(x,\eta),\phi(y,-\eta)\} \neq 0. \tag{5}$$

It turns out, however, that any two operators

$$\begin{aligned}A_{ij}(x) &= \{\phi_i(x,+),\phi_j(x,-)\},\\ A_{kl}(y) &= \{\phi_k(y,+),\phi_l(y,-)\}\end{aligned} \tag{6}$$

commute at space-like distances. To prove this use the representation (1)

$$\begin{aligned}[A_{ij}(x),A_{kl}(y)] &= \sum_{\xi,\eta,\delta,\epsilon}[\{\phi_i^{\xi}(x,\xi),\phi_j^{\eta}(x,-\eta)\},\{\phi_k^{\delta}(y,\delta),\phi_l^{\epsilon}(y,-\epsilon)\}]\\ &= \sum_{\xi,\eta,\delta,\epsilon}(2\pi)^{-6}\int\frac{d\mathbf{p}d\mathbf{q}d\,\mathbf{k}d\mathbf{r}}{4(p^0q^0k^0r^0)^{1/2}}\\ &\times e^{i\xi px+i\eta qx+i\delta ky+i\epsilon ry}I_{ijkl}^{\xi\eta\delta\epsilon}(\mathbf{p},\mathbf{q},\mathbf{k},\mathbf{r})\end{aligned} \tag{7}$$

and

$$\begin{aligned}I_{ijkl}^{\xi\eta\delta\epsilon}(\mathbf{p},\mathbf{q},\mathbf{k},\mathbf{r},) &= [\{a_i^{\xi}(\mathbf{p},\xi),a_j^{\eta}(\mathbf{q}-\eta)\},\{a_k^{\delta}(\mathbf{k},\delta),a_l^{\epsilon}(\mathbf{r},-\epsilon)\}].\end{aligned} \tag{8}$$

The structure relations (2) yield

$$\begin{aligned}I_{ijkl}^{\xi\eta\delta\epsilon}(\mathbf{p},\mathbf{q},\mathbf{k},\mathbf{r},) &= -\eta\delta_{-\eta,\delta}\delta_{j,k}(\mathbf{q}-\mathbf{k})\{a_i^{\xi}(\mathbf{p},\xi),a_l^{\epsilon}(\mathbf{r},-\epsilon)\}\\ &+\epsilon\delta_{\xi,-\epsilon}\delta_{i,l}\delta(\mathbf{p}-\mathbf{r})\{a_j^{\eta}(\mathbf{q},-\eta),a_k^{\delta}(\mathbf{k},\delta)\}.\end{aligned} \tag{9}$$

Inserting (9) in (7) and integrating over all 3-momenta, we finally obtain

$$\begin{aligned}[\{\phi_i(x,+),\phi_j(x,-)\},\{\phi_k(y,+),\phi_l(y,-)\}] &= -i\delta_{j,k}D(x-y)\{\phi_i(x,+),\phi_l(y,-)\}\\ &-i\delta_{i,l}D(x-y)\{\phi_j(x,-),\phi_k(y,+)\},\end{aligned} \tag{10}$$

where $D(x-y)$ is the Pauli–Jordan commutator function. Therefore the operators (6) commute for $(x-y)^2 < 0$. This immediately implies that any polynomials (or, more general-

[a)]Present address: Institute of Nuclear Research and Nuclear Energy, Boul. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/80/102560-02$1.00  

ly, analytical functions) of operators of the type (6) also commute at space-like distances.

Consider now an interaction based on a relativistic invariant hermitian Lagrangian density $L_{\rm int}$, which depends on the $A$–fields only through the symmetrized combinations (6). Suppose moreover that the other interacting (Bose or Fermi) fields commute with the $A$–fields. Then $L_{\rm int}$ is a local operator, i.e., for $(x-y)^2<0$

$$[L_{\rm int}(x),L_{\rm int}(y)]=0. \tag{11}$$

Therefore one proves in the usual way for the perturbation approach that the S–matrix

$$S=T\exp\left[i\int L_{\rm int}(x)\,dx\right] \tag{12}$$

is a covariant, unitary, and causal operator.[7]

The Fock spaces $W$ of the $A$–fields are defined from the requirements

$$a_i^-(\mathbf{p},\eta)a_j^+(\mathbf{q},\eta)|0\rangle=\delta_{ij}\delta(\mathbf{p}-\mathbf{q})z_{\pm}|0\rangle,\ \eta=\pm$$
$$a_i^-(\mathbf{p},\eta)|0\rangle=0,\quad \langle 0|a_i^+(\mathbf{p},\eta)=0, \tag{13}$$

where $|0\rangle$ is the no-particle state. A more detailed study[6] shows that the metric of $W$ is positive definite if and only if the numbers $z_n$ are nonnegative integers. A vector

$$a_{i_1}^+(\mathbf{p}_1,\eta_1)\cdots a_{i_n}^+(\mathbf{p}_n,\eta_n)|0\rangle \tag{14}$$

from $W$ is different from zero only if

$$-z_-\leqslant z=\eta_1+\ldots+\eta_n\leqslant z_+. \tag{15}$$

This is an analogue of the Pauli principle for the tensor $A$–fields. It shows that the charge $z$ of an arbitrary ensemble of particles can not be more than $z_+$ nor less than $-z_-$. Therefore special transitions between Bose states will be forbidden for the $A$–fields. The question whether the allowed reactions lead to new physical predictions remains to be investigated.

I would like to thank Professor H. D. Doebner for the valuable discussions and the several suggestions he made, which were taken into account.

[1]A. Gantchev and T. Palev, J. Math. Phys. **21**, 797 (1980).
[2]S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1960); C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[3]H. S. Green, Phys. Rev. **90**, 270 (1953).
[4]V. G. Kac, Adv. Math. **26**, 8 (1977).
[5]T. Palev, Thesis, Sofia (1976); Preprints JINR E2-11904, 11905, Dubna (1978); to appear in Rep. Math. Phys.
[6]T. Palev, Preprint JINR E2-11942, Dubna (1978).
[7]N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, English transl. (Interscience, New York, 1959).

# Unitarity of supergravity and $Z_2$ or $Z_2\times Z_2$ or $Z_2\times Z_2\times Z_2$ gradings of gauge and ghost fields

P. van Nieuwenhuizen
*Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794*

(Received 13 February 1980; accepted for publication 16 May 1980)

Unitarity can be proven from the usual $Z_2$ grading of gauge and ghost fields, or from a $Z_2\times Z_2$ grading, geometrically derived by Ne'eman and Thierry-Mieg, or from a $Z_2\times Z_2\times Z_2$ grading derived here. The claim that only the $Z_2\times Z_2$ grading leads to unitarity is incorrect. The opposite is shown to hold: signs due to different gradings are physically unobservable. We show how the $Z_2\times Z_2$ grading follows from the $Z_2$ grading by taking a product space.

## I. INTRODUCTION

In a series of recent papers, Ne'eman and Thierry-Mieg (NT) have proposed a geometrical derivation of BRST (Becchi–Rouet–Stora–Tyutin) transformations.[1] They arrive at an unusual $Z_2\times Z_2$ statistics for ghost and gauge fields. For example, a commuting ghost field $C^a$ and an anticommuting ghost field $C^\nu$ anticommute rather than commute with each other. They also claim to have completed the proof of unitarity of supergravity of Sterman, Townsend, and van Nieuwenhuizen.[2] In this proof, no signs due to the commutator of $C^\nu$ and $C^a$ were mentioned.

We show below that the $Z_2\times Z_2$ grading of NT, which divides fields into (integer spins/half-integer spins) and (physical/ghost) needs an anticommuting parameter which commutes with gauge fields but which uniformly commutes or anticommutes with ghost fields. Then we show that the $Z_2\times Z_2$ grading is simply obtained from the usual $Z_2$ grading by considering a product space. We are indebted to B. Zumino for showing this relation to us.

Finally, we explain why the signs due to ordering $C^\nu$ and $C^a$ cancel in the unitarity proof. The reason is, briefly, that in a cut graph one needs a Ward identity once on the left-hand side and once on the right-hand side, and any unusual sign on the left is cancelled by a similar sign on the right.

The claim of NT that the proof of unitarity is complete only with their $Z_2\times Z_2$ grading, and that this gives support for their theory, is not substantiated. Also the standard BRST formalism can be used to prove unitarity. The unobservability of the unusual NT signs is of the same kind as the unobservability of defining two different spinors to commute rather than to anticommute.

In Sec. 2 we complete the rules of the $Z_2\times Z_2$ grading and show that the usual quantum action is again BRST invariant, provided $\delta\phi^i$ and $lC^\alpha$ have the same commutation properties as $\phi^i$ and $\xi^\alpha$ and ($\xi^\alpha$ are the gauge parameters). In Sec. 3 we relate the $Z_2$ to the $Z_2\times Z_2$ formalism. In Sec. 4 we carefully analyze the signs in the unitarity proof.

## II. BRST INVARIANCE OF THE ACTION

Consider a classical action $I_{cl}$ depending on gauge fields $\phi^i$ which is invariant under local gauge transformations $\delta\phi^i = R^i{}_\alpha\xi^\alpha$ with $\xi^\alpha$ on the extreme right.[3] If extra signs appear for some $\alpha$ ("$\alpha$–dependent signs") we eliminate them by redefining $\xi^\alpha$. Also $i$–dependent signs are eliminated by redefining $R^i{}_\alpha$.

The BRST rules in the NT scheme for $\phi^i$ read

$$\delta\phi^i = R^i{}_\alpha lC^\alpha. \tag{2.1}$$

We write for later purposes $l$ to the left of $C^\alpha$. Possible $\alpha$–dependent signs due to positioning $l$ in this way are eliminated by redefining $C^\alpha$. No $i$–dependent signs are possible if the classical action $I_{cl}$ is to be invariant under (2.1). Moreover, $\phi^i$ and $\delta\phi^j$ must have the same statistics as $\phi^i$ and $\phi^j$, otherwise $I_{cl}$ will not be invariant in general. Let us see what consequences this has.

In the NT scheme, fields have a $Z_2\times Z_2$ grading. According to the first $Z_2$ grading, fields are divided into integer spin fields ($a_1 = 0$) and half-integer spin fields ($a_1 = 1$). The other $Z_2$ grading divides fields into physical fields ($a_2 = 0$) and ghost fields ($a_2 = 1$). Thus for two fields $A$ and $B$ with gradings $(a_1,a_2)$ and $(b_1,b_2)$ one has

$$A(a_1,a_2)B(b_1,b_2) = B(b_1,b_2)A(a_1,a_2)(-)^{a_1b_1+a_2b_2}. \tag{2.2a}$$

For $A=\phi^i$ and $B=C^\alpha$, one has $a_2=0$, so that the commutation relations between any ghost field $C^\alpha$ and a physical field $\phi^i$ are unusual if and only if $\phi^i$ is fermionic. Thus $\delta\phi^i$ commutes with $\phi^j$ in the same way as $\phi^i$ with $\phi^j$ if *$l$ commutes with $\phi^i$*, and in this case $I_{cl}$ is invariant under (2.1). In the usual scheme,[4] fields have a $Z_2$ grading

$$A(a_1,a_2)B(b_1,b_2) = B(b_1,b_2)A(a_1,a_2)(-)^{(a_1+a_2)(b_1+b_2)}. \tag{2.2b}$$

The most general gauge fixing term quadratic in gauge functions $F_\alpha$ is given by

$$\mathcal{L}(\text{fix}) = \tfrac{1}{2}F_\alpha\gamma^{\alpha\beta}F_\beta\,, \tag{2.3}$$

where $\gamma^{\alpha\beta}$ is independent of $\phi^i$.[5] Its variation is simply

$$\delta\mathcal{L}(\text{fix}) = F_\alpha\gamma^{\alpha\beta}F_{\beta,j}R^j{}_\gamma lC^\gamma, \tag{2.4}$$

where, $j$ denotes right differentiation. The ghost action and the BRST variation law for the antighost read

$$\mathcal{L}(\text{ghost}) = C^{*\beta}F_{\beta,j}R^j{}_\gamma C^\gamma, \tag{2.5}$$

$$\delta C^{*\beta} = -lF_\alpha\gamma^{\alpha\beta}. \tag{2.6}$$

Possible $\beta$–dependent signs in (2.6) are eliminated by redefining $C^{*\beta}$, but if (2.6) inserted into (2.5) is going to cancel (2.4), no $\alpha$–dependent signs can occur in (2.6), and no $\beta, j$, or

 0022-2488/80/102562-05$1.00  

$\gamma$ dependent signs can occur in (2.5). *Hence, the quantum action is the same as usual.* The unusual sign in (2.6) is due to the fact that pulling $l$ from right to the extreme left in (2.4) does not yield an overall minus sign. In the usual scheme, $\delta\phi^i = R^i{}_\alpha C^\alpha \Lambda$, and pulling $\Lambda$ from the extreme right to the extreme left in $\delta\mathcal{L}$ (fix) yields an overall minus sign since the variation of an action is always bosonic.

We must now show that the variation of $F_{\beta,j}R^j{}_\gamma C^\gamma$ vanishes. Since one can consider $F_\beta$ which are linear in $\phi^i$ as well as $F_\beta$ which are nonlinear in $\phi^i$, one has two requirements

$$(\delta F_{\beta,j})R^j{}_\gamma C^\gamma = 0 \quad \text{and} \quad F_{\beta,j}\delta(R^j{}_\gamma C^\gamma) = 0. \tag{2.7}$$

Together they state that the double variation of $F_\beta$ vanishes, i.e., nilpotency of BRST. The first requirement reads in full (writing $l_1$ rather than $l$)

$$F_{\beta,jk}R^k{}_\alpha l_1 C^\alpha R^j{}_\gamma C^\gamma = 0, \tag{2.8}$$

and multiplying on the right with $l_2$, one finds for a pair of $(j,k)$ *two* terms (no sum over $j$, $k$ now)

$$F_{\beta,jk}(R^k{}_\alpha l_1 C^\alpha)(R^j{}_\gamma C^\gamma l_2) + F_{\beta,kj}(R^j{}_\gamma l_1 C^\gamma)(R^k{}_\alpha C^\alpha l_2). \tag{2.9}$$

This pair of terms vanishes if $R^k{}_\alpha l_1 C^\alpha$ and $R^j{}_\gamma C^\gamma l_2$ have the same commutation properties as $\phi^k$ and $\phi^j$ and if in addition $l_1$ and $l_2$ anticommute *or* if $R^k{}_\alpha l_1 C^\alpha$ and $R^j{}_\gamma C^\gamma l_2$ have different commutation properties while $l_1$ and $l_2$ commute. Thus, for all $k$ and $j$, $\delta\phi^k$ and $\delta\phi^j$ must have the same or opposite commutation properties as $\phi^k$ and $\phi^j$. The commutation properties of $\delta\phi^k$ and $\delta\phi^j$ are

$$\delta_1\phi^j\delta_2\phi^k = \delta_2\phi^k\delta_1\phi^j(-)^{S_1+S_2+S_3} \tag{2.10}$$

$$S_1 = a_1(\phi^j)a_1(\phi^k) \tag{2.11}$$

$$S_2 = [a_1(C^\gamma)+1] + [a_1(C^\alpha)+1] + [a_1(C^\gamma)+1][a_1(C^\alpha)+1]+1 \tag{2.12}$$

$$S_3 = \sigma(l_1,C^\alpha) + \sigma(l_2,C^\gamma) + [a_1(C^\gamma)a_1(C^\alpha)+1] + \sigma(l_1,l_2). \tag{2.13}$$

The term $S_1$ would result if $\delta\phi^i$ and $\delta\phi^j$ had the usual commutation properties, the term $S_2$ accounts for the interchanging of $l_1C^\gamma$ and $C^\alpha l_2$ in the usual scheme, and thus in $S_1+S_2$ one has not yet specified how $l_1C^\gamma$ and $C^\alpha l_2$ commute (we have used that in the NT scheme $R^j{}_\gamma$ and $C^\alpha l_2$ commute as usual, which is equivalent to requiring that $\delta\phi^j$ and $\phi^k$ have the usual commutation properties). Finally, $S_3$ specifies how $l_1C^\gamma$ and $C^\alpha l_2$ commute in the NT scheme. The symbol $\sigma(l_1,C^\alpha)$ is $+1$ if $l_1$ and $C^\alpha$ commute and $-1$ if they anticommute.

Thus, $S_2+S_3$ must be even. This is the case if $\sigma(l_1,C^\alpha) + \sigma(l_2,C^\gamma)$ is even. Hence, *either $l$ commutes with all $C^\alpha$, or $l$ anticommutes with all $C^\alpha$, while $l_1$ and $l_2$ anticommute.*

We now return to the second term in Eq. (2.7). It requires that

$$R^j{}_{\gamma,k}(R^k{}_\alpha l C^\alpha)C^\gamma + R^j{}_\gamma \delta C^\gamma = 0. \tag{2.14}$$

This relation must be proved by using *closure of the gauge algebra*

$$R^j{}_{\gamma,k}R^k{}_\alpha(\eta^\alpha\xi^\gamma - \xi^\alpha\eta^\gamma) = R^j{}_\beta f^\beta{}_{\gamma\alpha}\eta^\alpha\xi^\gamma. \tag{2.15}$$

Putting $\eta^\alpha = lC^\alpha$ and $\xi^\gamma = C^\gamma l_2$, one can use Eq. (2.15) and (2.14) only if two properties hold

(i) these special $\eta^\alpha$ and $\xi^\gamma$ have the same commutation properties as the usual parameters $\eta^\alpha$ and $\xi^\gamma$. In this case Eq. (2.15) remains an identity;

(ii) the terms with $-\xi^\alpha\eta^\gamma$ is equal to the term with $\eta^\alpha\xi^\gamma$. In the usual scheme this is obvious ($-C^\alpha\Lambda_2 C^\gamma\Lambda = C^\alpha\Lambda C^\gamma\Lambda_2$ no matter now a single $\Lambda$ commutes with $C^\gamma$).

The condition (i) is equivalent to $S_3 = a_1(C^\alpha)a_1(C^\gamma)$ and is thus always satisfied. The condition (ii) is also satisfied, since $l_1$ and $l_2$ anticommute. It follows that

$$\delta C^\gamma = -\tfrac{1}{2}f^\gamma{}_{\alpha\beta}lC^\beta C^\alpha. \tag{2.16}$$

Thus, the usual quantum action is invariant under the modified rules [Eqs. (2.1), (2.6), and (2.16)]. The fields commute as in Eq. (2.2) and $l$ anticommutes with itself while $l$ commutes with $\phi^i$. Finally, $l$ can either commute with all $C^\alpha$ or it anticommutes with all $C^\alpha$.

These results strongly suggest that $l$ is an anticommuting parameter which acts in the second $Z_2$ grading or in a space outside both $Z_2$ gradings. Thus there are three possible BRST formalisms in which $l$ is always anticommuting with itself

(i) $l$ and $A$ $(a_1,a_2)$ anticommute if and only if $a_1+a_2=1$. This is the usual scheme[4];

(ii) $l$ and $A$ $(a_1,a_2)$ anticommute if and only if $a_2=1$. This is the NT scheme[1];

(ii) $l$ and $A$ $(a_1,a_2)$ always commute. This is a new scheme, proposed here. One may formally call this a $Z_2\times Z_2\times Z_2$ grading. The last $Z_2$ factor is nonzero only for $l$ itself.

## III. RELATION BETWEEN $Z_2$ GRADING AND $Z_2\times Z_2$ GRADING

We now show that one can derive the statistics rules in Eq. (2.2) from the usual statistics rules by making a direct product space. We are indebted to B. Zumino for showing us this possibility.[6]

Consider a Grassmann algebra with $Z_2$ grading, with even elements $x^j$ (the bosonic physical fields), odd elements $\theta^\alpha$ (the fermionic physical fields) and differentials $dx^j$ (even elements, the integer spin ghosts) and $d\theta^\alpha$ (odd elements, the half-integer spin ghosts). The commutation rules of the $Z_2$ grading of this Grassmann algebra produce the usual statistics for physical fields and ghost fields.

On the other hand, consider a Grassmann algebra with $Z_2\times Z_2$ grading, whose elements are

$$X^j = x^j\otimes\mathbf{1}, \quad \eta^\alpha = \theta^\alpha\otimes\sigma_3$$
$$DX^j = dx^j\otimes\sigma_1, \quad D\eta^\alpha = d\theta^\alpha\otimes(-i\sigma_2). \tag{3.1}$$

The $Z_2\times Z_2$ commutation table reads

| | $X^j$ | $\eta^\alpha$ | $DX^j$ | $D\eta^\alpha$ |
|---|---|---|---|---|
| $X^j$ | + | + | + | + |
| $\eta^\alpha$ | + | − | ⊕ | ⊖ |
| $DX^j$ | + | ⊕ | − | ⊖ |
| $D\eta^\alpha$ | + | ⊖ | ⊖ | + |

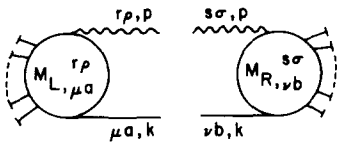



FIG. 1.

The + signs denote that the elements defining the row and column commute, the − signs that they anticommute, and one can even express $D$ in terms of $d$ by $D = d \times \sigma_1$. The unusual signs are encircled, and this table corresponds exactly to Eq. (2.2b) if one identifies $X^j(\eta_\alpha)$ with physical fields, and $DX^j(D\eta^\alpha)$ with ghosts, with integer (half-integer) spins.

This simple relation between the $Z_2$ and the $Z_2 \times Z_2$ grading leads one to expect that the encircled spins in the table are unobservable. That this is indeed the case we will show next.

## IV. UNITARITY OF SUPERGRAVITY AND STATISTICS OF GHOST FIELDS

In this section we will first discuss the proof of unitarity of supergravity using the ordinary $Z_2$ statistics. Afterwards we will discuss whether there are any changes if one uses instead $Z_2 \times Z_2$ statistics. In the diagrammatic proof of Sterman, Townsend, and van Nieuwenhuizen, no signs which distinguish between supersymmetry ghosts and coordinate ghosts were given. The reason is that such signs always cancel. To make this completely explicit, we will consider a graph consisting of a virtual gravitino and graviton, exchanged between blobs with external lines which are on-shell with physical polarizations. We will omit these explicit lines since they are the same in all cases considered. For a discussion of unitarity based on a Gupta-Bleuler type of approach, see Ref. 10.

The gauge fixing term is chosen as

$$\mathscr{L}(\text{fix}) = \tfrac{1}{4}\bar{\psi}\cdot\gamma\partial\gamma\cdot\psi - \tfrac{1}{4}\left[\partial_\mu(\sqrt{g}g^{\mu\nu})\right]^2 + \tfrac{1}{2}\lambda\,(e_{m\mu}\delta^\mu_n - m\leftrightarrow n)^2. \tag{4.1}$$

In general, $\mathscr{L}(\text{fix}) = \frac{1}{2}F_\alpha\gamma^{\alpha\beta}F_\beta$ and $\mathscr{L}(\text{ghost}) = C^{*\alpha}F_{\alpha,j}R^j_\beta C^\beta$. Since we normalize as $\delta\psi_\mu = \partial_\mu\epsilon +$ more, and $\mathscr{L}(\text{ghost}) = -\bar{\mathbf{C}}\partial C - C^{*\nu}\Box C^\nu +$ more, we find

$$F_\beta = \left[-(\gamma\cdot\psi)^a,\ \partial_\lambda(\sqrt{g}g^{\lambda\mu}),\ e^m_\mu\partial^{\mu n} - m\leftrightarrow n\right] \tag{4.2}$$

$$F_\alpha\gamma^{\alpha\beta} = \left[\tfrac{1}{2}\bar{\psi}\cdot\gamma\overleftarrow{\partial},\ -\tfrac{1}{2}\partial_\lambda(\sqrt{g}g^{\lambda\nu})\partial_{\nu\mu},\ \tfrac{1}{2}\lambda\,(e_{m\mu}\delta_{\mu n} - m\leftrightarrow n)\right]. \tag{4.3}$$

The two Ward identities we will need are both derived from

$$\delta\langle\phi^i C^{*\beta}\rangle = 0 = \langle(R^i{}_\alpha C^\alpha\Lambda)C^{*\beta}\rangle + \langle\phi^i(-\Lambda F_\alpha\gamma^{\alpha\beta})\rangle. \tag{4.4}$$

The first Ward identity follows if one puts $\phi^i = e^{s\sigma}$ and $C^{*\beta}$ equal to the supersymmetry antighost

$$-\tfrac{1}{2}\langle e_{s'\sigma'}\delta^{s's}\delta^{\sigma'\sigma}\bar{\psi}\cdot\gamma\overleftarrow{\partial}\rangle = \langle\partial^\sigma C^\alpha\delta^s_\alpha\bar{C}^a\rangle. \tag{4.5}$$

The second Ward identity is obtained for $\phi^i = \psi_{\mu' a}$, and $C^{*\beta}$ equal to the coordinate antighost

$$\tfrac{1}{2}\langle\psi_{\mu' a}\partial_\lambda(\sqrt{g}g^{\lambda\sigma})\rangle = \langle(\partial_{\mu'}C_a)C^{*\sigma}\rangle. \tag{4.6}$$

We omitted in Eq. (4.5) a term $\langle C^{sn}\delta^\sigma_n(\bar{\psi}\cdot\gamma\overleftarrow{\partial})\rangle$ since the Lorentz ghosts $C^{sn}$ do not propagate and hence do not contribute in the unitarity relations (see Ref. 7, where it is shown that this is true after rediagonalizing the Lorentz ghosts $C^{mn}$).

The propagators are normalized as usual[8]

$$P^{3/2}_{\mu a,\nu b}(k) = \tfrac{1}{2}(\gamma_\nu k\!\!\!/\gamma_\mu)_{ab}, \quad P^{1/2}{}_{ab}(k) = -k\!\!\!/_{ab}$$
$$P^{2,s\sigma,r\rho}(p) = -\tfrac{1}{2}i(\delta^{sr}\delta^{\sigma\rho} + \delta^{s\rho}\delta^{r\sigma} - \delta^{s\sigma}\delta^{r\rho}), \quad P^1_{\mu\nu} = -i\delta_{\mu\nu}. \tag{4.7}$$

We have taken only the propagator of the symmetric part of $e^{r\rho}$ since its antisymmetric part does not propagate.

Consider now the graph in Fig. 1, in which a gravitino and a graviton are exchanged. Let the $S$–matrix be given by

$$S = P^{2,s\sigma,r\rho}(p)M^{r\rho}_{\text{L},\mu a}P^{3/2}_{\mu a,\nu b}(k)M^{s\sigma}_{\text{R},\nu b} \tag{4.8a}$$
$$= \langle(e^{r\rho}M^{r\rho}_{\text{L},\mu a}\psi_{\mu a})(\bar{\psi}_{\nu b}M^{s\sigma}_{\text{R},\nu b}e^{s\sigma})\rangle. \tag{4.8b}$$

Lower indices refer to the lower lines in all graphs, upper indices to the upper lines, and $L$ and $R$ denote left and right, respectively. We will compare this graph to the graph in Fig. 2, in which the gravitino and the graviton are replaced by a super symmetry ghost and a coordinate ghost, respectively. Let the $S$–matrix for Fig. 2 be given by

$$S = N^\rho_{\text{L},a}P^{1/2}_{ab}(k)N^\sigma_{\text{R},b}P^{1,\sigma\rho}(p) \tag{4.9a}$$
$$= -\langle(C^{*\rho}N^\rho_{\text{L},a}C_a)(\bar{\mathbf{C}}_bN^\sigma_{\text{R},b}C_\sigma)\rangle. \tag{4.9b}$$

The minus sign is cancelled if one moves $C^{*\rho}$ from the extreme left to the extreme right since the effective action $\mathbf{C}^\rho N^\rho_{\text{L},a}C^a$ is itself always bosonic.

If both gravitino and graviton have unitary propagators, the theory is manifestly unitary. These unitary propagators are obtained from the renormalizable propagators in a very simple way, namely by replacing $\delta_{\mu\nu}$ by $\delta^T_{\mu\nu} = \epsilon^+_\mu\epsilon^-_\nu + \epsilon^-_\mu\epsilon^+_\nu$. We refer the reader to Ref. 2 for a detailed discussion. In one of the propagators in Fig. 1 one may replace the unitary propagator by the renormalizable propagator; we choose for this the graviton propagator. The gravitino propagator is now rewritten, using the on-shell identity[9]

$$\delta^T_{\mu\nu} = \delta_{\mu\nu} - (k_\mu\mathbf{k}_\nu + k_\nu\mathbf{k}_\mu)(k\cdot\mathbf{k})^{-1}, \tag{4.10}$$

where $\mathbf{k}_\mu$ is the time-reserved of $k_\mu$, as follows

$$P^{3/2}_{ab}(k)^{\text{unit}} = \left(\delta_{\nu\nu'} - \frac{k_\nu\mathbf{k}_{\nu'}}{k\cdot\mathbf{k}}\right)(\tfrac{1}{2}\gamma_{\nu'}k\!\!\!/\gamma_{\mu'}) \times(\delta_{\mu\mu'} - \mathbf{k}_{\mu'}k_\mu(k\cdot\mathbf{k})^{-1}). \tag{4.11}$$

We will consider the term in (4.11)

$$k_\mu(\tfrac{1}{2}\gamma_\nu k\!\!\!/\gamma_{\mu'})[-k_{\mu'}(k\cdot\mathbf{k})^{-1}] \tag{4.12}$$

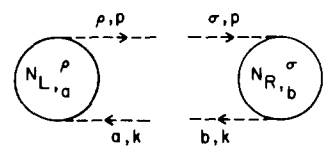



FIG. 2.

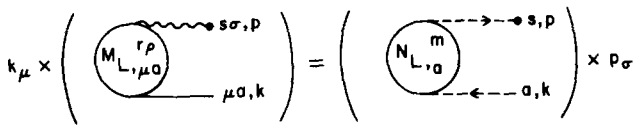


FIG. 3.

and show that if $k_\mu$ hits the left vertex $M_L$, then $M_L$ flips into $N_L$ which absorbs $k_\mu$ and emits a $p^\sigma$ upstairs. This $p^\sigma$ travels to the right, hits $M_R$, flips $M_R$ into $N_R$, which absorbs $p^\sigma$ and emits a $k_{\mu'}$ downstairs. Finally this $k_{\mu'}$ hits $-\mathbf{k}_{\mu'}(k\cdot\mathbf{k})^{-1}$ and produces $(-1)$ times the ghost graph.

We thus consider the expression (4.8a) substituting (4.12)

$$Z = (P^{2,s\sigma,r\rho}(p)M^{r\rho}_{L,\mu a}k_\mu)(\tfrac{1}{2}\gamma_\nu k\!\!\!/\gamma_{\mu'})_{ab} \times\left(-\frac{\bar{k}_{\mu'}}{k\cdot\mathbf{k}}\right)M^{s\sigma}_{R,\nu b}. \tag{4.13}$$

The first Ward identity [Eq. (4.5)] reads, in terms of the effective actions in Eq. (4.8b) and (4.9b), (see Fig. 3)

$$\begin{aligned}-\tfrac{1}{2}\langle e^{s\sigma}(e^{r\rho}M^{r\rho}_{L,\mu a'}\psi_{\mu a'})[\bar{\psi}_\nu\gamma_\nu(-ik\!\!\!/)]_a\rangle &= -P^{2,s\sigma,r\rho}(p)M^{r\rho}_{L,\mu a'}k_\mu(-ik\!\!\!/)_{a'a}\\ &= \langle\partial^\sigma C^s(C^{*m}N^m_{L,a'}C_{a'})\bar{C}_a\rangle\\ &= ip^\sigma P^{1,sm}(p)N^m_{L,a'}(-k\!\!\!/)_{a'a}.\end{aligned} \tag{4.14}$$

Hence, we may replace $P^2M_Lk$ in Eq. (4.13) by $p^\sigma P^1N_L$ and find

$$Z = (p^\sigma P^{1,sm}(p)N^m_{L,a'})\left(-\frac{\mathbf{k}_{\mu'}}{k,\mathbf{k}}\right) \times(\tfrac{1}{2}\gamma_\nu k\!\!\!/\gamma_{\mu'})_{ab}M^{s\sigma}_{R,\nu b}. \tag{4.15}$$

We now use the second Ward identity in Eq. (4.6) depicted in Fig. 4

$$\begin{aligned}\tfrac{1}{2}\langle\psi_{\mu'a}(\bar{\psi}_{\nu b}M^{s'\sigma}_{R,\nu b}e^{s'\sigma})\partial^\lambda(\sqrt{g}g^{\lambda s})\rangle &= (\tfrac{1}{2}\gamma_\nu k\!\!\!/\gamma_{\mu'})_{ab}M^{s\sigma}_{R,\nu b}p^\sigma\\ &= \langle\partial_{\mu'}C_a(\bar{C}_bN^{\sigma'}_{R,b}C^{\sigma'})C^{*s}\rangle\\ &= ik_{\mu'}(-k\!\!\!/)_{ab}N^{\sigma'}_{R,b}P^{1,\sigma's}(p).\end{aligned} \tag{4.16}$$

We used that $\langle e^{s'\sigma}\partial^\lambda(\sqrt{g}g^{\lambda s})\rangle = p^s\delta^{s\sigma}+p^\sigma\delta^{ss'}$ and that $M^{s\sigma'}_R$ can be taken symmetric in $s\sigma'$ [see below Eq. (4.7)]. Thus we find for $Z$

$$Z = P^{1,sm}(p)N^m_{L,a'}\left(-\frac{\mathbf{k}_{\mu'}}{k\cdot\mathbf{k}}\right)k_{\mu'}P^{1/2}_{ab}(k)N^{\sigma'}_{R,b}\delta^{\sigma's}$$
$$= (-1)\text{times(4.9a)}.$$

Hence, we find precisely the correct result: the term Eq. (4.12) leads to minus the ghost loop, the latter with a definite orientation.

Thus, as shown in Ref. 2, unitarity holds for supergravity if the BRST formalism, which yields the Ward identities, is based on a $Z_2$ grading of gauge and ghost fields.

We now discuss whether there are any modifications when one chooses the $Z_2\times Z_2$ grading of NT, or the $Z_2\times Z_2\times Z_2$ grading we have discussed. The quantum action being unchanged, Eqs. (4.2) and (4.3) are the same. The Ward identities become $\langle R^i_\alpha C^\alpha C^{*\beta}+\phi^i F_\alpha\gamma^{\alpha\beta}\rangle = 0$ since $l$ commutes with $\phi^i$. Also Eqs. (4.5) and (4.6) are not modified. In Eq. (4.8b) bringing $e^{s\sigma}$ from the right to the left yields Eq. (4.8a) without extra sign, since a physical integer spin field commutes with all other fields. In Eq. (4.9b), bringing $C^{*\rho}$ to the right again yields an extra minus sign in Eq. (4.9a), since $C^{*\rho}$ anticommutes with $C_a$ but commutes with $\psi_{\mu a}$ (in the usual $Z_2$ grading the opposite is true: $C^{*\rho}$ commutes with $C_a$ but anticommutes with $\psi_{\mu a}$). Thus Eqs. (4.9a) and (4.9b) are unmodified. The rest of the proof is numerical and independent of statistics.

Hence, also in the $Z_2\times Z_2$ or the $Z_2\times Z_2\times Z_2$ scheme, unitarity holds. We can say that the signs which are only due to a different grading, are not observable in the unitarity relation.

## V. CONCLUSIONS

There are three BRST formalisms. All three use the same quantum action and an anticommuting parameter, but whereas in the first scheme $\Lambda$ anticommutes with gravitino and coordinate ghosts, in the second scheme $l$ commutes with all gauge fields and anticommutes with all ghosts, while in the third scheme (derived here) $l$ commutes both with gauge and with ghost fields. In the second and third scheme, fields commute as in Eq. (2.2a).

The relation in Eq. (2.2a) can be obtained from the usual relations in Eq. (2.2b) by taking a direct product space (Sec. 3). This suggests that the signs which are different in different gradings are physically unobservable. This is indeed the case in the unitarity relation. A detailed analysis shows that in all three schemes, unitarity can be proven (Sec. 4).

Thus, the proof of Ref. 2 is not only complete for the $Z_2$ cases but applies equally well to the $Z_2\times Z_2$ cases and whether one prefers the usual $Z_2$ grading, or any of the other gradings is a matter of taste. In particular, one cannot state that the $Z_2\times Z_2$ scheme is the only correct scheme since only it leads to unitarity.

In loop calculations, the differences in statistics cancel. For example, a loop obtained by contracting $(\bar{C}^\nu C^a)(\bar{C}^bC^\mu)$ always has the same sign when one brings $C^\mu$ to the left of $\bar{C}^\nu$ since $C^\mu$ commutes twice (with $C^a$ and with $\bar{C}^b$).

The position of the anticommuting parameter in the ghost laws $\delta C^\alpha = -\frac{1}{2}f^\alpha{}_{\beta\gamma}C^\gamma\Lambda C^\beta$ (usual scheme) and $\delta C^\alpha = -\frac{1}{2}f^\alpha{}_{\beta\gamma}lC^\beta C^\gamma$ (NT scheme) is such that the statistics of the ghosts is the same as that of the structure constants in *either* case.

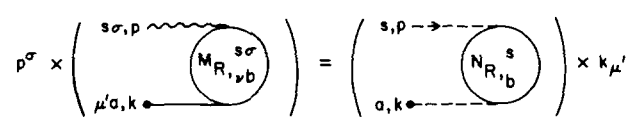


FIG. 4.

[1]J. Thierry-Mieg, CTP 68-733 and J. Math. Phys. (in press); Y. Ne'eman and J. Thierry-Mieg, Annals of Physics **123**, 2 (1979) 247, *idem* PNAS **77**, 2 (1980) 720, *idem*, Proceedings Second Marcel Grosman Meeting.
[2]G. Sterman, P. K. Townsend and P. van Nieuwenhuizen, Phys. Rev. D **17**, 1501 (1978).
[3]We use the notation of R. E. Kallosh, Nucl. Phys. B **141**, 141 (1978).
[4]C. Becchi, A. Rouet and R. Stora, Commun. Math. Phys. **42**, 127 (1975).
[5]The case that $\gamma^{\alpha\beta}$ depends on dynamical fields is under study by N. K. Nielsen and the author.
[6]B. Zumino, to be published. The different reality properties corresponding to the first two gradings are discussed here, also.
[7]M. K. Fung, P. van Nieuwenhuizen and D. R. T. Jones, Phys. Rev. D (to be published).
[8]For details of calculations in quantum supergravity, see a forthcoming "Physics Reports" by the author.
[9]The identity for $k^2 \neq 0$ can be used to analyze instantaneous Coulomb forces, (see E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. D (to be published).
[10]T. Kugo and I. Ojima, Phys. Lett. B **73**, 459 (1973); K. Nakanishi and M. Okawa, Progr. Theor. Phys. (Kyoto).

# Space spinors

Paul Sommers
*Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650*

(Received 18 July 1979; accepted for publication 15 October 1979)

Just as Maxwell's electromagnetic field equations govern the evolution of electric and magnetic spatial vectors if some choice of time function has been made, so also the neutrino equation and Dirac equation may be understood as governing the evolutions of certain spatial quantities. In this space-plus-time view of the spinor field equations, it is accurate and natural to regard a two-component spinor as the square root of a complex *spatial* null vector. The field equations are written in 3-plus -1 form for both the spinor fiedls and the corresponding null vector fields. A spatial null vector is of the form $\mathbf{M} = \mathbf{E} + i\mathbf{B}$, with $\mathbf{E}\cdot\mathbf{E} - \mathbf{B}\cdot\mathbf{B} = 0 = \mathbf{E}\cdot\mathbf{B}$, so it is also of the correct algebraic form for describing a null electromagnetic field. The time derivative of a squared neutrino field $M_a$, however, is $-i\,\mathrm{curl}M_a + \langle M\rangle^c D_a M_c$, compared with simply $-i\,\mathrm{curl}M_a$ for a source-free Maxwell field. Here $\langle M\rangle^c$ is the real spatial unit vector in the neutrino propagation direction $\mathbf{E}\times\mathbf{B}$, and $D_a$ is the spatial covariant derivative.

## 1. INTRODUCTION

The theory of electromagnetic fields was first understood as fields of electric and magnetic 3 vectors whose evolutions are governed by Maxwell's equations. The four dimensional relativistic formulation of Maxwell's equations does not invalidate the original understanding, but rather shows that the fields and their governing equations are of the same form no matter what choice is made for the decomposition of space–time into space and time. For other *integral*-spin relativistic fields and their field equations a similar understanding is readily available: It is a straightforward matter to interpret the equations as governing the evolutions of certain spatial tensor fields.

The purpose of this paper is to show that half-integral-spin field equations, such as the neutrino equation and the Dirac electron equation, are also evolution equations for *spatial* fields. In this picture a two-component spinor appears as a natural square root of a null complex 3 vector. Initial data for a neutrino field therefore consist of a complex null vector field tangent to the initial-data hypersurface. An additional bit of information must be supplied to specify the choice of sign in taking the square root. Alternatively, one can work directly with the squared spinor field, i.e., with the null vector field. This alternative is simpler in the sense that all relevant quantities are common tensor fields; it is awkward, however, inasmuch as the field equations, which are linear equations governing the spinor fields, become nonlinear equations for the squared fields.

The presentation in this paper is intended to be explicit enough that the results may be applied in numerical evolutions of space–times with spin-$\frac{1}{2}$ fields.

## 2. SPACE PLUS TIME

For the discussion of space spinors it will be helpful first to fix some notation and conventions.[1]

A time function $t$ on a space–time will be a smooth function whose hypersurfaces of constant value are spacelike. The future–pointing unit vector field normal to these constant-time hypersurfaces will be denoted by $n_a$. The lapse function $\alpha$ may be defined by $n_a = \alpha\nabla_a t$. The space–time metric $g_{ab}$ will have signature $(+\ -\ -\ -)$, so $n^a n_a = +1$; and note that the vector field $\alpha n^a$ is parametrized by $t$:

$$(\alpha n^a\nabla_a)t = (n^a)\alpha\nabla_a t = n^a n_a = 1. \tag{1}$$

The torsion-free covariant derivative defined by $g_{ab}$ will be denoted by $\nabla_a$.

The tensor $h_{ab} := g_{ab} - n_a n_b$ is the intrinsic metric tensor of the constant-time hypersurfaces. The contravariant form of the intrinsic metric is $h^{ab} = g^{ab} - n^a n^b$. (Here, and throughout, indices are raised using the space–time metric $g^{ab}$.) A *spatial* tensor is a tensor which has no nonzero contracted product with $n^a$ or $n_a$. The tensor $h_a{}^b = g_a{}^b - n_a n^b$ serves to project vectors or differential forms to spatial tensors, so arbitrary tensors can be projected to spatial tensors using products of this projection tensor. The spatial alternating tensor is given in terms of the space–time alternating tensor by $\epsilon_{bcd} = n^a\epsilon_{abcd}$.

For spatial tensors, the torsion-free covariant derivative $D_a$ arising from the metric $h_{ab}$ has a simple operational definition: Given a spatial tensor (for example, $S_b{}^{cd}$), its space-covariant derivative $(D_a S_b{}^{cd})$ is obtained by taking its space–time derivative $(\nabla_a S_b{}^{cd})$ and then projecting the resulting tensor:

$$D_a S_b{}^{cd} = h_a{}^k h_b{}^l h_m{}^c h_n{}^d \nabla_k S_l{}^{mn}. \tag{2}$$

Any tensor can be written in terms of purely spatial tensors together with the unit normal $n^a$ (or $n_a$). For example,

$$\begin{aligned}\nabla_{[a}n_{b]} &= \nabla_{[a}(\alpha\nabla_{b]}t) = (\nabla_{[a}\alpha)\nabla_{b]}t \\ &= (\nabla_{[a}\ln\alpha)n_{b]} = -n_{[a}\nabla_{b]}\ln\alpha \\ &= -n_{[a}(D_{b]}\ln\alpha + n_{b]}n^c\nabla_c\ln\alpha) \\ &= -n_{[a}D_{b]}\ln\alpha.\end{aligned} \tag{3}$$

(Here the square brackets mean skew symmetrization. Round brackets around indices will denote symmetrization.) Note also that $n^b\nabla_a n_b = 0$, since $n^b n_b = 1$. Moreover,

$$n^a\nabla_a n_b = n^a\nabla_b n_a + 2n^a\nabla_{[a}n_{b]} = -D_b\ln\alpha; \tag{4}$$

 0022-2488/80/102567-05$1.00  

and from these results, it easily follows that

$$\nabla_a n_b = -K_{ab} - n_a D_b \ln\alpha, \tag{5}$$

where $K_{ab}$ is a symmetric spatial tensor field. This last equation frequently comes in handy.

## 3. SPACE SPINOR ALGEBRA

A space spinor will be denoted as an ordinary two-component spinor $\mu_A$. In fact, it carries the same information as an ordinary spinor and may be identified with an ordinary spinor. The only difference is in the way the spinor is related to space–time geometry. The description of a space spinor utilizes some (arbitrary) time function, and thereby characterizes the spinor (up to sign) by way of exclusively spatial tensors. The standard picture[2] of (the square of) a spinor consists of a real null vector (flagpole) and a 2 plane (flag) spanned by the null vector and some orthogonal spacelike vector. This is equivalent to representing the square of the spinor $\mu_A$ by a self-dual null bivector, and one writes $M^{+}_{ab} = \mu_A\mu_B\epsilon_{B'A'}$. Any such self-dual null bivector is of the form $M^{+}_{ab} = M_{[a}l_{b]}$, where $l_a$ is a future-pointing real null vector and $M_a$ is a complex null vector orthogonal to $l_a$. If a time function has been defined, then $l_a$ and $M_a$ may be fixed uniquely by requiring that $h_a{}^b l_b$ be a *unit* vector and $M_a$ be a spatial vector. The complex spatial null vector $M_a$ is in fact equivalent to the self-dual null bivector $M^{+}_{ab}$, as follows. Given $M^a = E^a + iB^a$ (with $E^a$ and $B^a$ real), let $\langle M\rangle^a$ be the real unit spatial vector in the direction of $\mathbf{E}\times\mathbf{B}$ (i.e., in the direction of $\epsilon^a{}_{bc}E^bB^c$). Let $l^a$ be the future-pointing real null vector whose spatial projection is $\langle M\rangle^a$. Then $M^{+}_{ab} = M_{[a}l_{b]}$ is the self-dual null bivector corresponding to $M_a$. By virtue of a specified time function, the standard geometrical picture of a spinor therefore corresponds to a *spatial null vector.*

A phase transformation $\mu_A \mapsto e^{i\theta}\mu_A$ causes the transformation $M_a \mapsto e^{2i\theta}M_a$. The plane of $M^a$ (i.e, the subspace of the tangent space spanned by the real and imaginary parts of $M^a$) is invariant under this phase rotation, and so is the orthogonal null direction $l^a$. The rotation occurs within the plane of $M^a$. A rotation of $\pi$ radians in the phase of $\mu_A$ causes $E^a$ and $B^a$ to make a full rotation of $2\pi$ rad in this plane, so $\mu_A$ and $-\mu_A$ correspond to the same null vector $M^a$.

The correspondence between *null* spatial vectors and *null* self-dual bivectors suggests a further correspondence. Since a basis for complex spatial vectors can be selected from the null spatial vectors and since the corresponding null self-dual bivectors span the 3-complex-dimensional vector space of all self-dual bivectors, there is an identification between complex spatial vectors and self-dual bivectors (which is basis independent). This yields a natural isomorphism between the space of complex spatial vectors and the space of symmetric spinors $\theta_{AB} = \theta_{(AB)}$, since a self-dual bivector is of the form $\theta_{AB}\epsilon_{A'B'}$.

There is actually a more direct route to this isomorphism. The direct route exploits the fact that the unit normal vector $n^{AA'}$ defines an isomorphism from the space of conjugate (primed) spinors to the space of (unprimed) spinors:

$$\alpha_{A'} \mapsto \sqrt{2}n_B{}^{A'}\alpha_{A'}. \tag{6}$$

The inverse map is given by

$$\alpha_B \mapsto -\sqrt{2}n^{B}{}_{A'}\alpha_B, \tag{7}$$

since

$$-2n^B{}_{C'}n_B{}^{A'} = \epsilon_{C'}{}^{A'} = \delta^{A'}_{C'}.$$

The complex four-dimensional tangent space consists of vectors $V^{AA'}$ which can be mapped, via $n^a$, to unprimed valence-2 spinors:

$$V^{AB} := V^A{}_{A'}(\sqrt{2}n^{BA'}). \tag{8}$$

The vector $n^{AA'}$ itself becomes

$$n^{AB} = n^A{}_{A'}(\sqrt{2}n^{BA'}) = \frac{1}{\sqrt{2}}\epsilon^{AB}. \tag{9}$$

The spatial vectors are those vectors orthogonal to $n^a$, and hence the vectors $S_{AB}$ such that $S_{AB}\epsilon^{AB} = 0$. The spatial vectors are therefore precisely the symmetric spinors. The decomposition of a complex 4 vector into spatial and space-orthogonal components coincides with the decomposition of its valence-2 spinor into a symmetric spinor together with a scalar multiple of $\epsilon_{AB}$.

Let $\{o_A, \iota_A\}$ be a normalized spinor basis ($o_A\iota^A = 1$) chosen so that $n^{AA'} = (1/\sqrt{2})(o^A\bar{o}^{A'} + \iota^A\bar{\iota}^{A'})$. Then the isomorphism between primed spin space and unprimed spin space is given by the action on basis spinors:

$$\bar{o}_{A'} \mapsto \iota_A, \bar{\iota}_{A'} \mapsto -o_A, o_A \mapsto -\bar{\iota}_{A'}, \iota_A \mapsto \bar{o}_{A'}. \tag{10}$$

The spin space automorphisms ($o_A \mapsto \hat{o}_A$, $\iota_A \mapsto \hat{\iota}_A$) such that $\hat{o}_A\hat{\iota}^A = 1$ are a representation of the group SL(2,$C$). The subgroup for which $\hat{o}^A\bar{\hat{o}}^{A'} + \hat{\iota}^A\bar{\hat{\iota}}^{A'} = (1/\sqrt{2})n^{AA'}$ is SU(2). This is the subgroup of linear transformations preserving the form (10) of the isomorphism (6) and (7).

By exploiting a given time function it is now possible to dispense with primed spinors altogether. Any primed index automatically converts to an unprimed spinor index via its contracted product with $\sqrt{2}n^{AA'}$. A space–time tensor index, which is normally identified with a pair of spinor indices of which one is primed,[2] will here be identified with a pair of unprimed indices. The following equalitities are then sensible:

$$\begin{aligned} V_a = V_{AB} &= V_{(AB)} + V_{[AB]} = V_{(AB)} + \tfrac{1}{2}V_{CD}\epsilon^{CD}\epsilon_{AB} \\ &= V_{(AB)} + (V_c n^c)n_{AB} = V_{(AB)} + (V_c n^c)n_a \\ &= h_a{}^c V_c + (V_c n^c)n_a. \end{aligned} \tag{11}$$

In order to minimize confusion, tensor indices may be chosen from the odd-numbered letters of the alphabet. For example, $T_a{}^c{}_g = T_{AB}{}^{CD}{}_{GH}$.

For vectors which are not purely spatial, attention must be paid to the order of the indices. A spinor $V^{AB}$ could define two different vectors, depending on which index were converted to a primed index. According to the definition (8), a spinor $V^{AB}$ is identified with a vector $V^{AA'}$ by converting the *second* index to a primed index. The other vector $W^c := -\sqrt{2}n^{AC'}V_A{}^C$ would be related to $V^c$ by time reflection: $W^c = V^c - 2(n_a V^a)n^c$.

The index clumping rules apply also to higher valence

tensors. For example, the space–time metric $g_{ac}$ is

$$\begin{aligned} g_{ABCD} &:= (\sqrt{2}n_B{}^{A'})\,(\sqrt{2}n_D{}^{C'})g_{AA'CC'} \\ &= 2n_B{}^{A'}n_D{}^{C'}\epsilon_{AC}\epsilon_{A'C'} \\ &= \epsilon_{AC}\epsilon_{BD}. \end{aligned} \tag{12}$$

The spatial metric $h_{ac}$ is

$$\begin{aligned} h_{ABCD} &= g_{ABCD} - n_{AB}n_{CD} \\ &= \epsilon_{AC}\epsilon_{BD} - \tfrac{1}{2}\epsilon_{AB}\epsilon_{CD} \\ &= -\epsilon_{A(C}\epsilon_{D)B}. \end{aligned} \tag{13}$$

## 4. DERIVATIVES

The transcription of spinor field equations from their four dimensional spacetime form to a space-plus-time form requires the conversion of the spinor covariant derivative $\nabla_{AA'}$ to derivatives which can be evaluated with reference only to intrinsic spatial geometry plus a suitable time derivative.

A convenient choice for the spatial covariant derivative is the following:

$$D_{AB}\mu_C := \sqrt{2}n_{(B}{}^{A'}\nabla_{A)A'}\mu_C - \frac{1}{\sqrt{2}}K_{ABCD}\mu^D. \tag{14}$$

Here $K_{ABCD} := K_{ac}$ is the extrinsic curvature defined by Eq. (5). The derivative definition extends to spinors of arbitrary valence in accordance with the usual Leibnitz rule for the derivative of a product. The spinors $\epsilon_{CD}$ and $\epsilon^{CD}$ are $D_{AB}$ constant since they are $\nabla_{AA'}$ constant and $K_{ABCD} = K_{(AB)(CD)}$. Acting on functions, $D_{AB}$ is the spatial gradient: $D_{AB}f = h_a{}^c\nabla_c f$. Acting on a spatial vector $S_c$, the derivative $D_{AB}S_{CD}$ agrees with $D_aS_c$ as defined in Sec. 2. This can be checked by verifying $2n^B{}_{A'}n^D{}_{C'}D_{AB}S_{CD} = D_{AA'}S_{CC'}$, using $S_{CC'} = -\sqrt{2}n^D{}_{C'}S_{CD}$ and Eq. (5).

Although the definition (14) makes the spatial covariant derivative appear to be computationally complicated, it is actually simple. In specific calculations spinors are customarily referred to a normalized basis. If $\{o_A, \iota_A\}$ is such a basis ($o_A\iota^A = 1$), let

$$m_a := -o_Ao_B,\quad \bar{m}_a := \iota_A\iota_B,\quad z_a := \sqrt{2}o_{(A}\iota_{B)}. \tag{15}$$

Note that $m_a$ and $\bar{m}_a$ are the complex null vectors associated via Eq. (10) with $o_A\bar{\iota}_{A'}$, and $\iota_A\bar{o}_{A'}$, respectively, and $z^a$ is the real unit vector $z^a = i\epsilon^a{}_{bc}m^b\bar{m}^c$ ($\sqrt{-1}$ times the cross product of the two null vectors). Now let $S^a$ be any real spatial vector. To compute $D_So_C$ and $D_S\iota_C$ (i.e., $S^{AB}D_{AB}o_C$ and $S^{AB}D_{AB}\iota_C$) one need only compute the tensorial "rotation coefficients" $z^aD_Sm_a$ and $\bar{m}^aD_Sm_a$, where $D_Sm_a := S^bD_bm_a$. The $D_{AB}$ derivative is then given by the formulas

$$D_So_C = -\frac{1}{2}(\bar{m}^aD_Sm_a)o_C + \frac{1}{\sqrt{2}}(z^aD_Sm_a)\iota_C \tag{16}$$

and

$$D_S\iota_C = -\frac{1}{\sqrt{2}}(z^aD_S\bar{m}_a)o_C + \frac{1}{2}(\bar{m}^aD_Sm_a)\iota_C, \tag{17}$$

which are easily verified by transvecting with the basis spinors and using Eq. (15).

A convenient choice for the time derivative of a spinor field $\mu_A$ is

$$D_n\mu_A := n^{CC'}\nabla_{CC'}\mu_A - \frac{1}{\sqrt{2}}\mu^BD_{AB}\ln\alpha, \tag{18}$$

where $\alpha$ is the lapse function. The action of $D_n$ on spinors of higher valence is obtained by the Leibnitz rule. Note that $D_n\epsilon_{AB} = 0$, and $D_nf = n^a\nabla_a f$ for any function $f$. Like the spatial covariant derivative, $D_n$ has the property that its action on a spatial vector $S_a = S_{AB}$ is simple: $D_nS_a = h_a{}^bn^c\nabla_cS_b$. It is just the spatial projection of the space–time-covariant derivative with respect to the vector $n^a$. It is therefore easy to compute time derivatives by reference to a normalized spinor basis. The time derivatives of the basis spinors are given by Eqs. (16) and (17) with $D_S$ replaced by $D_n$ throughout.

In 3-plus-1 formalisms it is usually the Lie derivative $\mathcal{L}_{\alpha n}$ of spatial tensor fields which is most applicable as a "time derivative." For action on tensor fields, the derivative $D_n$ can easily be converted to $\mathcal{L}_{\alpha n}$ by insertion of known (nonderivative) terms. For example, using Eq. (5) and the formula for the Lie derivative of a one form, it is easy to check that

$$\mathcal{L}_{\alpha n}S_a = \alpha(D_nS_a - K_{ac}S^c). \tag{19}$$

Combining $D_n$ and $D_{AB}$ provides the spinorial 3-plus-1 transcription of $\nabla_{AA'}\mu_C$ (or $\nabla_{AB}\mu_C \equiv \sqrt{2}n_B{}^{A'}\nabla_{AA'}\mu_C$):

$$\begin{aligned} \nabla_{AB}\mu_C &= \nabla_{(AB)}\mu_C + \nabla_{[AB]}\mu_C \\ &= D_{AB}\mu_C + \frac{1}{\sqrt{2}}K_{ABCD}\mu^D \\ &\quad + \epsilon_{AB}\left(\frac{1}{\sqrt{2}}D_n\mu_C + \frac{1}{2}\mu^DD_{CD}\ln\alpha\right). \end{aligned} \tag{20}$$

## 5. FIELD EQUATIONS IN MINKOWSKI SPACE WITH CARTESIAN TIME

In Minkowski space there are available the Cartesian time functions for which $K_{ab} = 0$ and $\alpha = 1$. In that case the derivative formulas (14) and (18) simplify, and the structure of field equations is surveyable. As a first example, consider a neutrino field $\mu_A$ satisfying the field equation $\nabla_{AA'}\mu^A = 0$. This is equivalent to $\nabla_{AB}\mu^A = 0$ or, using Eq. (20),

$$D_n\mu_B = -\sqrt{2}D_{AB}\mu^A. \tag{21}$$

This last equation gives the time derivative of $\mu_B$ in terms of its spatial derivatives.

In view of the fact that a spinor is the square root of a spatial null vector, it is also appropriate to examine the evolution of the null vector field $M_a \equiv -\mu_A\mu_B$. Since $D_nM_a = -2\mu_{(A}D_n\mu_{B)}$, the field equation (21) gives

$$D_nM_a = 2\sqrt{2}\mu_{(A}D_{B)C}\mu^C. \tag{22}$$

The identity

$$\mu^CD_{AB}\mu_C = \mu^CD_{C(A}\mu_{B)} - \mu_{(A}D_{B)C}\mu^C \tag{23}$$

allows one to see that

$$2\mu_{(A}D_{B)C}\mu^C = D_{C(A}\mu_{B)}\mu^C - \mu^CD_{AB}\mu_C. \tag{24}$$

Making this substitution yields

$$D_n M_a = \sqrt{2} D_{C(A}\mu_{B)}\mu^C - \sqrt{2}\mu^C D_{AB}\mu_C. \tag{25}$$

Now, it is not hard to verify that the cross product[3] $-\epsilon_a{}^{bc}u_b v_c$ of spatial one forms $u_a$ and $v_a$ is given by $-i\sqrt{2}u_{C(A}v_{B)}{}^C$, and so curl$M_a$, i.e., $-\epsilon_a{}^{bc}D_b M_c$, is $-i\sqrt{2}D_{C(A}M^C_{B)}$, or $+i\sqrt{2}D_{C(A}\mu_{B)}\mu^C$. Therefore,

$$D_n M_a = -i\,\mathrm{curl}M_a - \sqrt{2}\mu^C D_{AB}\mu_C. \tag{26}$$

The final term on the right can be usefully re-expressed with the following notational convention: Any nonzero spatial null vector $\mathbf{M} = \mathbf{E} + i\mathbf{B}$ (with $\mathbf{E}\cdot\mathbf{E} = \mathbf{B}\cdot\mathbf{B}$ and $\mathbf{E}\cdot\mathbf{B} = 0$) defines the plane of $\mathbf{E}$ and $\mathbf{B}$ and, in particular, a unit spatial vector perpendicular to the plane in the direction of $\mathbf{E}\times\mathbf{B}$. Let $\langle M\rangle^a$ denote this real unit vector, so

$$\langle M\rangle^a := -i[M_d\bar{M}^d]^{-1}\epsilon^a{}_{bc}M^b\bar{M}^c. \tag{27}$$

Then the equation $\sqrt{2}\mu^C D_{AB}\mu_C = -\langle M\rangle^c D_a M_c$ is valid, as may be verified by introducing a spinor field $\lambda_A$ satisfying $\mu_A\lambda^A = 1$ and $\lambda^A n_A{}^{A'}\bar{\mu}_{A'} = 0$, so that $\langle M\rangle_c = \sqrt{2}\mu_{(C}\lambda_{D)}$. The "squared neutrino equation" is therefore

$$D_n M_a = -i\,\mathrm{curl}M_a + \langle M\rangle^c D_a M_c. \tag{28}$$

The unit vector $\langle M\rangle^a$ is in the direction of propagation of the neutrino field, it being aligned with the spatial projection of the null vector $\mu_A\bar{\mu}_{A'}$.

The neutrino can thus be pictured as a complex vector $\mathbf{E} + i\mathbf{B}$ (with $\mathbf{E}\cdot\mathbf{E} - \mathbf{B}\cdot\mathbf{B} = 0 = \mathbf{E}\cdot\mathbf{B}$) propagating in the direction of $\mathbf{E}\times\mathbf{B}$, in complete analogy with a null electromagnetic field. A null source-free Maxwell field may be described as a (divergence-free) spatial null vector $M_a$ which satisfies $D_n M_a = -i\,\mathrm{curl}M_a$. It is the nonlinear second term on the right side of Eq. (28) which distinguishes the squared neutrino equation from Maxwell's evolution equation.

The Dirac equation can be manipulated in much the same way. It is customary[4] to introduce a *pair* of spinor fields $\xi^A$ and $\eta_{A'}$; the Dirac equation is then the pair of equations

$$\nabla_{AA'}\xi^A = \frac{m}{\sqrt{2}}\eta_{A'} \text{ and } \nabla^{AA'}\eta_{A'} = -\frac{m}{\sqrt{2}}\xi^A, \tag{29}$$

where $m$ is the electron mass. With $\eta_B := \sqrt{2}n_B{}^{A'}\eta_{A'}$, these equations become

$$\nabla_{AB}\xi^A = \frac{m}{\sqrt{2}}\eta_B \text{ and } \nabla^{AB}\eta_B = -\frac{m}{\sqrt{2}}\xi^A. \tag{30}$$

Using Eq. (20), these yield

$$D_n\xi_B = -\sqrt{2}D_{AB}\xi^A + m\eta_B \text{ and}$$

$$D_n\eta_A = +\sqrt{2}D_{AB}\eta^B - m\xi_A. \tag{31}$$

Each spinor field satisfies an evolution equation similar to the neutrino equation but with the other spinor field serving as a driving term. One can picture the two spatial null vectors propagating and interacting. When the equations for the null vectors are written out, they are similar to the squared neutrino equation except that the equations for the Dirac fields each include an extra driving term proportional to $\xi_{(A}\eta_{B)}$. Explicitly, if $X_{AB} = -\xi_A\xi_B$ is the one null vector and $Y_{AB} = \eta_A\eta_B$ the other, then the coupling terms are proportional to $\mathbf{X}\times\mathbf{Y}$:

$$D_n X_a = -i\,\mathrm{curl}X_a + \langle X\rangle^c D_a X_c - i\sqrt{2}m(-X_d Y^d)^{-1/2}\epsilon_a{}^{bc}X_b Y_c, \tag{32}$$

$$D_n Y_a = i\,\mathrm{curl}Y_a - \langle Y\rangle^c D_a Y_c - i\sqrt{2}m(-X_d Y^d)^{-1/2}\epsilon_a{}^{bc}X_b Y_c. \tag{33}$$

By writing the Dirac equation tensorially in this way, there are several apparent sign ambiguities. If null vector fields $X_a$ and $Y_a$ are specified on an initial hypersurface, there are four corresponding pairs of spinor fields on the hypersurface since there are two possibilities for $\xi_A$ and two for $\eta_A$. After evolving the null vectors, however, there is only an overall sign ambiguity (i.e., freedom to change the sign of $\xi_A$ and $\eta_A$ together), because evolving the null vectors requires making a choice for the sign of $(-X_d Y^d)^{1/2} \equiv \xi_A\eta^A$.

## 6. FIELD EQUATIONS IN CURVED SPACETIMES WITH ARBITRARY TIME FUNCTIONS

In a generic space–time it is not possible to choose a time function for which $K_{ab} = 0$, and it sometimes inconvenient to choose it so that $\alpha = 1$. The neutrino equation and Dirac equations, when written in space-plus-time form, will then include additional terms containing the extrinsic curvature and lapse function gradient.

The linear neutrino equation $\nabla_{AA'}\mu^A = 0$ may be written, using Eq. (20), this way:

$$\epsilon^{AC}\left\{D_{AB}\mu_C + \frac{1}{\sqrt{2}}K_{ABCD}\mu^D + \epsilon_{AB}\left(\frac{1}{\sqrt{2}}D_n\mu_C + \frac{1}{2}\mu^D D_{CD}\ln\alpha\right)\right\} = 0. \tag{34}$$

From the symmetries $K_{ABCD} = K_{(AB)(CD)} = K_{CDAB}$ together with Eq. (13), one finds that $\epsilon^{AC}K_{ABCD} = \frac{1}{2}K\epsilon_{BD}$, where $K := K_a{}^a$. The neutrino equation is therefore

$$D_n\mu_B = -\sqrt{2}D_{AB}\mu^A + \frac{1}{2}K\mu_B - \frac{1}{\sqrt{2}}\mu^A D_{AB}\ln\alpha. \tag{35}$$

As in the previous section, this equation yields a squared neutrino equation governing the evolution of $M_a \equiv -\mu_A\mu_B$:

$$D_n M_a = -i\,\mathrm{curl}M_a + \langle M\rangle^c D_a M_c + KM_a - i\epsilon_a{}^{bc}M_b D_c\ln\alpha. \tag{36}$$

The Dirac equations (31) in this more general context become

$$D_n\xi_B = -\sqrt{2}D_{AB}\xi^A + \frac{1}{2}K\xi_B - \frac{1}{\sqrt{2}}\xi^A D_{AB}\ln\alpha + m\eta_B \tag{37}$$

and

$$D_n\eta_A = \sqrt{2}D_{AB}\eta^B + \frac{1}{2}K\eta_A + \frac{1}{\sqrt{2}}\eta^B D_{AB}\ln\alpha - m\xi_A. \tag{38}$$

(Caution: To arrive at this second equation, one evaluates $\nabla^{AA'}(-\sqrt{2}n^B_{A'}\eta_B) = \nabla^{AB}\eta_B - \sqrt{2}\eta_B\nabla^{AA'}n^B{}_{A'}$; that is to say, $\nabla^{AA'}\eta_{A'}$ is not simply $\nabla^{AB}\eta_B$. Also, note that $\nabla^{AB}\eta_B$ is not the same as $\nabla^{BA}\eta_B$.)

For the null vectors, $X_a = -\xi_A\xi_B$ and $Y_a = \eta_A\eta_B$, the evolution equations are

$$D_n X_a = -i\,\mathrm{curl}X_a + \langle X\rangle^c D_a X_c + KX_a - i\epsilon_a{}^{bc}X_b D_c \ln\alpha - i\sqrt{2}m(-X_d Y^d)^{-1/2}\epsilon_a{}^{bc}X_b Y_c \tag{39}$$

and

$$D_n Y_a = i\,\mathrm{curl}Y_a - \langle Y\rangle^c D_a Y_c + KY_a + i\epsilon_a{}^{bc}Y_b D_c \ln\alpha - i\sqrt{2}m(-X_d Y^d)^{-1/2}\epsilon_a{}^{bc}X_b Y_c. \tag{40}$$

## 7. DISCUSSION

The objective of the foregoing formulations is to expose the manner in which the neutrino and Dirac equations govern the evolution of spatial fields. The fields are spinor fields, but a spinor may be regarded as the square root of a *spatial* complex null vector. The only ambiguity in this picture of a spinor field is the overall sign: A null vector field has two square roots. The sign of a spinor field has no geometric manifestation. It can be determined only by comparison with other spinor fields. If *all* spinor fields underwent a phase change of $\pi$ rad, there would be no physically obvervable consequence. In practice, the sign of a spinor is fixed relative to some reference spinor. For example, if $o_A$ is a reference spinor, then the sign of $\mu_A$ is fixed by the direction of the complex vector $o_{(A}\mu_{B)}$. Reversing the sign of $\mu_B$ reverses the sign of this complex vector (provided $o_A$ is unchanged).

The field equations for spinor fields allow one to determine the fields at later (or earlier) times, given appropriate initial data on one hypersurface of constant time. For a neutrino field, the appropriate data are a spinor field (without constraints) on the hypersurface. Given a null vector field $M_a$ on one hypersurface, therefore, the field equation (36) serves to define the null vector field (and hence its square roots) all over space–time. For a Dirac field, initial data are *two* (unconstrained) null vector fields $X_a$ and $Y_a$ on the initial hypersurface. Their evolutions are governed by the field equations (39) and (40).

Of course, the physical fields are the spinor fields, not the null vector fields. Working with the square of the spinor field not only introduces a sign ambiguity, but also causes the linear spinor field equations to become nonlinear. It is usually desirable to work with the linear field equations (35) and (37) and (38) which govern the evolutions of spinor fields themselves. For such calculations it is convenient to employ a normalized basis $\{o_A, \iota_A : o_A\iota^A = 1\}$ related to spatial geometry by Eq. (15). It is possible to choose such a basis to satisfy $D_n o_A = 0 = D_n \iota_A$. Using a basis of this type, an arbitrary spinor $\mu_A = ao_A + b\iota_A$ is specified by a pair of complex numbers $a$ and $b$. The evolution of $\mu_A$ is determined by the time derivatives of $a$ and $b$; these derivatives are easily read out from equations like Eq. (35), using Eq. (16) and (17).

The 3-plus-1 interpretation of spinor field equations presents a couple of difficulties which the foregoing discussions are intended to circumvent. A spinor is frequently regarded as a set of components which are referred to a space–time tetrad. As such, it is not evident how to specify a spinor using only intrinsic spatial geometry. Instead, it is appropriate to think of a spinor basis as defining a spatial triad (15), or to regard a spinor as a square root of a spatial null vector. Another difficulty arises if one attempts to express the time derivative of a spinor field as a Lie derivative. Spinors are intimately related to space–time conformal geometry, and there is no satisfactory way to define a spinor Lie derivative except along a conformal Killing vector field. This difficulty is manifested in the fact that a null vector $M_a$ does not generally remain null when Lie transported. A derivative such as $D_n$, however, does provide a useful notion of the time derivative of a spinor field. Using equations like Eq. (19), $D_n$ derivatives of tensor fields can always be converted to convenient Lie derivatives.

[1]The notation here is comparable with that adopted in the following three references, which contain more discussion: R. Geroch, "General Relativity," lecture notes at the University of Chicago (1972) (duplicated but unpublished); L. Smarr, "The Structure of General Relativity with a Numerical Illustration: The Collision of Two Black Holes," Ph.D. dissertation, University of Texas (1974); L. Smarr, and J.W. York, Jr., Phys. Rev. D **17**, 2529 (1978).

[2]R. Penrose, *The Structure of Spacetime*, Battelle Rencontres, edited by C. Dewitt, and J.A. Wheeler (Benjamin, New York, 1968).

[3]The cross product of vectors $u^a$ and $v^a$ is $\epsilon^a{}_{bc}u^b v^c$. Using the metric to convert this to a one form gives $\epsilon_{abc}u^b v^c$, which is the *negative* of the cross product $(-\epsilon_a{}^{bc}u_b v_c)$ of the one forms $u_a$ and $v_a$. The difference in sign arises from the use here of a *negative* definite spatial metric.

[4]W.A. Bade, and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

# The Dirac inverse spectral transform: Kinks and boomerons [a)]

Jérôme JP. Leon [b)]

*Laboratoire de Physique Théorique, Université Mohamed V, Faculté des Sciences, Rabat-Maroc and Laboratoire de Physique Mathématique, Université des Sciences et Techniques du Languedoc, Montpellier cedex, France 34060*

(Received 9 August 1979; accepted for publication 19 October 1979)

The inverse spectral transform (IST) is derived when using the eigenvalue problem for the one-dimensional Dirac operator:
$(D) = i\sigma_3(d/dx) + i\begin{pmatrix} 0 & -\tilde{q} \\ \tilde{r} & 0 \end{pmatrix}$, $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,
where the potentials $\tilde{q}$ and $\tilde{r}$ have nonzero asymptotic values. The method used is of AKNS type. It is shown that the nonlinear evolution equations (NEE) obtained are of differential type at any order (and not of integro-differential type). Some particular solutions are studied, and it is shown that their special behavior is a direct consequence of the nonzero boundary condition on $(D)$.

## I. INTRODUCTION

This paper is devoted to the derivation of a new class of nonlinear evolution equations (NEE) whose particular feature gives rise, in certain cases, to "kinklike" solutions[1] and "boomerons" solutions,[2] that is to say, solitons that come back.

Some of the results have been previously reported in Ref. 3 and performed independently by Gerdjikov and Kulish in Ref. 4. In Ref. 3 the basic method is of the AKNS type[5] for the one-dimensional Dirac system:

$$\left\{ i\sigma_3 \frac{d}{dx} - iq_3\sigma_1 - q_1\sigma_2 + m\sigma_2 \right\} Y = \mathscr{E} Y, \tag{I.1}$$

that is to say, for a Zakharov–Shabat system[6] in which the potentials $\tilde{q} = -q_1 - iq_3 + m$ and $\tilde{r} = -q_1 + iq_3 + m$ go asymptotically to the same real constant $m$. In a different way, it is shown in Ref. 4 that the Zakharov–Shabat system whose potentials $\tilde{q}$ and $\tilde{r}$ verifies the asymptotic behavior

$$\tilde{r}\cdot\tilde{q} \underset{|x|\to\infty}{\to} m^2, \quad m^2 \in \mathscr{C} \tag{I.2}$$

possesses a complete class of Hamiltonian systems.

Since only the final results of the inverse problem are workful for I.S.T., we shall not develop its procedure which is already well known.[7] For a detailed treatment, the reader will refer itself to Ref. 4 or to unpublished work.[8] We prefer to focus our study at first on the derivation of the class of solvable NEE through the AKNS method with the asymptotic condition (I.2) on the potentials. Secondly, an important point which is often neglected is that the NEE obtained are of differential type even if the generator operator is of the integrodifferential type. This statement will be proved for the class of NEE obtained here (for the Z.S. case, see Ref. 9), and the structure of the demonstration surely hold for other classes of NEE. We shall thirdly study some special solutions, namely, kinks and boomerons, which correspond to an arbitrary discrete spectrum of the spectral data. It will be shown that the emergence of these special solutions is entirely owing to the nonzero asymptotic values of the potentials.

## II. THE INVERSE PROBLEM

We consider the one dimensional Dirac equation[10]

$$D: \left\{ i\sigma_3 \frac{d}{dx} + i\begin{pmatrix} 0 & -\tilde{q} \\ \tilde{r} & 0 \end{pmatrix} \right\} Y = \mathscr{E} Y, \tag{II.1}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

in which the potentials $\tilde{q}$ and $\tilde{r}$ verify the asymptotic boundary conditions

$$\tilde{r} \underset{\pm\infty}{\to} r^{\pm}, \quad \tilde{q} \underset{\pm\infty}{\to} q^{\pm}, \quad r^{+}q^{+} = r^{-}q^{-} = m^2, \quad m^2 \in \mathbb{C}. \tag{II.2}$$

In spite of energy $\mathscr{E}$ we better use the complex momentum $k$ defined by $k^2 = \mathscr{E}^2 - m^2$. Then the eigenvalue $\mathscr{E}$ appears as a double-valued function of $k$ which is represented by the two determinations $\mathscr{E} = \epsilon E$, $\epsilon = \pm$, where

$$E(k) = |k^2 + m^2|^{1/2} \times \exp\frac{i}{2}\{\mathrm{Arg}(k + im) + \mathrm{Arg}(k - im)\},$$

$$-\frac{\pi}{2} + \omega \leqslant \mathrm{Arg}(k + im) < \frac{3\pi}{2} + \omega,$$

$$\frac{-3\pi}{2} + \omega < \mathrm{Arg}(k - im) \leqslant \frac{\pi}{2} + \omega, \quad \omega = \mathrm{Arg}(m). \tag{II.3}$$

Thus, $D$ will represent two systems whose solutions are the eigenfunctions $Y^{\epsilon}(k,x)$ with the eigenvalues $\epsilon E$ ($\epsilon = \pm$).

The right and left Jost solutions of $D$ are defined by the asymptotic behaviors

$$\text{as } x \to +\infty: \begin{cases} \psi^{\epsilon}(k,x) \sim \psi_0^{\epsilon}(k,x) = \left(\dfrac{\epsilon E + k}{2k}\right)^{1/2} \begin{pmatrix} -iq^{+} \\ \epsilon E + k \\ 1 \end{pmatrix} e^{ikx}, \\ \bar{\psi}^{\epsilon}(k,x) \sim \bar{\psi}_0^{\epsilon}(k,x) = \left(\dfrac{\epsilon E + k}{2k}\right)^{1/2} \begin{pmatrix} 1 \\ ir^{+} \\ \epsilon E + k \end{pmatrix} e^{-ikx}, \end{cases} \tag{II.4}$$

[a)] This work has been done as part of the program "Recherche cooperative sur Programme no. 264: Etude interdisciplinaire des problèmes inverses."

[b)] Physique Mathématique et théorique, Equipe de recherche associée au C.N.R.S. no. 154.

 0022-2488/80/102572-07$1.00  

$$\text{as } x\to-\infty:\begin{cases}\phi^{\epsilon}(k,x)\sim\phi_0^{\epsilon}(k,x)=\left(\dfrac{\epsilon E+k}{2k}\right)^{1/2}\begin{pmatrix}1\\ \dfrac{ir^-}{\epsilon E+k}\end{pmatrix}e^{-ikx},\\ \tilde\phi^{\epsilon}(k,x)\sim\tilde\phi_0^{\epsilon}(k,x)=\left(\dfrac{\epsilon E+k}{2k}\right)^{1/2}\begin{pmatrix}\dfrac{iq^-}{\epsilon E+k}\\ -1\end{pmatrix}e^{ikx},\end{cases}\tag{II.5}$$

The functions $\psi^{\epsilon}(k,x)$ and $\phi^{\epsilon}(k,x)$ are analytic in $k$ in the upper cut half-plane $\{k/\mathrm{Im}(k)>0, k\notin[\mathrm{im},i\infty e^{i\omega}]\}$, continuous on the real axis, and continuous on the right on the cut. For real $k$, the reflexion coefficient to the right $R^{\epsilon}(k)$ and the transmission coefficient $T^{\epsilon}(k)$ are displayed by developing the solution $\phi^{\epsilon}(k,x)$ on the basic solutions $\psi^{\epsilon}(k,x)$ and $\bar\psi^{\epsilon}(k,x)$:

$$\phi^{\epsilon}(k,x)=\frac{1}{T^{\epsilon}(k)}\bar\psi^{\epsilon}(k,x)+\frac{R^{\epsilon}(k)}{T^{\epsilon}(k)}\psi^{\epsilon}(k,x),\tag{II.6}$$

$$\frac{1}{T^{\epsilon}(k)}=|\phi^{\epsilon}(k,x),\psi^{\epsilon}(k,x)|,\qquad \frac{R^{\epsilon}(k)}{T^{\epsilon}(k)}=|\bar\psi^{\epsilon}(k,x),\phi^{\epsilon}(k,x)|,\tag{II.7}$$

where the symbol $|f,g|$ denotes the determinant of the matrix made with column vectors $f$ and $g$. Starting with the analytic properties of $\phi$ and $\psi$, and formula (II.7), it can be shown that $1/T^{\epsilon}(k)$ is analytic in the upper cut half-plane. We shall assume throughout that the zeros $(k_n^{\epsilon}, n=1,\ldots,N^{\epsilon})$ of $1/T^{\epsilon}(k)$ are simple and not on the real axis. They correspond to the bound states to which we associate the coefficients

$$C_n^{\epsilon}=\frac{R^{\epsilon}(k)}{T^{\epsilon}(k)}\left\{\frac{d}{dk}T^{\epsilon}(k)^{-1}\right\}^{-1}\Big|_{k=k_n^{\epsilon}}.\tag{II.8}$$

The set of spectral data is then defined as

$$\mathcal{S}=\{R^{\epsilon}(k), k\in\mathbb{R}; k_n^{\epsilon}, C_n^{\epsilon}, (n=1,\ldots,N^{\epsilon}); \epsilon=\pm\}.\tag{II.9}$$

Conversely, $\mathcal{S}$, $m^2$, and $q^+$ being given, one can compute the potentials $\tilde q$ and $\tilde r$ via the so called inverse problem[8]:

$$[\sigma_3,K(x,x)]=\begin{pmatrix}0 & -(q-q^+)\\ (r-r^+) & 0\end{pmatrix},\quad r^+q^+=m^2,\tag{II.10}$$

where $K(x,y)$ is solution of the Marchenko equation

$$F(x+y)+K(x,y)+\int_x^{\infty}K(x,u)F(u+y)du=0,\quad y>x,\tag{II.11}$$

and $F(x+y)$ is given from the spectral data through

$$F(x,y)=\sum_{\epsilon}\left\{\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{kdk}{\epsilon E}R^{\epsilon}(k)\psi_0^{\epsilon}(k,x)^T\psi_0^{\epsilon}(k,y)\sigma_1-i\sum_n C_n^{\epsilon}\frac{k_n^{\epsilon}}{\epsilon E(k_n^{\epsilon})}\psi_0^{\epsilon}(k_n^{\epsilon},x)^T\psi_0^{\epsilon}(k_n^{\epsilon},y)\sigma_1\right\}.\tag{II.12}$$

The asymptotic behavior of $\tilde q$ and $\tilde r$ as $x$ goes to $-\infty$ is given from $q^+$ and $r^+$ by[4]

$$\log\frac{q^+}{r^-}=\log\frac{r^-}{r^+}=\sum_{\epsilon}\left\{\frac{i}{2\pi}\int_{-\infty}^{+\infty}\frac{dk}{\epsilon E}\times\log(1+|R^{\epsilon}(k)|^2)+\sum_1^{N^{\epsilon}}\log\frac{\epsilon E+k}{\epsilon E-k}\Big|_{k=k_n^{\epsilon}}\right\}.\tag{II.13}$$

## III. THE INVERSE SPECTRAL TRANSFORM

### A. Introduction and fundamental results

Assuming for $V={}^T(\tilde r,\tilde q)$ a $t$ dependance, we construct now a class of NEE solvable by I.S.T. for $D$.

We use the following notations: $F={}^T(F_1,F_2)$ and $G={}^T(G_1,G_2)$ being two functions of $\mathbb{R}$ in $\mathbb{C}^2$, we define the "scalar product" of $F$ and $G$ by

$$\langle F\,|\,G\rangle=\int_{-\infty}^{+\infty}(F(x)\,|\,G(x))dx=\int_{-\infty}^{+\infty}(F_1(x)G_1(x)+F_2(x)G_2(x))dx.\tag{III.1}$$

$Y^{\epsilon}={}^T(y_1^{\epsilon},y_2^{\epsilon})$ and $Z^{\epsilon}={}^T(z_1^{\epsilon},z_2^{\epsilon})$ being two solutions of $D$, we set $\Lambda^{\epsilon}={}^T(y_1^{\epsilon}z_1^{\epsilon},y_2^{\epsilon}z_2^{\epsilon})$. We shall use particularly

$$\Lambda_1^{\epsilon}(k,x)=T^{\epsilon}(k)^2\begin{pmatrix}\phi_1^{\epsilon}(k,x)^2\\ \phi_2^{\epsilon}(k,x)^2\end{pmatrix},\tag{III.2}$$

$$\Lambda_2^{\epsilon}(k,x)=T^{\epsilon}(k)^2\begin{pmatrix}\phi_1^{\epsilon}(k,x)\psi_1^{\epsilon}(k,x)\\ \phi_2^{\epsilon}(k,x)\psi_2^{\epsilon}(k,x)\end{pmatrix},\tag{III.3}$$

$$\Lambda_3^{\epsilon}(k_n^{\epsilon},x)=\left(\frac{d}{dk}\frac{1}{T^{\epsilon}(k)}\right)^{-1}\times\begin{pmatrix}\phi_1^{\epsilon}(k,x)\bar\psi_1^{\epsilon}(k,x)\\ \phi_2^{\epsilon}(k,x)\bar\psi_2^{\epsilon}(k,x)\end{pmatrix}\Big|_{k=k_n^{\epsilon}}.\tag{III.4}$$

The main tool of the inverse method consist at first of a set of relations between the scattering data and some "scalar products" from which arise $V(x,t)$ and $V_t(x,t)$ (the subscripts "$t$" or "$x$" mean partial derivatives with respect to $t$ or $x$, respectively):

$$\langle L^+V\,|\,\Lambda_1^{\epsilon}(k,x)\rangle=kR^{\epsilon}(k,t),\qquad \langle\sigma_3V_t\,|\,\Lambda_1^{\epsilon}(k,x)\rangle=R_t^{\epsilon}(k,t),\tag{III.5}$$

$$\langle L^+V\,|\,\Lambda_2^{\epsilon}(k,x)\rangle=0,\quad \langle\sigma_3V_t\,|\,\Lambda_2^{\epsilon}(k,x)\rangle=T_t(k,t),\tag{III.6}$$

$$\langle L^+V\,|\,\Lambda_3^{\epsilon}(k_n^{\epsilon},x)\rangle=k_n^{\epsilon}C_n^{\epsilon}(t),\tag{III.7}$$

where the operator $L^+$ is defined by

$$L^+=\frac{1}{2i}\begin{pmatrix}\dfrac{\partial}{\partial x}-2\tilde r\displaystyle\int_{-\infty}^{x}\tilde q & 2\tilde r\displaystyle\int_{-\infty}^{x}\tilde r\\ -2\tilde q\displaystyle\int_{-\infty}^{x}\tilde q & -\dfrac{\partial}{\partial x}+2\tilde q\displaystyle\int_{-\infty}^{x}\tilde r\end{pmatrix}.\tag{III.8}$$

Moreover, if $T_t(k,t)=0$, then the number $(N^{\epsilon})$ and the position $(k_n^{\epsilon})$ of the bound states are constant in time, and we have therefore

$$\langle \sigma_3 V_t \,|\Lambda_3^\epsilon(k_n^\epsilon, x)\rangle = \frac{d}{dt} C_n^\epsilon(t). \qquad \text{(III.7 bis)}$$

These scalar products are computed in Sec. III.B.

To achieve the inverse method, we need secondly a relationship between some evolution of $V(x,t)$, namely, NEE:

$$\sigma_3 V_t = \Omega(L^+)V, \qquad \text{(III.9)}$$

and the corresponding evolution of the spectral data. [$\Omega(\mathscr{E})/\mathscr{E}$ is an entire function of $\mathscr{E}$.] This relationship arises from the existence of operators $L$ and $L^+$ such that $(L^+)^n$, which acts in the $x$ space of potentials, is the adjoint of $(L)^n$ which acts in the $x$ space of functions $\Lambda^\epsilon$ and such that $\Lambda^\epsilon$ is the eigenfunction of $(L)^n$ with the eigenvalue $(\epsilon E)^n$. The fundamental properties of $L$ and $L^+$ are studied in Sec. III.C. These properties allow us to derive, in Sec. III.D, the following powerful relations: for any entire function $\Omega(\epsilon E)/\epsilon E$ of $\epsilon E$,

$$\langle \Omega(L^+)V \,|\Lambda_1^\epsilon\rangle = k\frac{\Omega(\epsilon E)}{\epsilon E} R^\epsilon(k,t), \qquad \text{(III.10)}$$

$$\langle \Omega(L^+)V \,|\Lambda_2^\epsilon\rangle = 0, \qquad \text{(III.11)}$$

$$\langle \Omega(L^+)V \,|\Lambda_3^\epsilon\rangle = k\frac{\Omega(\epsilon E)}{\epsilon E}\Big|_{k=k_n^\epsilon} C_n^\epsilon(t). \qquad \text{(III.12)}$$

If $V(x,t)$ evolves now according to the NEE (III.9), it is obvious to verify from Eqs. (III.5) and (III.10) and (III.6) and (III.11) that the spectral data evolves according to the trivially integrable equations

$$R_t^\epsilon(k,t) = k\frac{\Omega(\epsilon E)}{\epsilon E} R^\epsilon(k,t), \quad T_t^\epsilon(k,t) = 0, \qquad \text{(III.13)}$$

and thus

$$\frac{d}{dt} k_n^\epsilon = 0, \qquad \text{(III.14)}$$

which gives from Eqs. (III.7) and (III.12)

$$\frac{d}{dt} C_n^\epsilon(t) = k\frac{\Omega(\epsilon E)}{\epsilon E}\Big|_{k=k_n^\epsilon} C_n^\epsilon(t). \qquad \text{(III.15)}$$

At this point, the nonlinear evolution problem (III.9) together with some initial data $V_0(x) = V(x,t_0)$ is solvable by the inverse technique:

$$V_0(x) \to \mathscr{S}(t_0) \to \mathscr{S}(t) \to V(x,t),$$

i.e., first solve the direct scattering problem for $D$ at $t_0$, obtain then $\mathscr{S}(t)$ from $\mathscr{S}(t_0)$ by (III.13) and (III.15), and solve at last the inverse problem at $t$: obtain $V(x,t)$ from $\mathscr{S}(t)$ through inversion formulas (II.10)–(III.12).

The class of NEE solvable by this method is strongly related to the form of the linear eigenvalue problem used. The problem of finding an adequate scattering problem for a given NEE is still open, and it seems that, so long as one can construct new inverse problems, one can find new classes of NEE (exept when there exist a transformation which gives a class from another.[3,11])

## B. Computation of the scalar products

Let us start from euqation $D$ for $Y^\epsilon(k,x)$ and $Z^\epsilon(k,x)$, written into the following form:

$$\left.\begin{aligned} Y_x^\epsilon &= M^\epsilon Y^\epsilon \\ {}^T Z_x^\epsilon &= {}^T Z^\epsilon M^\epsilon \end{aligned}\right\}, \quad M^\epsilon = \begin{pmatrix} -i\epsilon E & \tilde{q} \\ \tilde{r} & i\epsilon E \end{pmatrix}, \qquad \text{(III.16)}$$

or into a "matrix Schrödinger" type form[3,12]:

$$\left.\begin{aligned} Y_{xx}^\epsilon &= NY^\epsilon \\ {}^T Z_{xx}^\epsilon &= {}^T Z^\epsilon {}^T N \end{aligned}\right\}, \qquad \text{(III.17)}$$

$$N = \begin{pmatrix} (\tilde{r}\tilde{q} - m^2) - k^2 & \tilde{q}_x \\ \tilde{r}_x & (\tilde{r}\tilde{q} - m^2) - k^2 \end{pmatrix}.$$

Using Eqs. (III.16) and (III.17), it is obvious to show the identity

$$\frac{d}{dx}({}^T Z^\epsilon i\sigma_2 Y_x^\epsilon - {}^T Z_x^\epsilon i\sigma_2 Y^\epsilon) = 2\,{}^T Z^\epsilon \begin{pmatrix} \tilde{r}_x & 0 \\ 0 & -\tilde{q}_x \end{pmatrix} Y^\epsilon. \qquad \text{(III.18)}$$

Integrating the two sides of Eq. (III.18) for $x$ on $\mathbb{R}$, one gets from the definition of $\Lambda^\epsilon$,

$$[{}^T Z^\epsilon i\sigma_2 Y_x^\epsilon - {}^T Z_x^\epsilon i\sigma_2 Y^\epsilon]_{-\infty}^{+\infty} = 2\langle \sigma_3 V_x \,|\Lambda^\epsilon\rangle. \qquad \text{(III.19)}$$

Since $\tilde{r}_x$ and $\tilde{q}_x$ are integrable on $\mathbb{R}$, and $\Lambda_n^\epsilon$ bounded for $n = 1,2,3$, then the right hand side of Eq. (III.19) exists. Its left hand side can be calculated for $\Lambda^\epsilon = \Lambda_{1,2,3}^\epsilon$ successively, by using the formulas (II.4) and (II.5). Moreover, using the identity immediately derived from definition (III.8) of $L^+$:

$$2iL^+ V = \sigma_3 V_x, \qquad \text{(III.20)}$$

one get the first part of formulas (III.5)–(III.7). Using now Eq. (III.16) and the fact that ${}^T M^\epsilon \sigma_2 + \sigma_2 M^\epsilon = 0$, it is easy to prove the following identity:

$$\frac{d}{dx}({}^T Z^\epsilon i\sigma_2 Y_t^\epsilon) = {}^T Z^\epsilon (i\sigma_2 M_t^\epsilon) Y^\epsilon = (\sigma_3 V_t \mid \Lambda^\epsilon). \qquad \text{(III.21)}$$

Since $\langle \sigma_3 V_t \,|\Lambda_n^\epsilon\rangle$ exists for $n = 1,2,3$, we integrate both sides of Eq. (III.21) on $\mathbb{R}$. The asymptotic behaviors of the Jost functions give rise to the second part of formulas (III.5) and (III.6). Making then the assumption that $k_n^\epsilon$ does not depend on time $t$ [which will be true as soon as $V$ evolves accroding to Eq. (III.9)], we deduce formula (III.7bis) from the identity (III.21).

## C. The operators $L$ and $L^+$

The existence and adequate properties of operators $L$ and $L^+$ allow us to go from the $x$ space, in which the evolution for $V$ is nonlinear and thus rather difficult to solve, to the $k$ space in which the evolution for the spectral data is trivially integrable. It is already a well known fact that this procedure is formally similar to the Fourier transform technique for linear cases.[5]

Let $L$ be the matrix operator

$$L = \frac{1}{2i}\begin{pmatrix} -\frac{\partial}{\partial x} - 2\,\tilde{q}\int_x^\infty \tilde{r} & -2\,\tilde{q}\int_x^\infty \tilde{q} \\ 2\tilde{r}\int_x^\infty \tilde{r} & \frac{\partial}{\partial x} + 2\,\tilde{r}\int_x^\infty \tilde{q} \end{pmatrix}, \qquad \text{(III.22)}$$

and $L^+$ the operator defined by Eq. (III.8). We have to determine at first the spaces $S$ and $S^+$ which act, respectively, on $L$ and $L^+$. From the definition of functions $F$ and $G$ and of $L$ and $L^+$, we have

$$L^+F = \frac{1}{2i}\sigma_3\frac{dF}{dx} - V\int_{-\infty}^{x}(\sigma_2 V \mid F)dx', \tag{III.23}$$

$$LG = -\frac{1}{2i}\sigma_3\frac{dG}{dx} - \sigma_2 V\int_{x}^{\infty}(V \mid G)dx'. \tag{III.24}$$

Hence,

$$S^+ = \{F : \mathbb{R}\to\mathbb{C}^2 \Big/ \int_{-\infty}^{x}(\sigma_2 V \mid F)dx' \text{ exists}\}, \tag{III.25}$$

$$S = \{G : \mathbb{R}\to\mathbb{C}^2 \Big/ \int_{x}^{\infty}(V \mid G)dx' \text{ exists}\}. \tag{III.26}$$

*Theorem* 1: $(L^+)^n V$ exists for all integer $n \geqslant 1$, and the components of this vectorial function are "polynomials" with respect to $\tilde{q}$, $\tilde{r}$ and their successive $x$ derivatives which vanish as $|x|$ goes to infinity (such a function will be called for short a polynomial).

To prove the theorem, we shall make use of the two following lemmas:

*Lemma* 1: If $F$ belongs to $S^+$, then

$$(L\sigma_2 - \sigma_2 L^+)F = \sigma_2 V\langle \sigma_2 V \mid F\rangle . \tag{III.27}$$

In particular, if $F$ is orthogonal to $\sigma_2 V$, we have

$$\sigma_2 L^+ F = L\sigma_2 F. \tag{III.28}$$

Property (III.27) is immediately derived from Eqs. (III.23) and (III.24). In the same way, and using partial integration, one can prove the second lemma:

*Lemma* 2: If $F$ belongs to $S^+$, $G$ to $S$, and $\lim_{x\to\infty}(F \mid \sigma_3 G) = 0$, then

$$\int_{-\infty}^{x}\{(F \mid LG) - (G \mid L^+F)\} = -(F \mid \sigma_3 G) - 2\int_{x}^{\infty}(V \mid G)\int_{-\infty}^{x}(i\sigma_2 V \mid F). \tag{III.29}$$

If moreover $\langle \sigma_2 V \mid F\rangle$ exists and $\lim_{|x|\to\infty}(F \mid \sigma_3 G) = 0$, then

$$\langle F \mid LG\rangle = \langle L^+F \mid G\rangle . \tag{III.30}$$

We are now able to prove theorem 1 by induction with the following induction hypothesis: $(H_n)$: $(L^+)^p V$, $\int_{-\infty}^{x}(\sigma_2 V \mid (L^+)^p V)$ exist and are polynomials.

$(H_n)$ is true for $n = 1$, and indeed we have

$$L^+ V = \frac{1}{2i}\begin{pmatrix} \tilde{r}_x \\ -\tilde{q}_x \end{pmatrix}, \tag{III.31}$$

$$\int_{-\infty}^{x}(\sigma_2 V \mid L^+ V) = -\tfrac{1}{2}(\tilde{r}\tilde{q} - m^2). \tag{III.32}$$

Supposing that $(H_n)$ is true, we prove now that $(H_{n+1})$ is then true. Since $(L^+)^n V$ belongs to $S^+$, then $(L^+)^{n+1} V$ exists. Let us insert $F = (L^+)^n V$ into formula (III.23):

$$(L^+)^{n+1} V = -\tfrac{1}{2}i\sigma_3\frac{d}{dx}\{(L^+)^n V\} - V\int_{-\infty}^{x}(\sigma_2 V \mid (L^+)^n V). \tag{III.33}$$

Thus $(L^+)^{n+1} V$ is a polynomial (use $H_n$). It remains to prove that $\int_{-\infty}^{x}(\sigma_2 V \mid (L^+)^{n+1} V)$ exists and is a polynomial.

We can use lemma 2 for $F = (L^+)^q V$ and $G = i\sigma_2 (L^+)^{n-q} V$, for $0 \leqslant q \leqslant n$. Indeed, since $(L^+)^q V$ and $(L^+)^{n-q} V$ are polynomials, we have

$$\lim_{x\to-\infty}((L^+)^q V \mid i\sigma_3\sigma_2 (L^+)^{n-q} V) = 0. \tag{III.34}$$

Hence, lemma 2 gives

$$\int_{-\infty}^{x}\{(i\sigma_2 (L^+)^{n-q} V \mid (L^+)^{q+1} V) - (Li\sigma_2 (L^+)^{n-q} V \mid (L^+)^q V)\} = (i\sigma_3\sigma_2 (L^+)^{n-q} V \mid (L^+)^q V) + 2\int_{x}^{\infty}(V \mid i\sigma_2 (L^+)^{n-q} V)\int_{x}^{\infty}(i\sigma_2 V \mid (L^+)^q V). \tag{III.35}$$

We have moreover $\langle V \mid i\sigma_2 (L^+)^{n-q} V\rangle = 0$ and thus

$$\int_{x}^{\infty}(V \mid i\sigma_2 (L^+)^{n-q} V) = -\int_{-\infty}^{x}(V \mid i\sigma_2 (L^+)^{n-q} V) \tag{III.36}$$

is a polynomial for $0 \leqslant q \leqslant n$ [for $q = n$, we have $(\sigma_2 V \mid V) = 0$]. Thus, the right hand side of Eq. (III.35) is a polynomial, namely, $P_q(V)$. Moreover, we can make use of Eq. (III.28) for $F = (L^+)^{n-q} V$:

$$L\sigma_2 (L^+)^{n-q} V = \sigma_2 (L^+)^{n-q+1} V. \tag{III.37}$$

Then formula (III.35) writes

$$\int_{-\infty}^{x}\{(i\sigma_2 (L^+)^{n-q} V \mid (L^+)^{q+1} V) - (i\sigma_2 (L^+)^{n-q+1} V \mid (L^+)^q V)\} = P_q(V). \tag{III.38}$$

Let's now add the formulas (III.38) written for $q = 1,2,\ldots,n$; we finally get

$$\int_{-\infty}^{x}\{(i\sigma_2 V \mid (L^+)^{n+1} V) - (i\sigma_2 (L^+)^{n+1} V \mid V)\} = \sum_{q=0}^{n} P_q(V). \tag{III.39}$$

However, since ${}^T(i\sigma_2) = -i\sigma_2$, we have $(i\sigma_2 (L^+)^{n+1} V \mid V) = -(i\sigma_2 V \mid (L^+)^{n+1} V)$. Then, $\int_{-\infty}^{x}(\sigma_2 V \mid (L^+)^{n+1} V)$ exists and is a polynomial, which ends the proof of theorem 1.

We shall make use of another important theorem:

*Theorem* 2: $(L)^n \sigma_2 V$ exists for all positive integer $n$, and we have

$$(L)^n \sigma_2 V = \sigma_2 (L^+)^n V. \tag{III.40}$$

*Proof*: Eq. (III.40) is true for $n = 1$ [use lemma 1 for $F = V$ and the fact that $(\sigma_2 V \mid V) = 0$].

Suppose now that Eq. (III.40) is true for $n$; we show that it is true for $n + 1$. We use lemma 1 for $F = (L^+)^n V$ and we notice that, from theorem 1, $\langle \sigma_2 V \mid (L^+)^n V\rangle = 0$, to show that

$$(L)^{n+1}\sigma_2 V = L\sigma_2 (L^+)^n V = \sigma_2 (L^+)^{n+1} V. \tag{III.41}$$

It obviously results from theorem 2 that

$$\langle V \mid (L)^n \sigma_2 V\rangle = \langle V \mid \sigma_2 (L^+)^n V\rangle = 0. \tag{III.42}$$

This last property allow us to prove another singular theorem:

*Theorem* 3:

$$(L^+)^n V = (\mathbf{L}^+)^n, \quad n \in \mathbf{N},$$

where the operator $\mathbf{L}^+$ is

$$\mathbf{L}^+ = \frac{1}{2i}\begin{pmatrix} \frac{\partial}{\partial x} + 2\tilde{r}\int_x^\infty \tilde{q} & -2\tilde{r}\int_x^\infty \tilde{r} \\ 2\tilde{q}\int_x^\infty \tilde{q} & -\frac{\partial}{\partial x} - 2\tilde{q}\int_x^\infty \tilde{r} \end{pmatrix}. \quad \text{(III.43)}$$

*Proof*: the above definition gives

$$\mathbf{L}^+ F = \frac{1}{2i}\sigma_3 \frac{dF}{dx} + \int_x^\infty (\sigma_2 V \mid F). \quad \text{(III.44)}$$

Using Eqs. (III.23) and (III.44),

$$(\mathbf{L}^+ - L^+)F = \langle \sigma_2 V \mid F \rangle, \quad \text{(III.45)}$$

and making $F = V$ in Eq. (III.45), we show that theorem 3 is true for $n = 1$ [use $(\sigma_2 V \mid V) = 0$]. Let us suppose that theorem 3 is true for $n$; we prove that it is true for $n+1$. To do that we set $F = (L^+)^n V = (\mathbf{L}^+)^n V$ in Eq. (III.45); we get

$$\{(\mathbf{L}^+)^{n+1} - (L^+)^{n+1}\} V = \langle \sigma_2 V \mid (L^+)^n V \rangle, \quad \text{(III.46)}$$

which is zero from Eq. (III.42).

### D. Evolution equations

With the help of the above theorems, we are now able to derive the following properties:

*Property* 1: For $p$ and $q$ integers such that $p \geqslant 1$, $q \geqslant 0$, we have

$$\lim_{A\to\infty} \int_{-\infty}^{A} ((L^+)^p V \mid L_A{}^q \Lambda_i^\epsilon) = (\epsilon E)^q \langle (L^+)^p V \mid \Lambda_i^\epsilon \rangle, \quad i = 1,2,3, \quad \text{(III.47)}$$

where $L_A$ is the cutoff operator deduced from $L$ by changing $\int_x^\infty$ into $\int_x^A$. The meaning of Eq. (III.47) is that $\Lambda_i^\epsilon$ is the eigenfunction of $L^q$ with the eigenvalue $(\epsilon E)^q$.

*Property* 2: For any positive integer $n$,

$$\lim_{A\to\infty} \int_{-\infty}^{A} (L^+ V \mid L_A{}^n \Lambda_i^\epsilon) = \langle (L^+)^{n+1} V \mid \Lambda_i^\epsilon \rangle. \quad \text{(III.48)}$$

The meaning of Eq. (III.48) is that $(L^+)^n$ is the adjoint operator of $(L)^n$.

*Proof of property* 1: Starting with the identities deduced from $D$:

$$i\frac{d}{dx}(Y_1^\epsilon Z_1^\epsilon) - i\tilde{q}(Y_1^\epsilon Z_2^\epsilon + Y_2^\epsilon Z_1^\epsilon) = 2\epsilon E Y_1^\epsilon Z_1^\epsilon, \quad \text{(III.49)}$$

$$i\frac{d}{dx}(Y_1^\epsilon Z_2^\epsilon + Y_2^\epsilon Z_1^\epsilon) = 2i(\tilde{r} Y_1^\epsilon Z_1^\epsilon + \tilde{q} Y_2^\epsilon Z_2^\epsilon), \quad \text{(III.50)}$$

and integrating Eq. (III.50) on the segment $[x, A]$, we prove that there exist scalar bounded functions $\delta_i(A)$ such that

$$L_A \Lambda_i^\epsilon = \epsilon E \Lambda_i^\epsilon + \delta_i(A)\sigma_2 V, \quad i = 1,2,3. \quad \text{(III.51)}$$

The functions $\delta_i(A)$ are given by

$$\delta_1(A) = \tfrac{1}{2} T^\epsilon(k)^2 ({}^T\phi^\epsilon(k,A)\sigma_1 \phi^\epsilon(k,A)), \quad \text{(III.52)}$$

$$\delta_2(A) = \tfrac{1}{2} T^\epsilon(k)^2 ({}^T\phi^\epsilon(k,A)\sigma_1 \psi^\epsilon(k,A)), \quad \text{(III.53)}$$

$$\delta_3(A) = \tfrac{1}{2}\left[\frac{d}{dk}\frac{1}{T^\epsilon(k)}\right]^{-1} \cdot ({}^T\phi^\epsilon(k,A)\sigma_1 \bar{\psi}^\epsilon(k,A))\Big|_{k=k_n^\epsilon}. \quad \text{(III.54)}$$

At this point, one must notice that, contrary to the case $r^\pm = q^\pm = 0$, the functions $\delta_{1,2}(A)$ have no limit as $A \to \infty$. This is the reason why we must use the cutoff operator $L_A$ inspite of $L$.

By induction on Eq. (III.51), we get

$$L_A{}^n \Lambda_i^\epsilon = (\epsilon E)^n \Lambda_i^\epsilon + \sum_{l=1}^{n} (\epsilon E)^{n-l} \delta_i(A) L_A{}^{l-1}\sigma_2 V, \quad n \in \mathbf{N}. \quad \text{(III.55)}$$

Thus, for $p \geqslant 1$ and $q \geqslant 0$:

$$\int_{-\infty}^{A} ((L^+)^p V \mid L_A{}^q \Lambda_i^\epsilon) = (\epsilon E)^q \int_{-\infty}^{A} ((L^+)^p V \mid \Lambda_i^\epsilon) + \sum_{l=1}^{q} (\epsilon E)^{q-l} \delta_i(A) \times \int_{-\infty}^{A} ((L^+)^p V \mid L_A{}^{l-1}\sigma_2 V). \quad \text{(III.56)}$$

When $A$ goes to infinity, the second term of right hand side of Eq. (III.56) goes to zero. Indeed we have that $\delta_i(A)$ is a bounded function of $a$, and

$$-\int_{-\infty}^{A} ((L^+)^p V \mid L_A{}^{l-1}\sigma_2 V) \underset{A\to\infty}{\to} \langle (L^+)^p V \mid L^{l-1}\sigma_2 V \rangle = 0,$$

by using theorem 2 and Eq. (III.42). This ends the proof of property 1.

*Proof of property* 2: If $F$ belongs to $S^+$ and $G$ is a bounded function, one can show by partial integration that

$$\int_{-\infty}^{A} \{(F \mid L_A G) - (L^+ F \mid G)\} = -(F \mid \sigma_3 G)(A). \quad \text{(III.57)}$$

Setting $F = (L^+)^p V$ and $G = (L_A)^{q-1}\Lambda_i^\epsilon$ in Eq. (III.57) leads to

$$\int_{-\infty}^{A} \{((L^+)^p V \mid (L_A)^q \Lambda_i^\epsilon) - ((L^+)^{p+1} V \mid (L_A)^{q-1}\Lambda_i^\epsilon)\} = -((L^+)^p V \mid \sigma_3 (L_A)^{q-1}\Lambda_i^\epsilon)(A). \quad \text{(III.58)}$$

When $A$ goes to infinity, the right hand side of Eq. (III.58) goes to zero, and we immediately obtain Eq. (III.48) by induction.

At this point, we use properties 1 and 2 to show that

$$\langle (L^+)^{n+1} V \mid \Lambda_i^\epsilon \rangle = (\epsilon E)^n \langle L^+ V \mid \Lambda_i^\epsilon \rangle. \quad \text{(III.59)}$$

*Property* 3: For any entire function $\Omega(\epsilon E)/(\epsilon E)$ of $\epsilon E$, we have

$$\langle \Omega(L^+) \cdot V \mid \Lambda_i^\epsilon \rangle = \frac{\Omega(\epsilon E)}{\epsilon E} \langle L^+ V \mid \Lambda_i^\epsilon \rangle. \quad \text{(III.60)}$$

Use now Eq. (III.60) together with Eqs. (III.5)–(III.7) to get formulas (III.10)–(III.12).

At this point the nonlinear evolution problem (III.9) is solved by I.S.T. However, one must notice that we have admitted the existence of a solution of Eq. (III.9) and that the inverse problem is solvable for any value of $t$. To prove the existence of a solution of Eq. (III.9), one would deduce from the evolution of $\mathcal{S}$ the NEE (III.9). This would be possible with the help of a closure theorem for the functions $\Lambda_i^\epsilon$.

(Such a result had been obtained by Kaup in the case $r^{\pm} = q^{\pm} = 0$.[15])

## IV. SOLITONS

Solitons are known to be the solutions of NEE when the associated spectral data reduce to the discrete spectrum only. We shall study here two representative cases corresponding to a set of spectral data consisting of one and two bound states, and no reflection coefficient. Although we do not want to enter in a complete study of the properties of these solutions, we shall point out the strong differences between the Zakharov–Shabat (ZS) case and the Dirac (D) case (we shall thus simplify the hypothesis).

### A. Single bound state: "kinklike" solution

Suppose that $\mathcal{S}$ consist in a single zero of $1/T^{+}(k)$ and $R^{\epsilon}(k) = 0$ for real $k$. The "Fourier transform" of $\mathcal{S}$ is then from Eq. (II.12):

$$F(x+y) = -iMC\, e^{ik(x+y)}, \tag{IV.1}$$

$$M = \frac{1}{2E}\begin{pmatrix} -iq^{+} & \dfrac{-(q^{+})^2}{E+k} \\ E+k & -iq^{+} \end{pmatrix},$$

where $C$ is the bound state constant (II.8). The noninversible matrix $M$ possesses the property

$$M\left(M + \frac{iq^{+}}{E}\right) = 0. \tag{IV.2}$$

Writing now the inversion formula (II.11), one can see that the kernel $K(x,y)$ becomes separable; thus, we set $K(x,y) = A(x)e^{iky}$. The inversion equation reduces to

$$-iCM + A(x)\left\{e^{-ikx} + \frac{C}{2k} M\, e^{ikx}\right\} = 0. \tag{IV.3}$$

The system (IV.3) of four equations reduces, by using property (IV.2), to a system of two equations.

The solution of the inverse problem is given by

$$[\sigma_3, A(x)e^{ikx}] = \begin{pmatrix} 0 & -\tilde{q} + q^{+} \\ \tilde{r} - r^{+} & 0 \end{pmatrix}, \tag{IV.4}$$

which gives

$$\tilde{q}(x) = q^{+}\frac{(Cq^{+}/2ikE)(E-k/E+k)e^{ikx} + e^{-ikx}}{(Cq^{+}/2ikE)e^{ikx} + e^{-ikx}}, \tag{IV.5}$$

$$\tilde{r}(x) = r^{+}\frac{(Cq^{+}/2ikE)(E+k/E-k)e^{ikx} + e^{-ikx}}{(Cq^{+}/2ikE)e^{ikx} + e^{-ikx}}. \tag{IV.6}$$

Before going further on, notice that $\tilde{q}$ and $\tilde{r}$ possess the right asymptotic behavior [i.e., formula (II.13) holds].

Assuming now that $\tilde{r}$ and $\tilde{q}$ evolves according to the NEE (III.9), then the constant $C$ evolves according to Eq. (III.15), i.e.,

$$C(t) = C_0 \exp\left(k\frac{\Omega(E)}{E}t\right). \tag{IV.7}$$

Setting

$$\frac{C_0 q^{+}}{2ikE} = e^{2a}, \quad \frac{E-k}{E+k} = e^{2b}, \tag{IV.8}$$

we can write $\tilde{q}$ in the form

$$\tilde{q}(x,t) = q^{+}e^{a}\cosh b\left\{+1(thb)th\left[ik\left(x - \frac{i}{2}\frac{\Omega(E)}{E}t\right) + a\right]\right\}. \tag{IV.9}$$

This "th" form is known as a kink[11] which moves with constant speed $v = \mathrm{Re}\{(-i/2E)\Omega(E)\}$.

In the ZS case a single pole does lead to a soliton solution.[15] Indeed, setting here $q^{\pm} = r^{\pm} = 0$, we get the trivial solution $\tilde{q} = \tilde{r} = 0$. Thus, the D case is strongly different from the ZS case not only because of the emergence of a solution when the spectral data reduce in one single bound state, but also because of the particular feature of this solution, which was reserved until now to solutions of the sine-Gordon equation.

### B. Two bound states: "boomeron" solution

We now assume that the spectral data consist of two bound states located at $k^{+}$ and $k^{-}$ [zeros of $1/T^{+}(k)$ and $1/T^{-}(k)$, respectively], and no reflection coefficient. We shall assume moreover for simplicity that $q^{\pm} = r^{\pm} = m$, $m$ real. In this case the stability of the inverse problem demands $k^{+} = k^{-}$. Indeed, Eq. (I.13) leads to

$$\frac{E^{+}+k^{+}}{E^{+}-k^{+}}\frac{E^{-}-k^{-}}{E^{-}+k^{-}} = 1, \tag{IV.10}$$

that is to say,

$$\frac{(k^{+})^2}{(k^{-})^2} = \frac{(k^{+})^2 + m^2}{(k^{-})^2 + m^2}. \tag{IV.11}$$

Since $\mathrm{Im}(k^{\pm}) > 0$, the solution of Eq. (IV.11) is $k^{+} = k^{-}$.

To simplify at most the results, we set moreover $k^{+} = k^{-} = iu$ $(m > u > 0)$. The method is essentially the same as that used in Eq. (IV.a); thus, we give only the final result:

$$\tilde{q}(x,t) - m = 2u\left[\left(\frac{u}{E} + i\right)e^{iu(\Omega_e/E)t} + \left(\frac{u}{E} - i\right)e^{-iu(\Omega_e/E)t}\right] \times \cosh^{-1}\left\{2u\left(x - x_0(t) - \frac{i}{2E}\Omega t - w\right)\right\}, \tag{IV.12}$$

where $\Omega_e$ and $\Omega_0$ are, respectively, the even and odd parts of $\Omega(E)$, where

$$w = \frac{1}{2u}\log\frac{C_0}{2u}, \tag{IV.13}$$

and where $x_0(t)$ is a solution of

$$\cosh\left[2u\left(x_0 + w + \frac{iu}{2E}\Omega_0 t\right)\right] = \cosh\left[2u\left(\frac{i}{2E}\Omega_0 t + w\right)\right] + \frac{m}{E}\cosh\left(i\frac{u}{E}\Omega_e t\right). \tag{IV.14}$$

Notice that we have assumed that $\Omega(E)$ is purely imaginary, in order to have a real dispertion relation.[5]

The most remarkable feature of these solutions is that the center of motion of the envelope soliton has a speed that evolves in time as soon as $m$ is different from zero. [Setting $m = 0$ in Eq. (IV.12) leads to $x_0 = 0$ and thus the soliton moves with constant speed $(i/2E)\Omega_0$.] Indeed, this speed is given by

$$v(t) = \{(i/2E)\Omega_0 \sinh[2u((i/2E)\Omega_0 t + w)] + (im/2E)\Omega_e \sinh(iu/E)\Omega_e t\}/ \sinh[2u(x_0 + w + (iu/2E)\Omega_0 t)] . \quad \text{(IV.15)}$$

At this point, one must notice that the position $[x_0 + (iu/2E)\Omega_0 t]$ of the center of the soliton has a definite sign. Indeed, we have from Eq. (IV.14)

$$\cosh\left[2u\left(x_0 + w + \frac{iu}{2E}\Omega_0 t\right)\right] \geqslant 1 + \frac{m}{E} > 1, \quad \text{(IV.16)}$$

and the continuity of the function $[x_0 + (iu/2E)\Omega_0 t]$ ensures that it has constant sign. Without loss of generality, we suppose that it is positive. Then the sign of $v(t)$ is that of its numerator. It is possible to show that $v(t)$ changes its sign when passing through the value zero, that it is asymptotically constant, and that $v(t = +\infty) = -v(t = -\infty)$. This solution is then a soliton that comes back; it is called a boomeron.

Let us list for instance a few properties of the boomeron of the nonlinear Schrödinger equation which is obtained for $\Omega(E) = 4iE^2$, and which is

$$\tilde{q}_t = i\tilde{q}_{xx} - 2i\tilde{q}^2\tilde{q}^* + 2im^2\tilde{q} \quad (\tilde{r} = \tilde{q}^*).$$

$v(t)$ becomes an odd function of $t$ because $\Omega_0 = 0$, and goes asymptotically to $\pm 4E^2$ at $\pm\infty$. Another difference between the ZS and D cases arises then from the fact that the corresponding solution in the ZS case ($m = 0$) does not evolves in time, because we have assumed that $k = iu$ and $\Omega_0 = 0$ (see Ref. 5, p. 279).


## ACKNOWLEDGMENTS

The author is especially grateful to Professor P.C. Sabatier for helpful discussions, and to Professor Baghdadi for the interest he had shown in this work. He would like to thank Dr. M. Jaulent for his help in the achievement of this paper.



[1]A general survey of the theory of solitons is given by A.C. Scott, F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1449 (1973).

[2]We follow in this appelation F. Calogero and A. Degasperis. See for instance their review paper in Nuovo Cimento B **39**, 1 (1977).

[3]M. Jaulent and J.JP. Leon, Lett. Nuovo Cimento **23**, 129 (1978). Another proof has been given by the use of results of Ref. 2 in Lett. Nuovo Cimento **23**, 137 (1978).

[4]V.S. Gerdjikov and P.P. Kulish, Preprint of the J.I.N.R. (1978).

[5]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).

[6]V.E. Zakharov and A.B. Shabat, Sov. Phys. JETP **34**, 62 (1972) [Zh. Eksp. Teor. Fiz. **61**, 118 (1971)].

[7]See for instance K. Chadan and P.C. Sabatier, "*Inverse Problems in Quantum Scattering Theory*" (Springer, New York, 1977).

[8]J. Leon, Thèse de Doctorat de Troisième Cycle, U.S.T.L., Montpellier (1978).

[9]A similar result has been obtained in a very different way for the ZS case by F. Calogero and A. Degasperis in the Appendix of Nuovo Cimento B **32**, 201 (1976).

[10]This equation for $r^{\pm} = q^{\mp} = m$, $m \in \mathbb{R}$ is transformed under the unitary transformation $Y \to \frac{1}{2}(\sigma_2 + \sigma_3)Y$ from the Dirac equation used in its radial version by M. Gasymov and B.M. Levitan, Dokl. Akad. Nauk SSSR **167**, 495 (1966). For the derivation of the radial Dirac equation see M.C. Barthelemy, Ann. Inst. Henri Poincaré **6**, 365 (1967); and **7**, 115 (1967).

[11]M. Jaulent and I. Miodek, Lett. Nuovo Cimento **20**, 655 (1977).

[12]M. Wadati and T. Kamijo, Prog. Theor. Phys. **52**, 397 (1974).

[13]D.J. Kaup, J. Math. Anal. Appl. **54**, 849 (1976).

[14]Kinks are known as solutions of the sine Gordon equation. For an example, one can refer to Ref. 1 or 5, p. 280.

[15]See Ref. 5, p. 278.

# Parametrizations of unitary matrices and related coset spaces

A. J. Macfarlane
*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England*

(Received 14 August 1979; accepted for publication 15 October 1979)

Explicit forms of U3 and U4 matrices as functions of a single adjoint vector are displayed. Parametrizations of the coset spaces $U(N+r)/(UN\times Ur)$ are discussed, most explicitly for $r=1$ and 2, and related, for $N=3$ and 4, to the results for U3 and U4 matrices.

## 1. INTRODUCTION

There has always been interest in and a genuine need for suitable explicit forms of parametrizations of unitary matrices. In the past, this was true for Lagrangian theories of chiral symmetry based on nonlinear realizations of $SUN\times SUN$.[1-4] At present, it holds true for the study of field theories, which possess instanton solutions. This area includes study of $CP^N$ models,[5,6] and of generalized $\sigma$ models.[7] The former may be viewed[8] as nonlinear realizations of an $SU(N+1)$ symmetry in which a $UN$ subgroup is realized linearly; the latter may be similarly viewed as nonlinear realizations of a $U(N+r)$ symmetry in which a $UN\times Ur$ subgroup is realized linearly.

This paper displays and gives proofs (rather than derivations) of a variety of results to which, for ease of reference, we attach Roman numerals, regarding unitary matrices that have been discovered at various times during the last ten years. Section 2 gives results for U3 matrices, written in terms of a single octet vector. The simplest result, result III below, has been given before.[3] Section 4 applies results given in Sec. 3 regarding SU4 adjoint vectors, to give results for U4 matrices written in terms of a single adjoint vector. These results are new. Section 5 discusses the coset space $SU(N+1)/UN$ writing $SU(N+1)$ matrices in terms of sets of Goldstone fields for nonlinear realizations of $SU(N+1)$ in which $UN$ is realized linearly.[8] Result VIII underlies[8] recent work of Cremmer and Scherk.[9] The $N=3$ case of result IX has been used by Salam and Strathdee.[10] Section 6 discusses the coset space $U(N+r)/(UN\times Ur)$ writing $U(N+r)$ matrices in terms of appropriate sets of Goldstone fields. Completely explicit results are given only for $r=2$, and arbitrary $N$, but the methods used extend with increasing complication to higher values of $r$. The results of Sec. 6, which are new, have been used in Ref. 7. The relationship of results of later sections to those of earlier sections is also pointed out. For example, results III and IV of Sec. 2 reduce upon appropriate specialization of the octet vector in terms of which they are given, to the $N=3$ cases of results VIII and IX of Sec. 5. Result VI of Sec. 4 similarly reduces to the $N=4$ case of result XI of Sec. 6. Some of the discussion of Sec. 5 could have alternatively been given by reduction of the $r=1$ case of the results of Sec. 6.

Finally we make reference to an important and elegant paper by Barnes and Delbourgo,[11] which deals with the construction of $SUN$ quantities out of a single adjoint $SUN$ vector. This paper is essentially complementary to the present work. It would provide an alternative approach to the derivation of results such as are given in this paper, but the nontrivial task of extracting simple explicit forms from its very general discussion was not undertaken there.

## 2. PARAMETRIZATION OF U3 MATRICES

We use standard notation for $\lambda$ matrices, the tensor $d_{ijk}$, etc., and refer to previous work for their properties.[10]

Let $M_i$ denote a real octet vector, then as usual $N_i=d_{ijk}M_jM_k$ is real octet vector in general linearly independent of $M_i$. Also $X=M_iM_i$, $Y=M_iN_i$ are SU3 invariants. To prove our results, it will suffice to note the following

$$\begin{aligned} &d_{ijk}M_jN_k=\tfrac{1}{3}XM_i,\\ &d_{ijk}N_jN_k=\tfrac{2}{3}YM_i-\tfrac{1}{3}XN_i,,\\ &N_iN_i=\tfrac{1}{3}X^2. \end{aligned} \tag{2.1}$$

Our various parameterizations are labelled by Roman numerals throughout the paper.

I. If $B$, $C$, and $D$ are three real functions of the SU3 invariants $X$ and $Y$ related by

$$1+B^2+2CD=XC^2, \tag{2.2}$$

then $U(M)$ given by

$$\begin{aligned} \Omega U(M)&=D-\tfrac{1}{3}CX+i(\tfrac{1}{3}C^2Y+BD)\\ &\quad-i\lambda.M(1+iB)+\lambda.NC, \end{aligned} \tag{2.3}$$

$$\begin{aligned} |\Omega|^2&=(1+B^2)(X+D^2)\\ &\quad+\tfrac{2}{3}BCY(2+CD)+\tfrac{1}{9}C^4Y^2, \end{aligned} \tag{2.4}$$

is a unitary matrix.

*Proof*: The result follows straightforwardly from the usual rules for $\lambda$-matrices and (2.1).

The general result I is complicated chiefly because $B$ contains $Y$ as a factor, as can be seen from parity considerations when $M_i$ is pseudoscalar. Result I contains two arbitrary functions of $X$ and $Y$, and we may dispose of this arbitrariness to achieve simpler forms for $U(M)$. Thus, we have:

II. Set $B=0$ in I

For arbitrary $C(X)$, and hence $D(X)$, this keeps down to a minimum dependence on the awkward SU3 invariant $Y$. We may dispose of the remaining freedom in II in various ways.

Set $C=\frac{1}{2}$ and replace $M$ by $M'=-2M$ in II, dropping the primes thereafter. This gives

$$\begin{aligned} &\text{III. } \omega U(M)=1-\tfrac{1}{3}X-\tfrac{2}{3}iY+2i\lambda.M-2\lambda.N,\\ &\qquad|\omega|^2=(1+X)^2+\tfrac{4}{9}Y^2. \end{aligned} \tag{2.5}$$

 0022-2488/80/102579-04$1.00  

Similarly, set $D = 1$ in II, and replace $M$ by $-M$. This gives

$$\omega U(M) = \tfrac{1}{3}(2 + R) - \tfrac{1}{3}iC^2Y + i\lambda.M + \lambda.NC,$$
$$|\omega|^2 = R^2 + \tfrac{1}{9}C^4Y^2, \tag{2.6}$$
$$C = -(1 + R)^{-1}, \quad R = (1 + X)^{1/2}.$$

Other results comparable with IV can be obtained from II. Result III was given originally in our previous work,[3] where indication of its origin was given, as opposed to mere verification of the unitarity of $U(M)$ defined by (2.5). Result I above arose by the same method appropriately streamlined to minimize algebra.

## 3. SU4 $d$ TENSORS AND ADJOINT VECTORS

$\lambda$ matrices for SU4 are the natural generalization of those for SU3, and give rise to

$$\lambda_i\lambda_j = \tfrac{1}{2}\delta_{ij} + (d_{ijk} + if_{ijk})\lambda_k. \tag{3.1}$$

The tensors $d$ and $f$ for SU4 naturally have those properties listed in reference twelve as holding for all SU$n$. However, they fail to satisfy the result

$$d_{m(ij}d_{k)lm} = \tfrac{1}{3}\delta_{(ij}\delta_{k)l}, \tag{3.1a}$$

true for the SU3 $d$ tensors. Here and below, total symmetry must by imposed on a set of indices surrounded by round brackets. To state the property typical of SU4 that its $d$ tensors obey, proceed as follows. Introduce the $r$th rank isotropic tensor $d^{(r)}_{ij\cdots st}$, with $r$ indices, by means of

$$d^{(r+1)}_{ij\cdots stu} = d^{(r)}_{ij\cdots sh}\, d_{htu}, \tag{3.2}$$

where

$$d^{(2)}_{ij} = \delta_{ij},$$

and hence

$$d^{(3)}_{ijk} \equiv d_{ijk}.$$

In terms of such a notation for SU3, Eq. (A) can be written as $d^{(4)}_{(ijk)l} = 1/3\delta_{(ij}\delta_{k)l}$. For SU4, the corresponding result is

$$d^{(5)}_{(ijkl)m} = \tfrac{1}{2}\delta_{(ij}d_{kl)m} + \tfrac{1}{6}d_{(ijk}\delta_{l)m}. \tag{3.3}$$

This result was first given (to this author's knowledge) by Sudbery,[12] who proved the corresponding result for general SU($n$) and displayed the SU4 and SU5 versions of it. His proof used the Cayley–Hamilton theorem for SU$n$ in much the same way (technically streamlined both in notation and in outlook) as used in Ref. 13 for deriving Eq. (A), above, for SU3.

We go on to use (3.3) to discuss the vector and tensor world that can be realized using a single SU4 (adjoint) vector $M_i$. There are in general two vectors linearly independent of $M_i$, namely,

$$N_i = d_{ijk}M_jM_k, \tag{3.3a}$$
$$P_i = d^{(4)}_{ijlm}M_jM_lM_m = d_{ijk}M_jN_k, \tag{3.4}$$

and there are three SU4 invariants

$$X = M_iM_i, \tag{3.5}$$
$$Y = M_iN_i = d_{ijk}M_iM_jM_k, \tag{3.6}$$
$$Z = M_iP_i = N_iN_i = d^{(4)}_{ijkl}M_iM_jM_kM_l. \tag{3.7}$$

We may exploit (3.3) to show that all the further naturally occurring vectors and scalars are respectively linear combinations of $M$, $N$,and $P$, and functions of $X$, $Y$, and $Z$. The results in question are as follows

$$(Q_i =)\, d_{ijk}M_jP_k = \tfrac{1}{2}XN_i + \tfrac{1}{6}YM_i, \tag{3.8}$$
$$d_{ijk}N_jN_k = \tfrac{2}{3}YM_i, \tag{3.9}$$
$$d_{ijk}N_jP_k = \tfrac{1}{2}ZM_i + \tfrac{1}{6}YN_i, \tag{3.10}$$
$$d_{ijk}P_jP_k = \tfrac{1}{3}XYM_i + \tfrac{1}{2}ZN_i - \tfrac{1}{3}YP_i, \tag{3.11}$$
$$N_iP_i = \tfrac{2}{3}XY, \tag{3.12}$$
$$P_iP_i = \tfrac{1}{2}ZX + \tfrac{1}{6}Y^2 \quad (= Q_iM_i). \tag{3.13}$$

These results will enable us to state and prove our results regarding the parametrization of U4 matrices in terms of a single SU4 (adjoint) vector.

## 4. PARAMETRIZATIONS OF U4 MATRICES

The most general parametrization of a U4 matrix in terms of a single adjoint vector $M_i$ (and quantities built, as in Sec. 3, out of it) involves three arbitrary functions of the SU4 invariants $X$, $Y$, and $Z$. Having seen result I for SU3, one would expect this to be very complicated and not very useful. Accordingly, we seek results that keep to a minimum dependence on the higher order invariants $Y$ and $Z$. We begin with

V. If $U \in$ U4 is expanded in the form

$$U(M) = \delta + (\alpha M + \beta N + \gamma P).\lambda, \tag{4.1}$$

then $UU^\dagger$ is a multiple of the identity if

$$-i\alpha = \beta(2A + XE),$$
$$-i\gamma = -\beta E,$$
$$-i\delta = \alpha A - \tfrac{1}{6}Y\beta E - \tfrac{1}{4}Z\gamma E, \tag{4.2}$$

and the real functions $A$ and $E$ are related by

$$2AE + \tfrac{1}{2}XE^2 = 1. \tag{4.3}$$

*Proof*: This is achieved by direct use of the results of section three to simplify $UU^\dagger$.

The simplest special case found is a close relative of result III for U3. To reach it, set $E = \frac{1}{2}$, find $A = 1 - \frac{1}{8}X$, and change $M$ to $-2M$. We thus obtain

VI.

$$\omega U = 1 - \tfrac{1}{4}X^2 + \tfrac{1}{2}Z - \tfrac{1}{3}iY + 2i\lambda.M(1 + \tfrac{1}{2}X) - 2\lambda.N - 2i\lambda.P,$$
$$|\omega|^2 = [(1 + \tfrac{1}{2}X)^2 - \tfrac{1}{2}Z]^2 + 4Y^2/9. \tag{4.4}$$

## 5. THE COSET SPACE SU($N+1$)/U$N$

It is easy to discuss a general parametrization of the space, one in which $U(K) \in$ SU($N + 1$) is written in terms of set of $N$ complex scalar fields, the Goldstone fields of a nonlinear realization of SU($N + 1$) in which a U$N$ subgroup is realized linearly.

Thus consider

$$U(K) = \begin{pmatrix} e + hK\overline{K} & fK \\ -f\overline{K} & g \end{pmatrix}, \tag{5.1}$$

in which $K$ is a one column matrix of $N$ complex fields, $K\bar{K}$ is an $N\times N$ matrix, and $e, f, g$, and $h$ are real functions of the single U$N$ invariant $X=\bar{K}K$ that can be built out of $K$. It is easy to verify

VII. $U(K)$, given by (5.1), belongs to SU$(N+1)$ for arbitrary $f(X)$, if

$$e=1,\quad g=(1-Xf^2)^{1/2},\qquad h=-f^2(1+g)^{-1}. \tag{5.2}$$

Simple special cases arise by appropriate choice of $f$.

VIII. If $f=\alpha^{-1}, \alpha^2=1+X$, then $U(K)\equiv U_1(K)$, where

$$\alpha U_1(K)=\begin{pmatrix}\alpha\mathbb{1}-(1+\alpha)^{-1}K\bar{K} & K\\ -\bar{K} & 1\end{pmatrix}. \tag{5.3}$$

IX. If $f=2/1+X$, then $U(K)\equiv U_2(K)=U_1(K)^2$, where

$$(1+X)U_2(K)=\begin{pmatrix}(1+X)\mathbb{1}-2K\bar{K} & 2K\\ -2\bar{K} & 1-X\end{pmatrix}. \tag{5.4}$$

These results can also be obtained for the case of $N=3$ from the results of Sec. 2. For this purpose, set

$$M_a=0,\quad a=1,2,3,8,$$
$$M_\alpha=K_\alpha,\quad \alpha=4,5,6,7,$$

and make contact with the notation of (5.1) by setting

$$i\lambda.M=i\lambda_\alpha K_\alpha=\begin{pmatrix}0 & K\\ -\bar{K} & 0\end{pmatrix}. \tag{5.5a}$$

Then one calculates

$$N_\alpha=0,\quad N.\lambda=\begin{pmatrix}K\bar{K}-\frac{2}{3}X\mathbb{1} & 0\\ 0 & -\frac{1}{3}X\end{pmatrix},$$
$$X=\bar{K}K,\quad Y=0. \tag{5.5b}$$

Insertion of (5.5) into results IV and III of Sec. 2 reproduces the results VIII and IX of this section for $N=3$. Similarly one relates results of Sec. 4 to the results of this section when $N=4$.

## 6. THE COSET SPACE $U(N+r)/(UN\times Ur)$

The required parametrizations are obtained by expressing $U\in Un$, $n=N+r$, as a function of $2Nr$ real or $Nr$ complex scalar fields, for example the Goldstone fields $K$ of a nonlinear realization of U$n$ in which U$N\times$U$r$ is realized linearly. The aim is not to achieve results of the greatest generality but rather to discuss parametrizations which are either manifestly simple in appearance of else known to underly specially simple versions of the theories in which they occur.

Aiming to build unitary matrices in the form,

$$\begin{pmatrix}J & M\\ -M^\dagger & L\end{pmatrix},$$

where $J$, $M$, and $L$ are respectively $N\times N$, $N\times r$ and $r\times r$ matrices, a natural guess would be to write $M=K$ and determine $J$ and $L$ in terms of $K$ in such a way as to achieve unitarity. In discussing $U(n,r)$ $\sigma$-models[7] it was found that, while this choice was manageable, a choice, vastly superior choice for fields $K$ either in two or four dimensions was achieved, by study of self-duality equations in two dimensions.

Thus, set

$$U(K)=\begin{pmatrix}J & KL\\ -LK^\dagger & L\end{pmatrix}, \tag{6.1}$$

and impose

$$J^2+KL^2K^\dagger=1, \tag{6.2a}$$
$$JK=KL, \tag{6.2b}$$
$$L^2+LK^\dagger KL=1, \tag{6.2c}$$

to make $U(K)$ unitary. In general below, we assume $r<N$ without providing those minor changes that are required when $r=N$. For given $r$, $K^\dagger K$ obeys a Cayley–Hamilton equation $f(K^\dagger K)=0$, where $f(x)$ is a polynomial in $x$ of degree $r$. It then follows that the Cayley–Hamilton equation of $KK^\dagger$ is a polynomial of degree $r+1$, in general. Likewise $J$ and $L$ are polynomials respectively of degree $r+1$ in $KK^\dagger$, and $r$ in $K^\dagger K$. Further (6.2c) can be written $L^2(1+K^\dagger K)=1$.

We illustrate by consideration of the case $r=2$. Then $K^\dagger K$ obeys

$$(K^\dagger K)^2=SK^\dagger K-\Delta, \tag{6.3}$$

where the U$N\times$U$r$ invariants $S$ and $\Delta$ are given by

$$S=\mathrm{Tr}K^\dagger K, \tag{6.4a}$$
$$\Delta=\det K^\dagger K, \tag{6.4b}$$

and $KK^\dagger$, in consequence of (6.3), obeys

$$(KK^\dagger)^3=S(KK^\dagger)^2-\Delta KK^\dagger. \tag{6.5}$$

It is now easy to solve (6.2), initially for $L^2$ and $J^2$, obtaining

$$L^2=a+bK^\dagger K, \tag{6.6a}$$
$$J^2=1-aKK^\dagger-b(KK^\dagger)^2, \tag{6.6b}$$

where

$$a=(S+1)/\Gamma,\quad b=-1/\Gamma,\quad \Gamma=\Delta+S+1. \tag{6.6c}$$

To make (6.1) an explicit parametrization, we must now obtain $L$ and $J$ from $L^2$ and $J^2$. It is easy to complete this calculation; as the result is quite complicated, we do not display it. A manifestly simple parametrization of U$(N+r)/$(U$N\times$U$r$) can be obtained by considering $\hat{U}(K)=U(K)^2$, which obviously is unitary if $U(K)$ itself is. This yields

X.

$$\hat{U}(K)=\begin{pmatrix}1-2KL^2K^\dagger & 2KL^2\\ -2L^2K^\dagger & 2L^2-1\end{pmatrix}, \tag{6.7}$$

belongs to U$n$, and is given explicitly as a function of $K$ with the aid of $L^2$, given by (6.6a) and (6.6c) alone.

The parametrization (6.7) is notable because it does not involve any square roots.

In the special case of $N=r=2$, (6.7) is replaced by

XI.

$$\hat{U}(K)=\frac{1}{\Gamma}\times\begin{pmatrix}S+1-\Delta-2KK^\dagger & 2(S+1)K-2KK^\dagger K\\ -2(S+1)K^\dagger+2K^\dagger KK^\dagger & S+1-\Delta-2K^\dagger K\end{pmatrix}. \tag{6.8}$$

This result also arises naturally from results for U4 matrices. To see this, introduce a set of SU4 $\lambda$ matrices in obvious generalization of the Gell–Mann $\lambda$-matrices of SU3. Specialize the SU4 adjoint vector $M_i$ by setting

$$M_a = 0, \quad M_\alpha = K_\alpha,$$

where $a$ labels the values 1,2,3,8,13,14,15, corresponding to the generators of S(U2 $\times$ U2), of $i$, and $\alpha$ labels the rest. To connect $K_\alpha$ to the matrix $K$ of Goldstone fields set

$$iM.\lambda = \begin{pmatrix} 0 & K \\ -K^\dagger & 0 \end{pmatrix}. \tag{6.9a}$$

Then we calculate

$$N_\alpha = 0,$$

$$N.\lambda = \begin{pmatrix} KK^\dagger - \tfrac{1}{2}S & 0 \\ 0 & K^\dagger K - \tfrac{1}{2}S \end{pmatrix}, \tag{6.9b}$$

$$X = S, \quad Y = 0, \quad Z = N_i N_i = \tfrac{1}{2}(S^2 - 4\Delta), \quad P_a = 0, \tag{6.9c}$$

$$iP.\lambda + \tfrac{1}{2}SiM.\lambda = \begin{pmatrix} 0 & KK^\dagger K \\ -K^\dagger K K^\dagger & 0 \end{pmatrix}. \tag{6.9d}$$

Now insertion of the four results (6.9) into (4.4) reproduces (6.8). The U4 parametrization that reduces to $U(K)$ using (6.9) has not been found.


[1] P. Chang and F. Gursey, Phys. Rev. **164**, 1752 (1967).
[2] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1968).
[3] A.J. Macfarlane, A. Sudbery, and P.H. Weisz, Proc. R. Soc. London A **314**, 217 (1970).
[4] K.J. Barnes, P.H. Dondi, and S.C. Sarker, Proc. R. Soc. London A **330**, 389 (1972).
[5] A. Belavin and A.M. Polyakov, JETP Lett. **22**, 245 (1975).
[6] M. Lüscher, Phys. Lett. B **78**, 465 (1978).
[7] A.J. Macfarlane, Phys. Lett. B **82**, 239 (1979).
[8] A.J. Macfarlane, Nuc. Phys. B **152**, 145 (1979).
[9] E. Cremmer and J. Scherk, Phys. Lett. B **74**, 341 (1978).
[10] A. Salam and J. Strathdee, Phys. Rev. **184**, 1750 (1969).
[11] K.J. Barnes and R. Delbourgo, J. Phys. A **5**, 1043 (1972).
[12] A. Sudbery, Ph.D. Thesis, Cambridge University, 1968.
[13] A.J. Macfarlane, A. Sudbery, and P.H. Weisz, Commun. Math. Phys. **11**, 77 (1968).

# Field fluctuations in a two-phase random medium

Mark J. Beran [a)]

*Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

(Received 27 August 1979; accepted for publication 13 June 1980)

We consider here the problem of determining the mean square fluctuations in a statistically homogeneous isotropic two-phase dielectric random medium. An expression is derived for a weighted sum of the mean square fluctuations in each phase in terms of the effective dielectric constant. From this expression bounds are derived for the mean square fluctuations in each phase. An assumption is then made to allow us to obtain exact expressions for the mean square fluctuations in a particular phase.

## 1. INTRODUCTION

In a dielectric medium the electric and displacement fields (**E** and **D** respectively) are governed by the equations

$$\nabla\cdot\epsilon\mathbf{E} = 0\,, \tag{1}$$

$$\nabla\times\mathbf{E} = 0\,, \tag{2}$$

$$\mathbf{D} = \epsilon\mathbf{E}\,, \tag{3}$$

where $\epsilon$ is considered to be a random function of position. (The same equations govern heat conductivity, electrical conductivity, etc.)

The effective dielectric constant $\epsilon^*$, for an infinite homogeneous isotropic random medium is defined by the relation

$$\langle\mathbf{D}\rangle = \epsilon^*\langle\mathbf{E}\rangle\,, \tag{4}$$

where the brackets $\langle\ \rangle$, denote an ensemble average, $\langle\mathbf{D}\rangle$ and $\langle\mathbf{E}\rangle$ are constant vectors. (We note that $\epsilon^*$ may also be defined by $\langle\mathbf{D}\cdot\mathbf{E}\rangle = \epsilon^*\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle$. It may be shown[1] that $\langle\mathbf{D}\cdot\mathbf{E}\rangle = \langle\mathbf{D}\rangle\cdot\langle\mathbf{E}\rangle$ for an infinite homogeneous isotropic random medium.) The theory of effective constants has been investigated by many authors[1-6] in great detail.

Here we wish to determine information about the mean square field fluctuations in terms of $\epsilon^*$. For a general medium, the variance of the **E** field is defined as

$$\sigma_E^2 = \frac{\langle(\mathbf{E}-\langle\mathbf{E}\rangle)\cdot(\mathbf{E}-\langle\mathbf{E}\rangle)\rangle}{\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle}\,, \tag{5}$$

with a similar expression for $\sigma_D^2$.

A two-phase medium is characterized by the phase dielectric constants $\epsilon_1$ and $\epsilon_2$ and volume fractions $v_1$ and $v_2$ ($v_1 + v_2 = 1$). We choose here $\epsilon_1 > \epsilon_2$. The geometry of the medium may be characterized by a probability density functional or the $n$ moments $\langle\epsilon(\mathbf{x}_1)\cdots\epsilon(\mathbf{x}_n)\rangle$ for all $n$. We define here mean square field fluctuations in a particular phase (i) as

$$\sigma_{E_i}^2 = \frac{\langle(\mathbf{E}_i-\langle\mathbf{E}_i\rangle)\cdot(\mathbf{E}_i-\langle\mathbf{E}_i\rangle)\rangle}{\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle} \tag{6}$$

(no summation on $i$).

The $\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle$ normalization in the denominator may be replaced by $\langle\mathbf{E}_i\rangle\cdot\langle\mathbf{E}_i\rangle$ when desired, since as we point out below $\langle\mathbf{E}_i\rangle$ and $\langle\mathbf{E}\rangle$ are simply related. Later in the paper we shall consider the alternative normalization. There is a similar definition for $\sigma_{D_i}^2$.

In Sec. 2 we shall derive the equation

$$\frac{v_1\epsilon_1}{\langle\epsilon\rangle}\sigma_{E_1}^2 + \frac{v_2\epsilon_2}{\langle\epsilon\rangle}\sigma_{E_2}^2 = \frac{\epsilon_1\epsilon_2}{v_1v_2(\epsilon_1-\epsilon_2)^2} \times\left(1-\frac{\epsilon^*}{\langle\epsilon\rangle}\right)\left(\epsilon^*\left\langle\frac{1}{\epsilon}\right\rangle - 1\right) \tag{7}$$

and a similar one for the **D** field.

In Sec. 3 we shall find bounds for $\sigma_{E_i}^2$ and $\sigma_{D_i}^2$ and present an assumption which allows us to obtain $\sigma_{E_i}^2$ or $\sigma_{D_i}^2$ exactly. The importance of the quantities $\epsilon^*/\langle\epsilon\rangle$ and $\epsilon^*(\langle 1/\epsilon\rangle)$ will be emphasized.

## 2. DERIVATION OF EQ. (7)

In a previous paper,[7] we derived the equation

$$(\langle\epsilon\rangle - \epsilon^*)\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle = \langle\epsilon(\mathbf{E}-\langle\mathbf{E}\rangle)\cdot(\mathbf{E}-\langle\mathbf{E}\rangle)\rangle\,, \tag{8}$$

and a similar one for **D** with $\epsilon$ replaced by $\alpha = 1/\epsilon$. In that paper, Eq. (8) was used to obtain bounds for $\sigma_E^2$ and $\sigma_D^2$. Mendelsohn[8] also has recently treated the problem of bounds.

We shall now derive Eq. (7) from Eq. (8). For a two-phase material

$$\langle\epsilon(\mathbf{E}-\langle\mathbf{E}\rangle)\cdot(\mathbf{E}-\langle\mathbf{E}\rangle)\rangle = \epsilon_1v_1\langle(\mathbf{E}_1-\langle\mathbf{E}\rangle)\cdot(\mathbf{E}_1-\langle\mathbf{E}\rangle\rangle + \epsilon_2v_2\langle(\mathbf{E}_2-\langle\mathbf{E}\rangle)\cdot(\mathbf{E}_2-\langle\mathbf{E}\rangle)\rangle\,. \tag{9}$$

Noting that $\langle\mathbf{E}\rangle = v_1\langle\mathbf{E}_1\rangle + v_2\langle\mathbf{E}_2\rangle$, we find from Eqs. (8) and (9) after some algebra

$$\left(1-\frac{\epsilon^*}{\langle\epsilon\rangle}\right) = \frac{v_1\epsilon_1}{\langle\epsilon\rangle}\sigma_{E_1}^2 + \frac{v_2\epsilon_2}{\langle\epsilon\rangle}\sigma_{E_2}^2 + \frac{v_1v_2}{\langle\epsilon\rangle}\frac{(\epsilon_1v_2+\epsilon_2v_1)(\langle\mathbf{E}_1\rangle-\langle\mathbf{E}_2\rangle)\cdot(\langle\mathbf{E}_1\rangle-\langle\mathbf{E}_2\rangle)}{\langle\mathbf{E}\rangle\cdot\langle\mathbf{E}\rangle}\,. \tag{10}$$

The term $\langle\mathbf{E}_1\rangle - \langle\mathbf{E}_2\rangle$ may be simplified by noting the relations

$$\langle\mathbf{E}\rangle = v_1\langle\mathbf{E}_1\rangle + v_2\langle\mathbf{E}_2\rangle \tag{11}$$

[a)]Permanent address: School of Engineering, Tel Aviv University, Tel-Aviv, Israel.

 0022-2488/80/102583-03$1.00  

$$\langle \mathbf{D} \rangle = \epsilon^* \langle \mathbf{E} \rangle = v_1 \langle \mathbf{D}_1 \rangle + v_2 \langle \mathbf{D}_2 \rangle$$
$$= v_1 \epsilon_1 \langle \mathbf{E} \rangle + v_2 \epsilon_2 \langle \mathbf{E}_2 \rangle . \quad (12)$$

From Eqs. (11) and (12) we find

$$\langle \mathbf{E}_1 \rangle = \frac{(\epsilon^* - \epsilon_2)}{v_1(\epsilon_1 - \epsilon_2)} \langle \mathbf{E} \rangle , \quad (13a)$$

$$\langle \mathbf{E}_2 \rangle = \frac{(\epsilon_1 - \epsilon^*)}{v_2(\epsilon_1 - \epsilon_2)} \langle \mathbf{E} \rangle , \quad (13b)$$

$$\langle \mathbf{E}_1 \rangle - \langle \mathbf{E}_2 \rangle = \frac{1}{v_1 v_2 (\epsilon_1 - \epsilon_2)} (\epsilon^* - \langle \epsilon \rangle) \langle \mathbf{E} \rangle . \quad (13c)$$

Substituting Eq. (13c) into Eq. (10) yields Eq. (7) after further simplification.

An analogous procedure for the $D$ field yields

$$\frac{v_1 \alpha_1}{\langle \alpha \rangle} \sigma_{D_1}^2 + \frac{v_2 \alpha_2}{\langle \alpha \rangle} \sigma_{D_2}^2 = \frac{\alpha_1 \alpha_2}{v_1 v_2 (\alpha_1 - \alpha_2)^2} \times \left(1 - \frac{\alpha^*}{\langle \alpha \rangle}\right) \cdot \left(\alpha^* \left\langle \frac{1}{\alpha} \right\rangle - 1\right) \quad (14)$$

where

$$\alpha_i \equiv 1/\epsilon_i ,$$
$$\alpha^* \equiv 1/\epsilon^* .$$

When $v_i \epsilon_1 \gg v_j \epsilon_2$ ($i = 1,2; j = 1,2$), Eqs. (7) and (14) become respectively:

$$\sigma_{E_1}^2 + \frac{v_2 \epsilon_2}{v_1 \epsilon_2} \sigma_{E_2}^2 = \frac{\epsilon_2}{v_1 v_2 \epsilon_1} \left(1 - \frac{\epsilon^*}{v_1 \epsilon_1}\right) \left(\frac{v_2 \epsilon^*}{\epsilon_2} - 1\right), \quad (15)$$

$$\frac{v_1}{v_2} \frac{\alpha_1}{\alpha_2} \sigma_{D_1}^2 + \sigma_{D_2}^2 = \frac{\alpha_1}{v_1 v_2 \alpha_2} \left(1 - \frac{\alpha^*}{v_2 \alpha_2}\right) \left(\frac{v_1 \alpha^*}{\alpha_1} - 1\right). \quad (16)$$

## 3. BOUNDS AND EXACT EXPRESSIONS FOR $\sigma_{E_i}^2$, $\sigma_{D_i}^2$, $\tilde{\sigma}_{E_i}^2$ AND $\tilde{\sigma}_{D_i}^2$

### 3.1 Bounds

In Eq. (7) both terms on the left-hand side are positive. Thus, each term is less than the right-hand side and we find the upper bounds

$$\sigma_{E_1}^2 \leqslant \frac{\epsilon_2 \langle \epsilon \rangle}{v_1^2 v_2 (\epsilon_1 - \epsilon_2)^2} \left(1 - \frac{\epsilon^*}{\langle \epsilon \rangle}\right) \left(\epsilon^* \left\langle \frac{1}{\epsilon} \right\rangle - 1\right), \quad (17)$$

$$\sigma_{E_2}^2 \leqslant \frac{\epsilon_1 \langle \epsilon \rangle}{v_1 v_2^2 (\epsilon_1 - \epsilon_2)^2} \left(1 - \frac{\epsilon^*}{\langle \epsilon \rangle}\right) \left(\epsilon^* \left\langle \frac{1}{\epsilon} \right\rangle - 1\right). \quad (18)$$

Similarly for the $D$ field we have

$$\sigma_{D_1}^2 \leqslant \frac{\alpha_2 \langle \alpha \rangle}{v_1^2 v_2 (\alpha_1 - \alpha_2)^2} \left(1 - \frac{\alpha^*}{\langle \alpha \rangle}\right) \left(\alpha^* \left\langle \frac{1}{\alpha} \right\rangle - 1\right), \quad (19)$$

TABLE I. Upper bounds for $\sigma_{E_i}^2$ and $\sigma_{D_i}^2$ for HS limits of $\epsilon^*$ ($v_i \epsilon_1 \gg v_j \epsilon_2$).

| | $\epsilon_L^*$ | $\epsilon_U^*$ |
|---|---|---|
| $\sigma_{E_1}^2 \leqslant$ | $\frac{\epsilon_2}{\epsilon_1} \frac{2}{v_2}$ | $\frac{2v_2}{(3 - v_1)^2}$ |
| $\sigma_{E_2}^2 \leqslant$ | $\frac{2v_1}{v_2^2}$ | $\frac{\epsilon_1}{\epsilon_2} \frac{2v_1}{(3 - v_1)^2}$ |
| $\sigma_{D_1}^2 \leqslant$ | $\frac{\epsilon_1}{\epsilon_2} \frac{2v_2}{(1 + 2v_1)^2}$ | $\frac{v_2}{2v_1^2}$ |
| $\sigma_{D_2}^2 \leqslant$ | $\frac{2v_1}{(1 + 2v_1)^2}$ | $\frac{\epsilon_2}{\epsilon_1} \frac{1}{2v_1}$ |

TABLE II. Upper bounds for $\tilde{\sigma}_{E_i}^2$ and $\tilde{\sigma}_{D_i}^2$ for HS limits of $\epsilon^*$ ($v_i \epsilon_1 \gg v_j \epsilon_2$).

| | $\epsilon_L^*$ | $\epsilon_U^*$ |
|---|---|---|
| $\tilde{\sigma}_{E_1}^2 \leqslant$ | $\frac{2}{9v_2}\left(\frac{\epsilon_1}{\epsilon_2}\right)$ | $\frac{v_2}{2}$ |
| $\tilde{\sigma}_{E_2}^2 \leqslant$ | $2v_1$ | $\frac{2v_1}{9(3 - v_1)^2} \frac{\epsilon_1}{\epsilon_2}$ |
| $\tilde{\sigma}_{D_1}^2 \leqslant$ | $\frac{2}{9} \frac{v_2}{(1 + 2v_1)} \frac{\epsilon_1}{\epsilon_2}$ | $\frac{v_2}{2}$ |
| $\tilde{\sigma}_{D_2}^2 \leqslant$ | $2v_1$ | $\frac{2v_1}{9} \frac{\epsilon_1}{\epsilon_2}$ |

$$\sigma_{D_2}^2 \leqslant \frac{\alpha_1 \langle \alpha \rangle}{v_1 v_2^2 (\alpha_1 - \alpha_2)^2} \left(1 - \frac{\alpha^*}{\langle \alpha \rangle}\right) \left(\alpha^* \left\langle \frac{1}{\alpha} \right\rangle - 1\right). \quad (20)$$

We are most interested in these bounds for $\epsilon_1 \gg \epsilon_2$. Using again the condition $v_i \epsilon_1 \gg v_j \epsilon_2$, which means that we are not considering extremes in either volume fraction, we find

$$\sigma_{E_1}^2 \leqslant \frac{\epsilon_2}{\epsilon_1} \frac{1}{v_1 v_2} \left(1 - \frac{\epsilon^*}{v_1 \epsilon_1}\right) \left(\frac{v_2 \epsilon^*}{\epsilon_2} - 1\right), \quad (21)$$

$$\sigma_{E_2}^2 \leqslant \frac{1}{v_2^2} \left(1 - \frac{\epsilon^*}{v_1 \epsilon_1}\right) \left(\frac{v_2 \epsilon^*}{\epsilon_2} - 1\right), \quad (22)$$

$$\sigma_{D_1}^2 \leqslant \frac{1}{v_1^2} \left(1 - \frac{\alpha^*}{v_2 \alpha_2}\right) \left(\frac{v_1 \alpha^*}{\alpha_1} - 1\right), \quad (23)$$

$$\sigma_{D_2}^2 \leqslant \frac{\alpha_1}{\alpha_2} \frac{1}{v_1 v_2} \left(1 - \frac{\alpha^*}{v_2 \alpha_2}\right) \left(\frac{v_1 \alpha^*}{\alpha_1} - 1\right). \quad (24)$$

The bounds are dependent upon where $\epsilon^*$ lies with respect to $\epsilon_1$ and $\epsilon_2$. Hashin and Shtrikman[3] have shown that

$$1 + \frac{v_1}{1/(\beta - 1) + (1 - v_1)/3} \leqslant \frac{\epsilon^*}{\epsilon_2} \leqslant \beta + \frac{(1 - v_1)}{1/(1 - \beta) + v_1/3\beta}, \quad (25)$$

where $\beta = \epsilon_1/\epsilon_2$.

For $v_i \epsilon_1 \gg v_j \epsilon_2$, Eq. (25) reduces to

$$\epsilon_L^* \equiv \frac{(1 + 2v_1)\epsilon_2}{(1 - v_1)} \leqslant \epsilon^* \leqslant \frac{2v_1}{(3 - v_1)} \epsilon_1 \equiv \epsilon_U^* . \quad (26)$$

In Table I we show how the upper bound for $\sigma_{E_i}^2$ and $\sigma_{D_i}^2$ depends on $\epsilon_L^*$ and $\epsilon_U^*$. Roughly speaking, we note that the lower bound, $\epsilon_L^*$, corresponds to a geometry in which phase (1) materials are inclusions in the matrix of phase (2). The upper bound, $\epsilon_U^*$, corresponds to a geometry where phase (2) materials are inclusions in the matrix of phase (1).

As we stated above, it is also useful to consider the quantity

$$\tilde{\sigma}_{E_i}^2 = \frac{\langle (\mathbf{E}_i - \langle \mathbf{E}_i \rangle) \cdot (\mathbf{E}_i - \langle \mathbf{E}_i \rangle) \rangle}{\langle \mathbf{E}_i \rangle \cdot \langle \mathbf{E}_i \rangle}, \quad (27)$$

with a similar definition for $\tilde{\sigma}_{D_i}^2$. Here the phase fluctuations are normalized with respect to the mean field in the particular phase. The results for $v_i \epsilon_1 \gg v_j \epsilon_2$ are given in Table II.

The bounds in Tables I and II are correct, but the question arises as to whether or not the bounds are restrictive. For example, in Table II we see that four of the bounds depend upon $\epsilon_1/\epsilon_2$. If the bounds are indeed suitably restrictive

(i.e., "good" bounds), we may expect materials for which $\tilde{\sigma}^2_{E_i}(\epsilon^* = \epsilon^*_L)$ may be quite large.

The bounds that do not depend upon $\epsilon_1/\epsilon_2$ seem reasonably restrictive from a physical point of view since we would expect fluctuations in a particular phase to be of the order of magnitude of the mean field in the phase. It is true that for isolated spherical inclusions, or for the composite spheres Hashin and Shtrikman used to prove the realizability of their bounds, there are no fluctuations in the sphere itself. This occurs, however, only because of the special nature of the sphere and in general, fluctuations occur in inclusions.

Without further calculations, numerical solutions or experimental evidence, we cannot at this point answer the question of whether or not materials exist for which, say, $\tilde{\sigma}^2_{E_1} \sim O(\epsilon_1/\epsilon_2)$ when $\epsilon^* = \epsilon^*_L$. We would expect, however, that for a large class of materials this is not so and that $\tilde{\sigma}^2_{E_1}$ and $\tilde{\sigma}^2_{E_2}$ (or $\tilde{\sigma}^2_{D_1}$ and $\tilde{\sigma}^2_{D_1}$) should be of the same order of magnitude. We might expect this to be so, for example, if the geometry of the inclusions was not vastly different from the matrix geometry. In the next section, we shall show the consequences of making the *a priori* assumption that $\tilde{\sigma}^2_{E_1}$ and $\tilde{\sigma}^2_{E_2}$ are both of the same order of magnitude.

### 3.2 Exact expressions for $\sigma^2_{E_i}$, $\sigma^2_{D_i}$, $\tilde{\sigma}^2_{E_i}$, and $\tilde{\sigma}^2_{D_1}$

In this section, we make the assumption that $\tilde{\sigma}^2_{E_1}$ and $\tilde{\sigma}^2_{E_2}$ are of the same order in $\epsilon_1/\epsilon_2$ if $v_2 = O(v_1)$. [This implies that $\tilde{\sigma}^2_{D_1}$ and $\tilde{\sigma}^2_{D_2}$ are also of the same order. We interpret the condition $v_2 = O(v_1)$ here to be $.2 \lesssim v_i \lesssim .8$.] We assume that this is true over the whole range of $\epsilon^*$. That is, we assume that this is true if phase (1) may be considered to be an inclusion phase in the matrix phase (2) ($\epsilon^* \approx \epsilon^*_L$) or vice-versa ($\epsilon^* \approx \epsilon^*_U$).

If we assume that $\tilde{\sigma}^2_{E_1}$ and $\tilde{\sigma}^2_{E_2}$ are the same order of magnitude, then the relative magnitude of the terms

$$\sigma^2_{E_1} \quad \text{and} \quad \frac{v_2\epsilon_2}{v_1\epsilon_1}\sigma^2_{E_2}$$

in Eq. (15) depends strongly on the value of $\epsilon^*$. It may be shown that if $\epsilon^* = O(\epsilon^*_L)$ then

$$\frac{v_2\epsilon_2}{v_1\epsilon_1}\sigma^2_{E_2} \gg \sigma^2_{E_1} \tag{28}$$

and if $\epsilon^* = O(\epsilon^*_U)$

$$\sigma^2_{E_1} \gg \frac{v_2\epsilon_2}{v_1\epsilon_1}\sigma^2_{E_2}. \tag{29}$$

TABLE III. Exact expressions for $\sigma^2_{E_i}$ and $\sigma^2_{D_i}$ $(v_i\epsilon_1 \gg v_j\epsilon_2)$.

| | $\epsilon^* = O(\epsilon^*_L)$ | $\epsilon^* = O(\epsilon^*_U)$ |
|---|---|---|
| $\sigma^2_{E_1}$ | | $\frac{\epsilon^*}{v_1\epsilon_1}\left(1 - \frac{\epsilon^*}{v_1\epsilon_1}\right)$ |
| $\sigma^2_{E_2}$ | $\frac{1}{v_2^2}\left(\frac{v_2\epsilon^*}{\epsilon_2} - 1\right)$ | |
| $\sigma^2_{D_1}$ | | $\frac{1}{v_1^2}\left(\frac{v_1\epsilon_1}{\epsilon^*} - 1\right)$ |
| $\sigma^2_{D_2}$ | $\frac{1}{v_2}\frac{\epsilon_2}{\epsilon^*}\left(1 - \frac{\epsilon_2}{v_2\epsilon^*}\right)$ | |

TABLE IV. Exact expressions for $\tilde{\sigma}^2_{E_i}$ and $\tilde{\sigma}^2_{D_i}$ $(v_i\epsilon_1 \gg v_j\epsilon_2)$.

| | $\epsilon^* = O(\epsilon^*_L)$ | $\epsilon^* = O(\epsilon^*_U)$ |
|---|---|---|
| $\tilde{\sigma}^2_{E_1}$ | | $\frac{v_1\epsilon_1}{\epsilon^*}\left(1 - \frac{\epsilon^*}{v_1\epsilon_1}\right)$ |
| $\tilde{\sigma}^2_{E_2}$ | $\left(\frac{v_2\epsilon^*}{\epsilon_2} - 1\right)$ | |
| $\tilde{\sigma}^2_{D_1}$ | | $\left(\frac{v_1\epsilon_1}{\epsilon^*} - 1\right)$ |
| $\tilde{\sigma}^2_{D_2}$ | $\frac{v_2\epsilon^*}{\epsilon_2}\left(1 - \frac{\epsilon_2}{v_2\epsilon^*}\right)$ | |

If then $\epsilon^* = O(\epsilon^*_L)$ we find from Eq. (15)

$$\sigma^2_{E_2} = \frac{1}{v_2^2}\left(\frac{v_2\epsilon^*}{\epsilon_2} - 1\right), \tag{30}$$

while if $\epsilon^* = O(\epsilon^*_U)$

$$\sigma^2_{E_1} = \frac{\epsilon^*}{v_1\epsilon_1}\left((1 - \frac{\epsilon^*}{v_1\epsilon_1}\right). \tag{31}$$

That is, in these limits we can obtain exact expressions for $\sigma^2_{E_2}$ or $\sigma^2_{E_1}$. Similar expressions can be obtained for $\sigma^2_{D_1}$ or $\sigma^2_{D_2}$. If on the other hand $\epsilon^* = O(\sqrt{\epsilon_1\epsilon_2})$, both terms are of the same order of magnitude and no exact expressions can be obtained. In Tables III and IV, we present exact expressions for the variances.

It is interesting to note that we only obtain exact expressions when $\sigma^2$ and $\tilde{\sigma}^2$ are of the same order of magnitude. That is, for that phase in which the mean phase field is of the same order of magnitude as the mean field.

In closing this section, we want to emphasize again that the results in Tables III and IV are exact results only if the assumptions $\tilde{\sigma}^2_{E_1} = O(\tilde{\sigma}^2_{E_2})$ and $\tilde{\sigma}^2_{D_1} = O(\tilde{\sigma}^2_{D_1})$ are correct. The validity of this assumption has yet to be firmly established for any class of materials.

## 4. CONCLUSIONS

In Eqs. (17)–(24), we present bounds for the quantities $\sigma^2_{E_i}$ and $\sigma^2_{D_i}$. Using Eqs. (13a)–(13c) and similar expressions for $\langle D_1\rangle$ and $\langle D_2\rangle$ are similar bounds can be obtained for $\tilde{\sigma}^2_{E_i}$ and $\tilde{\sigma}^2_{D_i}$.

When $\epsilon^* = \epsilon^*_L$ (the lower HS bound) or $\epsilon^* = \epsilon^*_U$ (the upper HS bound), bounds for $\sigma^2$ and $\tilde{\sigma}^2$ are given in Tables I and II when $v_i\epsilon_1 \gg v_j\epsilon_2$.

If $\tilde{\sigma}^2_{E_1} = O(\tilde{\sigma}^2_{E_2})$ exact expressions for $\sigma^2$ and $\tilde{\sigma}^2$ are given in Tables III and IV when $\epsilon^* = O(\epsilon^*_L)$ or $\epsilon^* = (\epsilon^*_U)$ for $v_i\epsilon_1 \gg v_j\epsilon_2$. The range of validity of this assumption has not yet been firmly established.


[1] M. Beran, *Statistical Continuum Theories* (Wiley, New York, 1968).
[2] W. F. Brown, J. Chem. Phys. **23**, 1514 (1955).
[3] Z. Hashin and S. Shtrikman, J. Appl. Phys. **33**, 3125 (1962).
[4] M. Hori, J. Math. Phys. **18**, 487 (1977); and six previous articles.
[5] M. N. Miller, J. Math. Phys. **10**, 1988 (1969).
[6] M. ElSayed and J. J. McCoy, Fibre Sci. Technol. **7**, 281 (1974).
[7] M. Beran, J. Appl. Phys. **39**, 5712 (1968).
[8] K. S. Mendelson, J. Appl. Phys. **48**, 25 (1977).

# Localized nonuniform patterns in a diffusion-reaction model with autocatalysis and the Langmuir–Hinshelwood saturation law

L. L. Bonilla and M. G. Velarde
*Departamento de Física de Fluidos, Universidad Autónoma de Madrid, Cantoblanco (Madrid) Spain*

(Received 17 July 1979; accepted for publication 26 October 1979)

The WKB (Wentzel-Kramers–Brillouin) method is used to predict the onset of *localized* dissipative structures in a one-dimensional reactor where a diffusion-reaction process with autocatalysis and the Langmuir-Hinshelwood (Michaelis-Menten, Holling) saturation law takes place.

## 1. INTRODUCTION

In reactor kinetics one usually deals with diffusion-reaction processes which involve autocatalytic pathways and/or steps obeying saturation laws, like the Langmuir–Hinshelwood law (see Ref. 1 for a comprehensive and recent monograph). The latter law also corresponds to the Michaelis–Menten kinetics in enzyme-controlled biochemical reactions, and to the Holling law in some population dispersion and competition problems in ecology. In a previous paper[2] (hereafter called I) the present authors have considered the onset of nonuniform distribution of reactants in a one-dimensional vessel. They restricted consideration, however, to the case where the structure spreads over the whole container. In the present note we extend the analysis to the case of dissipative structures appearing only as patches in certain parts of the reactor.[3] The model problem considered in I contains three relevant reactants $A$, $X$, and $Y$ whose spatio-temporal evolution is governed by the following set of partial differential equations:

$$\frac{\partial A}{\partial t} = -A + D_A \frac{\partial^2 A}{\partial r^2}, \tag{1.1a}$$

$$\frac{\partial X}{\partial t} = XY - X(1+qX)^{-1} + \theta D \frac{\partial^2 X}{\partial r^2}, \tag{1.1b}$$

$$\frac{\partial Y}{\partial t} = A - XY + D \frac{\partial^2 Y}{\partial r^2}, \tag{1.1c}$$

that for a one-dimensional geometry ($0 \leqslant r \leqslant 1$) we have written in dimensionless form. $D$ refers to the diffusion of $Y$, $D = D_Y$. $\theta$ is the ratio of the diffusion coefficient of $X$ to $D_Y$. $\theta D_Y = D_X$. $D_A$ is the diffusion coefficient of $A$. We shall take $D_A > D, D_X$ and, thus, and adiabatic elimination of $A$ will be made. Using Dirichlet's boundary conditions (b.c.)

$$A(r=0,t) = A(r=1,t) = \bar{A}, \tag{1.2}$$

where $\bar{A}$ is some given constant, Eq. (1.1a) has the following nonuniform steady solution

$$A(r) = \bar{A}\,\mathrm{sech}(2D_A^{1/2})^{-1}\cosh(r-\tfrac{1}{2})D_A^{-1/2}. \tag{1.3}$$

For simplicity we choose the following b.c. for (1.1b) and (1.1c)

$$X(0,t) = X(1,t) = \bar{A}/(1-q\bar{A}) \tag{1.4a}$$

$$Y(0,t) = Y(1,t) = 1 - q\bar{A}. \tag{1.4b}$$

This choice elimitates the appearance of diffusional boundary layers at both extremes of the reactor line. Thus, for low enough values of $D$, the system (1.1) possesses the following steady solution

$$X_s(r) = A(r)/[1-qA(r)]^{-1} + \{qA''(r) - [\theta A(r)/[1-qA(r)]]'\}D/[1-qA(r)] + O(D^2), \tag{1.5a}$$

$$Y_s(r) = 1 - qA(r) - \{2q[1-qA(r)]A''(r)/A(r) - \theta[A(r)/[1-qA(r)]]'\}D + O(D^2). \tag{1.5b}$$

Here and later on a prime ($'$) over a quantity merely denotes differentiation with respect to the spatial variable, $r$.

## 2. STABILITY OF THE PRIMARY NONUNIFORM STEADY DISTRIBUTION OF REACTANTS

To test the stability of (1.5) we consider disturbances $x$ and $y$ upon $X_s$ and $Y_s$. To the first approximation we have

$$\left(\frac{\partial}{\partial t} - L(\gamma,r)\right)u = 0, \tag{2.1}$$

where

$$u = (x,y), \tag{2.2a}$$

$$L(\gamma,r) \equiv \begin{pmatrix} Y_s - [1+qX_s(r)]^{-2} + \theta D\dfrac{\partial^2}{\partial r^2} & X_s(r) \\ -Y_s(r) & -X_s(r) + D\dfrac{\partial^2}{\partial r^2} \end{pmatrix}. \tag{2.2b}$$

Here the b.c. are

$$x(0,t) = x(1,t) = y(0,t) = y(1,t) = 0. \tag{2.2c}$$

Posing

$$u(r,t) = e^{\sigma t}\Xi(r), \tag{2.3}$$

with $\Xi \equiv (\xi(r), \eta(r))$, Eq. (2.1) leads to the following eigenvalue problem

$$[L(\gamma,r) - \sigma I]\Xi(r) = 0, \tag{2.4}$$

with the b.c. (2.2c). It should be noted that the coefficients in Eq. (2.4) depend on $A(r)$ which are *slow* varying quantities in the *fast* varying spatial scale where $X$ and $Y$ change. Thus the WKB method[4] can be used to solve (2.4). We pose

$$\Xi(r) \sim e^{i\omega(r,D^{1/2})/D^{1/2}} F(r,D^{1/2}), \tag{2.5a}$$

where $w$ and $F$ are slow varying functions over the scale $D^{1/2}$. For a self-consistent approach we shall assume that

 0022-2488/80/102586-04$1.00  

$$w(r,D^{1/2}) = w(r) \tag{2.5b}$$

and

$$F(r,D^{1/2}) = F_0(r) + D^{1/2}F_1(r) + DF_2(r) + O(D^{3/2}). \tag{2.5c}$$

Insertion of (2.5) into (2.4) yields the following hierarchy of equations:

$$HF_0 = 0, \tag{2.6a}$$

$$HF_1 = \begin{pmatrix} i\theta(w'' f_0 + 2w' f_0') \\ i(w'' g_0 + 2w' g_0') \end{pmatrix}, \tag{2.6b}$$

$$\cdots\cdots,$$

where $F_0 = (f_0, g_0)$, and

$$H = \begin{pmatrix} \sigma + \theta w'^2 - qA(1-qA) & -A/(1-qA) \\ 1 - qA & \sigma + A/(1-qA) + w'^2 \end{pmatrix}. \tag{2.7}$$

A necessary and sufficient condition for the existence of a nontrivial solution to (2.6a) is that (2.7) be singular. The vanishing of the determinant of (2.7) leads to a second-order equation in $w'^2$ whose solution is

$$\begin{aligned} 2\theta w'^2 = &-\left[\theta\left(\sigma + \frac{A}{1-qA}\right) + \sigma - qA(1-qA)\right] \\ &\pm \left\{\left[\theta\left(\sigma + \frac{A}{1-qA}\right) + \sigma - qA(1-qA)\right]^2 \right. \\ &- 4\theta\left[\sigma\left(\sigma + \frac{A}{1-qA} - qA(1-qA)\right)\right. \\ &\left.\left. + A(1-qA)\right]\right\}^{1/2}. \end{aligned} \tag{2.8}$$

On the other hand, if we choose an appropriate constant, $c$, the following relation holds:

$$(1-qA)c + \sigma + \theta w'^2 - qA(1-qA) = 0. \tag{2.9}$$

We also have

$$(1-qA)f_0 + \left(\sigma + w'^2 + \frac{A}{1-qA}\right)g_0 = 0 \tag{2.10a}$$

and

$$w''(\theta f_0 + c g_0) + 2w'(\theta f_0 + c g_0)' = 0. \tag{2.10b}$$

The latter equation (2.10b) yields

$$(\theta f_0 + c g_0)w'^{1/2} = \text{const.} \tag{2.11a}$$

Introducing (2.11a) into (2.10a) we get

$$\begin{aligned} g_0 = \text{const}\ & w'^{-1/2}/\{c - \theta[(\sigma + w'^2)(1-qA) + A] \\ & /(1-qA)^2\}. \end{aligned} \tag{2.11b}$$

Thus, up to some constants which depend on the b.c. of the problem, both $f_0$ and $g_0$ are determined. However, for given b.c. not all values of $\sigma$ will allow a nontrivial solution. Rather $\sigma$ is the eigenvalue of the problem to be solved.

Notice that the WKB method is valid provided the following relation does *not* hold

$$c = \theta[(\sigma + w'^2)(1-qA) + A](1-qA)^{-2}. \tag{2.12}$$

Otherwise $f_0$ and $g_0$ will be ill-defined and the asymptotic representation (2.5) will not be acceptable.

The values of $r$ at which (2.12) holds are the *turning* points in the WKB method. Using (2.9) these *turning* points are the solutions of the equation

$$2\theta w'^2 + \theta[\sigma + A/(1-qA)] + \sigma - qA(1-qA) = 0. \tag{2.13}$$

Thus the *turning* points make zero the discriminant in (2.8). We have

$$\begin{aligned} &\left[\theta\left(\sigma + \frac{A}{1-qA}\right) + \sigma - qA(1-qA)\right]^2 \\ &= 4\theta\left[\sigma\left(\sigma + \frac{A}{1-qA} - qA(1-qA)\right) + A(1-qA)\right], \end{aligned} \tag{2.14}$$

which for $\sigma = 0$ (strictly speaking for $\mathrm{Re}\sigma = 0$), yields the turning points located on the marginal stability curve for the onset of non-uniform spatial patterns.[5]

## 3. LOCALIZED STRUCTURES ALONG THE LINE REACTOR

For $\sigma = 0$, the solutions to (2.14) are the roots of the polynomial

$$\begin{aligned} Z(q) = q^4 + \frac{4\theta A - 2}{A}q^3 + \frac{1 - 10\theta A}{A^2}q^2 \\ + \frac{10\theta}{A^2}q + \theta\frac{\theta A - 4}{A^3}, \end{aligned} \tag{3.1}$$

and are the *turning* points belonging to the marginal curve (2.14) at vanishing $\sigma$. Note that with our assumptions[2] we have $0 < \theta < 1$, and $q > 0$. We shall denote by $q_c$ a value on this marginal curve (also denoted the curve of neutral stability). The discriminant of (2.7) is equal to $Z(q)/(1-qA)^2$, and has same sign as $z(q)$.

According to the Harriot–Descartes theorem, in the segment $0 < \theta A < 4$ the equation $z(q) = 0$ possesses one *negative* real solution and either one or three *positive* real ones. In the following to fix ideas, we shall restrict consideration to the numerical values $\bar{A} = 1$, $D_A = 10^{-2}$, and $D = 5 \times 10^{-4}$, for which the minimum of $A(r)$ in $0 \leqslant r \leqslant 1$ is

$$A = 0.013. \tag{3.2}$$

For the given set of data we have three *positive* real solutions. The negative one shall be disregarded as it corresponds to negative values of kinetic constants. Then, at fixed $\theta$ the following results hold:

(i) $Z(q)$ has three real positive roots in the interval[6]

$$A \leqslant A \leqslant A_*, \tag{3.3}$$

and a single real positive root and two complex conjugate ones in the remaining

$$A_* < A \leqslant \bar{A}; \tag{3.4}$$

(ii) Denoting by $q_1, q_2, q_3$ the solutions of $Z(q) = 0$ such that for $A$ belonging to (3.3.)

$$0 < q_1(A) \leqslant q_2(A) < q_3(A), \tag{3.5}$$

then the only positive root in (3.4) in $q_3$;

(iii) $q_2(A)$ and $q_3(A)$ are decreasing functions in their definition domain whereas $q_1(a)$ is decreasing in

$$\underline{A} \leqslant A < A_+, \tag{3.6}$$

and increasing in

TABLE I. Roots of (3.1) and related numerical values corresponding to the various cases of localized dissipative structures depicted in Fig. 1.

| $\theta$ | $A_+$ | $A_*$ | $q_3(\bar{A})$ | $q_1(A_+)$ | $q_1(A_*)=q_2(A_*)$ | $q_3(A_*)$ | $q_1(\underline{A})$ | $q_2(\underline{A})$ | $q_3(\underline{A})$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.983 | – | 0.909 | 0.664 | – | – | 5.454 | 76.823 | 76.824 |
| 0.2 | 0.492 | 0.530 | 0.833 | 1.328 | 1.408 | 1.706 | 7.666 | 76.722 | 76.724 |
| 0.3 | 0.328 | 0.353 | 0.768 | 1.991 | 2.113 | 2.562 | 9.347 | 76.620 | 76.624 |
| 0.4 | 0.246 | 0.265 | 0.712 | 2.655 | 2.815 | 3.412 | 10.754 | 76.517 | 76.525 |
| 0.5 | 0.197 | 0.212 | 0.663 | 3.319 | 3.519 | 4.265 | 11.988 | 76.416 | 76.426 |
| 0.6 | 0.164 | 0.176 | 0.619 | 3.983 | 4.237 | 5.139 | 13.100 | 76.314 | 76.328 |
| 0.7 | 0.140 | 0.151 | 0.579 | 4.647 | 4.939 | 5.989 | 14.118 | 76.213 | 76.229 |
| 0.8 | 0.123 | 0.132 | 0.543 | 5.311 | 5.649 | 6.852 | 15.064 | 76.112 | 76.131 |
| 0.9 | 0.109 | 0.117 | 0.510 | 5.974 | 6.371 | 7.733 | 15.951 | 76.011 | 76.033 |

$$A_+ < A \leqslant A_* , \tag{3.7}$$

which defines $A_+$ ;

(iv) In the vicinity of both $q_1$ and $q_3$ the polynomial $Z(q)$ is an increasing function of $q$, whereas in the neighborhood of $q_2$ it is a decreasing function of $q$;

(v) For any real value of $q$, in the limit $\theta = 0$, we have $Z(q) \geqslant 0$. Then $Z(q)$ has two double real roots, one at $q = 0$ and the other at $q = A^{-1}$. For $\theta$ going to infinity $z(q)$ has no real root.

Thus, using these properties and the numerical estimates given in Table I the following cases appear:

*Case I*: When $q_c$ is such that either

$$q_2(\underline{A}) < q_c < q_3(\underline{A}), \tag{3.8a}$$

or

$$q_3(\bar{A}) < q_c < q_1(A_+), \tag{3.8b}$$

the only *positive* real root of $z(q)$ which can be set equal to $q_c$ is $q_3$. Then we have two values of $r$, $r_1$, and $r_2$ such that

$$q_c = q_3\{A(r_i)\}, \quad i = 1,2,. \tag{3.9}$$

In the region $r_1 < r < r_2$ we have $q_c < q_3(A)$ and $Z(q_c)$ is negative. Thus $w'^2(q = q_c)$ is complex and the exponential in Eq. (2.5a) oscillates in that region of the reactor. Thus, we have bifurcation to a secondary solution. However, for $r$ belonging to the set $(0,r_1) \cup (r_2,1)$ we have $q_c > q_3(A)$ and $z(q)$ is positive. Then as $w'^2(q = q_c)$ is negative and the exponential in Eq. (2.5a) does not oscillate in that region of the reactor, the only solution to (2.4) which satisfies the b.c. is the primary one (1.5). Thus we have a *localized* dissipative structure as depicted in Fig. 1.(a).

*Case II*: When

$$q_1(A_+) < q_c < q_1(A_*), \tag{3.10}$$

the only positive real roots of $z(q)$ are $q_1$ and $q_3$. Then there exist two values of $A$, ${}^1_1A$, and ${}^1_2A$; ${}^1_1A < {}^1_2A$; such that

$$q_1({}^1_iA) = q_c, \quad i = 1,2, \tag{3.11a}$$

and a third value ${}^3A$; ${}^1_1A < {}^1_2A < {}^3A$; such that

$$q_3({}^3A) = q_c, \tag{3.11b}$$

as shown in Table I. $q_1(A)$ is a decreasing function of $A$ in a vicinity of ${}^1_1A$ whereas it is increasing in a neighborhood of ${}^1_2A$. Using the notation

$${}^1_iA = A(r = {}^ir_1) = A(r = {}^ir_2 = 1 - {}^ir_1), \quad i = 1,2,$$

and

$${}^3A = A(r = {}^3r_1) = A({}^3r_2 = 1 - {}^3r_1),$$

which yields

$${}^1r_1 < {}^2r_1 < {}^3r_1 < {}^3r_2 < {}^2r_2 < {}^1r_2,$$

a similar argument to that used in the preceding case produces the solution illustrated in Fig. 1.(c). We have a dissipative structure in three patches.

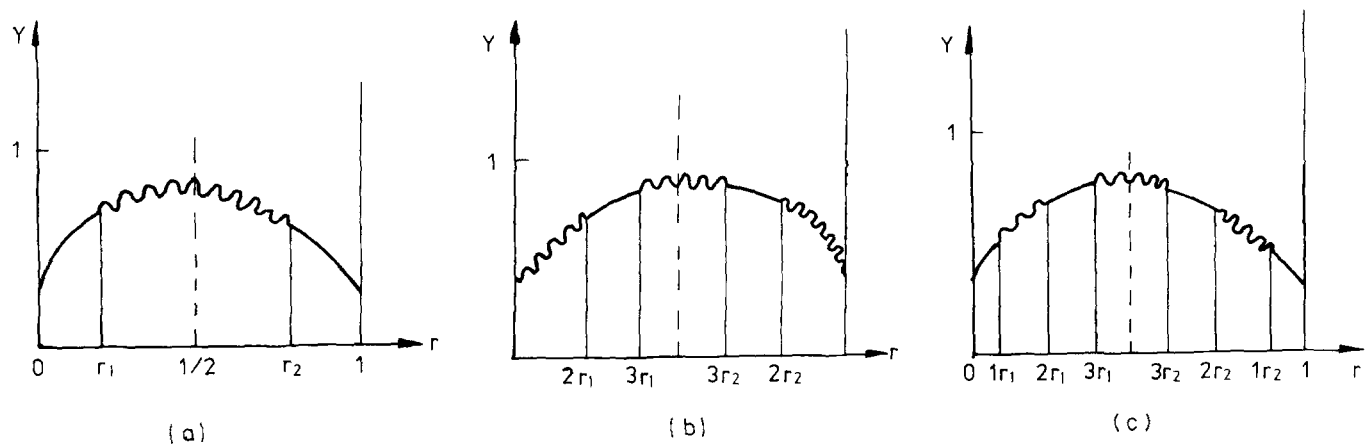



FIG. 1. Various possible dissipative structures in a one-dimensional reactor ($0 \leqslant r \leqslant 1$): (a) A localized single spatial pattern somewhere in the center of the vessel. (b) Three localized structures. (c) Alternative location for three structures. For illustration we have given only the expected distribution of one of the reactants.

*Case III*: When

$$q_1(A_*) < q_c < q_1(\underline{A}), \tag{3.12}$$

there are values of $A$ for which any of the three roots $q_i(A)$, $i = 1,2,3$ can be set equal to $q_c$. Let these values of $A$ be ${}^iA$, $i = 1,2,3$. Each value ${}^iA$ will appear in two points of the reactor ${}^ir_1$ and ${}^ir_2 = 1 - {}^ir_1$ symmetric with respect to the middle point, $r = \frac{1}{2}$. As the three roots $q_i(A)$, $i = 1,2,3$, are decreasing functions in the region (3.12) with $A$ belonging to $(\underline{A}, A_*)$ we have

$${}^1A > {}^2A > {}^3A. \tag{3.13}$$

A discussion similar to that given for case I yields again the distribution depicted in Fig. 1.(c). Note that for the value $\theta = 0.1$ no $A_*$ can be defined and $Z(q)$ has three positive real roots in the whole domain $(\underline{A}, \overline{A})$. However, the distribution of Fig. 1.(c) corresponds to $\theta = 0.1$ provided $q_1(A_+) < q_c < q_1(\underline{A})$.

*Case IV*: When

$$q_1(\underline{A}) < q_c < q_2(\underline{A}), \tag{3.14}$$

only $q_2$ and $q_3$ can be set equal to $q_c$. Thus ${}^1r_1 = 0$, ${}^1r_2 = 1$, in Fig. 1.(c), and the actual distribution of reactants corresponds to the picture given in Fig. 1.(b). Note that the ordering (a), (b), (c) in Fig. 1 corresponds to the $q_c$ decreasing from $q_3(A)$ to $q_1(A_+)$.

*Case V*: When

$$q_c < q_3(\overline{A}), \tag{3.15}$$

for all values of $A$ we have: $Z(q_c) < Z\{q_3(A)\} = 0$. Then the exponential in (2.5a) oscillates in all points in the reactor vessel. We obtain bifurcation to a global dissipative structure which looks like those described in paper I.

Lastly, the case

$$q_c > q_3(\underline{A}), \tag{3.16}$$

has not been considered because for all values of $A$ we have $Z(q_c) > Z\{q_3(A)\} = 0$. Then the exponential in (2.5a) does not oscillate at any point in the reactor vessel. The only solution (2.4) satisfying the b.c. (1.4) is the trivial one. Thus, values of $q$ above $q_3(\underline{A})$ cannot belong to the neutral stability curve (2.14).

## 4. CONCLUSION

In the model problem here discussed we have found either *none*, *one*, or *three* spatially structured patches for a reactive process occurring along a line (one-dimensional) reactor, e.g., along a catalytic wire. The number of such peculiar patches, which indeed affects the efficiency of the reaction-diffusion process, depends on the number of *positive* real roots of the polynomial (3.1). On the other hand, that number is the number of turning points found with the WKB method. It also appears that the *maximum* number of patches is essentially related to the reaction kinetics involved in the process.[7]


## ACKNOWLEDGMENT

The results reported in this note were obtained in the course of research sponsored by the Instituto de Estudios Nucleares (Spain).



[1] R. Aris, *The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts* (Clarendon, Oxford, England, 1976), Vol. I, Sec. 4.7.3; Vol. II, Secs. 7.9 and 8.8.

[2] L.L. Bonilla and M.G. Velarde, J. Math. Phys. **20**, 2692 (1979).

[3] For a general introduction to the subject see G. Nicolis and I. Prigogine, *Self-Organization in Nonequalibrium Systems* (Wiley, New York, 1977).

[4] See, for instance, L.I. Schiff, *Quantum Mechanics* (McGraw–Hill, New York, 1968), Chapter 8. For a related problem to the case discussed in the present paper see J.A. Boa and D.S. Cohen, SIAM J. Appl. Math. **30**, 123 (1976).

[5] When $\mathrm{Im}\sigma \neq 0$ we expect bifurcation to spatio-temporal structures.

[6] The value $A_*$ corresponds to the equals sign between $q_1$ and $q_2$ in the relation (3.5). At $A = A_*$ both roots are real, positive, and equal, whereas for values of $A_*$ above they are complex conjugate.

[7] For the *Brusselator*, discussed in Ref. 3, there is a single turning point. This is a consequence of the expansion to order $D^{1/2}$ used in (3.1). However, when $D_A$ approaches $D$, higher order corrections bring a second turning point.

# Resonance of nonaxial symmetric modes in circular microstrip disk antenna [a)]

W. C. Chew and J. A. Kong
*Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

(Received 27 January 1980; accepted for publication 18 April 1980)

Resonant frequencies of the nonaxial symmetric modes in a microstrip disk are computed using two approaches: Galerkin's method and a perturbative approach. The perturbative approach is good when the substrate of the microstrip disk is thin compared to its radius and when the dielectric constant of the substrate is high. Galerkin's method can be used to compute the resonant frequency to high accuracy but the perturbative approach is more efficient for thin substrate and large dielectric constant. In applying Galerkin's method, the problem is first cast into a vector dual integral equation using vector Hankel transform (VHT). Using VHT, it is also shown that the magnetic-wall model is only good when the substrate is of zero thickness. Using zero-order current distribution on the disk, we also derive the radiation field and radiation pattern. Also, by taking into account the radiation loss, the resonant frequencies are complex. We find discrepancies when we compare our results for the resonant frequency shifts with that obtained by quasistatic approach.

## I. INTRODUCTION

The applications of microstrip disks as antenna elements[1-4] and resonators[5-9] have aroused interest in finding a way to predict the resonant frequencies of a microstrip disk. Watkins[1] obtained the approximate resonant frequencies for a circular microstrip disk by using a magnetic-wall resonator model ignoring the fringing field effect. Several authors have obtained corrections to Watkins' formula using a quasistatic argument.[1,6,8] Improvements to the magnetic-wall model have also been obtained using qualitative arguments.[3,7] However, at resonance, the disk size is of the order of the wavelength concerned and the validity of such arguments is doubtful. Itoh[9] performed a full-wave analysis to find the resonant frequencies of a rectangular microstrip disk. The disk was enclosed in a waveguide so that the resonant frequencies are real.

Recently, Chew and Kong[10] studied the resonant frequencies of the axial-symmetric mode in a circular microstrip disk resonator using dual integral equation formalism and a perturbative approach. The rigorous approaches employed in Refs. 9 and 10 show discrepancies with the quasistatic approach. The dual integral equation formalism to mixed boundary value problems has been of historic interest.[11-13] The basis function expansion method to the solution of mixed boundary value problem was first employed by Tranter[12] and subsequently used by Itoh and Mittra,[14] Borkar and Yang,[8] and Chew and Kong[10,15] to the microstrip problem.

The problem of the nonaxial symmetric mode resonance in a circular microstrip disk has been formulated,[10] but has not been solved. It is shown that solution to such a problem becomes tractable if we devise a new kind of Hankel transform called the vector Hankel transform (see Appendix). By doing so, the nonaxial symmetric mode problem can be cast into a vector dual integral equation which can be solved by vector basis function expansion method. We also derive a perturbation formula for the resonant frequencies of the disk resonator and expressions for its radiation fields.

## II. INTEGRAL EQUATIONS FOR THE NONAXIAL SYMMETRIC MODES

The problem of finding the natural modes of a circular microstrip disk as shown in Fig. 1 has been formulated in terms of a set of integral equations.[10] In the formulation, a general form in terms of Hankel transform is written for the $E_z$ and $H_z$ components of the TM and TE waves, respectively. The tangential components of TM and TE fields can thus

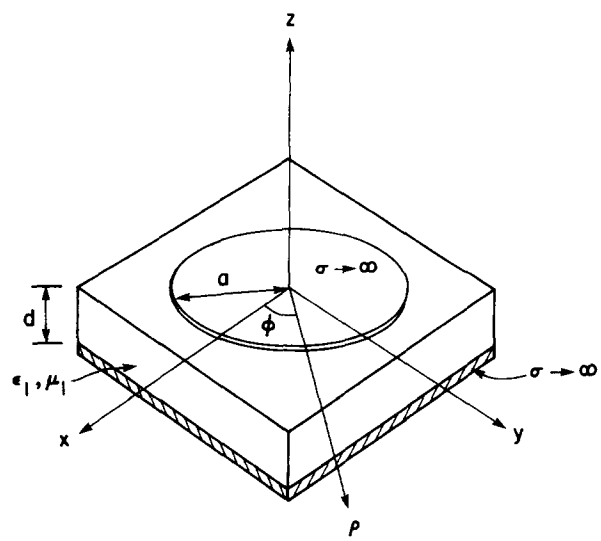



FIG. 1. Geometrical configuration.

[a)]This work is supported by Schlumberger–Doll Research Center and the Joint Services Electronics Program under Contract DAAG-29-78-C-0020.

 0022-2488/80/102590-09$1.00  

be obtained.[16] The mixed boundary condition imposed on the tangential field components at the plane that contains the disk results in

$$\epsilon_\rho(\rho)=\frac{E_\rho(\rho)}{\cos n\phi}=\int_0^\infty e_n^c(k_\rho)ik_z(1-R^{\mathrm{TM}})\times J_n'(k_\rho\rho)\,dk_\rho+i\omega\mu\int_0^\infty h_n^s(k_\rho)(1+R^{\mathrm{TE}})\times\frac{n}{k_\rho\rho}J_n(k_\rho\rho)\,dk_\rho=0,\quad \rho<a, \tag{1}$$

$$\epsilon_\phi(\rho)=-\frac{E_\phi(\rho)}{\sin n\phi}=\int_0^\infty e_n^c(k_\rho)ik_z(1-R^{\mathrm{TM}})\times\frac{n}{k_\rho\rho}J_n(k_\rho\rho)\,dk_\rho+i\omega\mu\int_0^\infty h_n^s(k_\rho)(1+R^{\mathrm{TE}})\times J_n'(k_\rho\rho)\,dk_\rho=0,\quad \rho<a, \tag{2}$$

$$\kappa_\phi(\rho)=\frac{K_\phi(\rho)}{\sin n\phi}=2\int_0^\infty h_n^s(k_\rho)ik_z\times J_n'(k_\rho\rho)\,dk_\rho+2i\omega\epsilon\int_0^\infty e_n^c(k_\rho)\times\frac{n}{k_\rho\rho}J_n(k_\rho\rho)\,dk_\rho=0,\quad \rho>a, \tag{3}$$

$$\kappa_\rho(\rho)=-\frac{K_\rho(\rho)}{\cos n\phi}=2\int_0^\infty h_n^s(k_\rho)\frac{nik_z}{k_\rho\rho}\times J_n(k_\rho\rho)\,dk_\rho+2i\omega\epsilon\int_0^\infty e_n^c(k_\rho)\times J_n'(k_\rho\rho)dk_\rho=0,\quad \rho\geqslant a, \tag{4}$$

where $e_n^c(k_\rho)$ and $h_n^s(k_\rho)$ are unknowns to be determined.

$$R^{\mathrm{TM}}=\frac{i\epsilon_1k_z\cos k_{1z}d-\epsilon k_{1z}\sin k_{1z}d}{i\epsilon_1k_z\cos k_{1z}d+\epsilon k_{1z}\sin k_{1z}d}, \tag{5a}$$

$$R^{\mathrm{TE}}=\frac{i\mu_1k_z\sin k_{1z}d+\mu k_{1z}\cos k_{1z}d}{i\mu_1k_z\sin k_{1z}d-\mu k_{1z}\cos k_{1z}d}, \tag{5b}$$

$$k_z=\sqrt{\omega^2\mu\epsilon-k_\rho^2},\quad k_{1z}=\sqrt{\omega^2\mu_1\epsilon_1-k_\rho^2}, \tag{5c}$$

and $K_\phi(\rho)$ and $K_\rho(\rho)$ are surface current components on the disk.

In general, hybrid modes exist unless $n=0$, i.e., if the mode is axial symmetric. Using vector Hankel transform (VHT) (see Appendix), Eqs. (1)–(4) can be written more concisely as

$$\bar{\epsilon}(\rho)=\begin{bmatrix}\epsilon_\rho(\rho)\\ \epsilon_\phi(\rho)\end{bmatrix}=\int_0^\infty k_\rho\,dk_\rho\,\bar{\bar{H}}_n(k_\rho\rho)\cdot\begin{bmatrix}e_n^c(k_\rho)\dfrac{ik_z}{k_\rho}(1-R^{\mathrm{TM}})\\ i\omega\mu\dfrac{h_n^s(k_\rho)}{k_\rho}(1+R^{\mathrm{TE}})\end{bmatrix}=0,\quad \rho<a, \tag{6}$$

$$\bar{\kappa}(\rho)=\begin{bmatrix}\kappa_\rho(\rho)\\ \kappa_\phi(\rho)\end{bmatrix}=\int_0^\infty k_\rho\,dk_\rho\,\bar{\bar{H}}_n(k_\rho\rho)\cdot\begin{bmatrix}2i\omega\epsilon\dfrac{e_n^c(k_\rho)}{k_\rho}\\ 2h_n^s(k_\rho)\dfrac{ik_z}{k_\rho}\end{bmatrix}=0,\quad \rho>a, \tag{7a}$$

and

$$\kappa_\rho(\rho\to a^-)=0. \tag{7b}$$

From the reciprocal relationship of VHT [see Eqs. (A1) and (A2)], the VHT of $\bar{\kappa}(\rho)$ is

$$\bar{K}(k_\rho)=\begin{bmatrix}K_1(k_\rho)\\ K_2(k_\rho)\end{bmatrix}=\begin{bmatrix}2i\omega\epsilon\dfrac{e_n^c(k_\rho)}{k_\rho}\\ 2h_n^s(k_\rho)\dfrac{ik_z}{k_\rho}\end{bmatrix}. \tag{8}$$

As such, Eqs. (6) and (7) can be written as a set of vector dual integral equations

$$\bar{\epsilon}(\rho)=\int_0^\infty k_\rho\,dk_\rho\,\bar{\bar{H}}_n(k_\rho\rho)\cdot\bar{\bar{G}}(k_\rho)\cdot\bar{K}(k_\rho)=0,\quad \rho<a, \tag{9}$$

$$\bar{\kappa}(\rho)=\int_0^\infty k_\rho\,dk_\rho\,\bar{\bar{H}}_n(k_\rho\rho)\cdot\bar{K}(k_\rho)=0,\quad \rho>a, \tag{10}$$

where

$$\bar{\bar{G}}(k_\rho)=\begin{bmatrix}\dfrac{k_z}{2\omega\epsilon}(1-R^{\mathrm{TM}}) & 0\\ 0 & \dfrac{\omega\mu}{2k_z}(1+R^{\mathrm{TE}})\end{bmatrix} \tag{11}$$

is related to the dyadic Green's function in $k_\rho$-*space*.

## III. ZEROTH-ORDER THEORY

In the limit when $d\to 0$, we can show from (11) that

$$\bar{\bar{G}}(k_\rho)\sim-\frac{id}{\omega\epsilon_1}\begin{bmatrix}k_1^2-k_\rho^2 & 0\\ 0 & k_1^2\end{bmatrix}. \tag{12}$$

With this approximation for $\bar{\bar{G}}(k_\rho)$, we can show that the current distribution predicted by the magnetic-wall model[5] is a solution of (9) and (10). The current distribution derived from the magnetic-wall model is given by

$$\bar{\kappa}(\rho)=\begin{cases}A\begin{bmatrix}J_n'\left(\beta_{nm}\dfrac{\rho}{a}\right)\\ \dfrac{na}{\beta_{nm}\rho}J_n\left(\beta_{nm}\dfrac{\rho}{a}\right)\end{bmatrix} & \rho<a,\\ 0 & \rho>a,\end{cases} \tag{13}$$

where $J_n'(\beta_{nm}) = 0$ satisfy the edge condition (7b). The VHT of $\bar{\kappa}(\rho)$ can be found easily

$$\bar{K}(k_\rho) = A\int_0^a \rho\, d\rho\, \bar{\bar{H}}_n(k_\rho\rho)\cdot\begin{bmatrix} J_n'\left(\beta_{nm}\frac{\rho}{a}\right) \\ \frac{na}{\beta_{nm}\rho}J_n\left(\beta_{nm}\frac{\rho}{a}\right)\end{bmatrix}$$

$$= A\beta_{nm}J_n(\beta_{nm})\begin{bmatrix} \frac{1}{(\beta_{nm}/a)^2 - k_\rho^2}J_n'(k_\rho a) \\ \frac{na}{\beta_{nm}^2 k_\rho}J_n(k_\rho a)\end{bmatrix}. \quad (14)$$

Substituting $\bar{K}(k_\rho)$ into (9), with approximation (12), and $k_1 = \beta_{nm}/a$, we arrive at

$$\bar{\epsilon}(\rho) = -\frac{idA}{\omega_{nm}\epsilon_1}\beta_{nm}J_n(\beta_{nm})\int_0^\infty k_\rho\, dk_\rho \cdot\begin{bmatrix} J_n'(k_\rho a)J_n'(k_\rho\rho) + \frac{n^2}{k_\rho^2\rho a}J_n(k_\rho a)J_n(k_\rho\rho) \\ \frac{nJ_n'(k_\rho a)J_n(k_\rho\rho)}{k_\rho\rho} + \frac{nJ_n(k_\rho a)J_n'(k_\rho\rho)}{k_\rho a}\end{bmatrix}, \quad (15)$$

where $\omega_{nm} = (\beta_{nm}/a)/\sqrt{\mu_1\epsilon_1}$. Using identities (A5) and (A6), we arrive at

$$\bar{\epsilon}(\rho) = -\frac{idA}{\omega_{nm}\epsilon_1}(\beta_{nm}/a)J_n(\beta_{nm})\begin{bmatrix}\delta(\rho-a) \\ 0\end{bmatrix}. \quad (16)$$

Hence, Eqs. (9) and (10) are satisfied by (13). We can thus conclude that the solution of (9) and (10) is in fact the solution of the magnetic-wall model when $d\to 0$.

## IV. GALERKIN'S METHOD

Noting that when $d\to 0$, the current distribution on the disk approaches that of a magnetic-wall model, we can find a complete set of vector basis functions to approximate the current distribution. By noting that the currents due to the TE and TM modes of a magnetic-wall cavity form a complete set, the current distribution of the $n$th mode of the microstrip disk resonator can be represented as

$$\bar{\kappa}(\rho) = \begin{cases} \sum_{m=1}^{\infty} a_m \begin{bmatrix} J_n'\left(\beta_{nm}\frac{\rho}{a}\right) \\ \frac{na}{\beta_{nm}\rho}J_n\left(\beta_{nm}\frac{\rho}{a}\right)\end{bmatrix} + \sum_{p=1}^{\infty} b_p \begin{bmatrix} \frac{na}{\alpha_{np}\rho}J_n\left(\alpha_{np}\frac{\rho}{a}\right) \\ J_n'\left(\alpha_{np}\frac{\rho}{a}\right)\end{bmatrix}, & \rho < a, \\ 0, & \rho > a, \end{cases} \quad (17)$$

where $J_n'(\beta_{nm}) = 0$ and $J_n(\alpha_{np}) = 0$. The orthogonality of the above mode vectors can be proved easily. In practice, we need only to pick the $M$ terms of the first series and the $P$ terms of the second series resulting in

$$\bar{\kappa}(\rho) = \sum_{m=1}^{M} a_m \bar{k}_{nm}(\rho) + \sum_{p=1}^{P} b_p \bar{f}_{np}(\rho), \quad \rho < a, \quad (18)$$
$$= 0, \quad \rho > a,$$

where

$$\bar{k}_{nm}(\rho) = \begin{bmatrix} J_n'\left(\beta_{nm}\frac{\rho}{a}\right) \\ \frac{na}{\beta_{nm}\rho}J_n\left(\beta_{nm}\frac{\rho}{a}\right)\end{bmatrix}, \quad (19a)$$

and

$$\bar{f}_{np}(\rho) = \begin{bmatrix} \frac{na}{\alpha_{np}\rho}J_n\left(\alpha_{np}\frac{\rho}{a}\right) \\ J_n'\left(\alpha_{np}\frac{\rho}{a}\right)\end{bmatrix}. \quad (19b)$$

The VHT of $\bar{\kappa}(\rho)$ can be found easily, giving

$$\bar{K}(k_\rho) = \sum_{m=1}^{M} a_m \bar{K}_{nm}(k_\rho) + \sum_{p=1}^{P} b_p \bar{F}_{np}(k_\rho), \quad (20)$$

where

$$\bar{K}_{nm}(k_\rho) = \beta_{nm}J_n(\beta_{nm})\begin{bmatrix} \frac{J_n'(k_\rho a)}{(\beta_{nm}/a)^2 - k_\rho^2} \\ \frac{na}{\beta_{nm}^2 k_\rho}J_n(k_\rho a)\end{bmatrix}, \quad (21a)$$

$$\bar{F}_{np}(k_\rho) = \frac{k_\rho a J_n'(\alpha_{np})}{k_\rho^2 - (\alpha_{np}/a)^2}\begin{bmatrix} 0 \\ J_n(k_\rho a)\end{bmatrix}. \quad (21b)$$

Equations (10) and (7a) are automatically satisfied by the choice of $\bar{\kappa}(\rho)$ in (18). The choice is also asymptotically good when $d\to 0$ because only one term is needed in the limit when $d\to 0$. Substituting (20) into (9), we obtain

$$\sum_{m=1}^{M} a_m \int_0^\infty k_\rho\, dk_\rho\, \bar{\bar{H}}_n(k_\rho\rho)\cdot\bar{\bar{G}}(k_\rho)\cdot\bar{K}_{nm}(k_\rho) + \sum_{p=1}^{P} b_p \int_0^\infty k_\rho\, dk_\rho\, \bar{\bar{H}}_n(k_\rho\rho)\cdot\bar{\bar{G}}(k_\rho)\cdot\bar{F}_{np}(k_\rho) = 0, \quad \rho < a. \quad (22)$$

To find $a_m$ and $b_p$, we find the inner product of the above with $\bar{k}_{nj}(\rho)$ and $\bar{f}_{nk}(\rho)$, i.e., multiplying the above with $\rho\bar{k}_{nj}^T(\rho)$ and $\rho\bar{f}_{nk}^T(\rho)$ and integrating from 0 to $a$ for $J = 1, 2, \ldots, M$, $k = 1, 2, \ldots, P$. Using Parseval's theorem (see Appendix), (22) becomes

$$\sum_{m=1}^{M} a_m \int_0^\infty k_\rho\, dk_\rho\, \bar{K}_{nj}^T(k_\rho)\cdot\bar{\bar{G}}(k_\rho)\cdot\bar{K}_{nm}(k_\rho)$$

$$+\sum_{p=1}^{P} b_p \int_0^\infty k_\rho\, dk_\rho\, \overline{\overline{K}}_{nj}^{T}(k_\rho)$$

$$\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{F}_{np}(k_\rho)=0,\quad J=1,\ 2,\ \ldots,\ M, \tag{23a}$$

$$\sum_{m=1}^{M} a_m \int_0^\infty k_\rho\, dk_\rho\, \overline{F}_{nk}^{T}(k_\rho)\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{\overline{K}}_{nm}(k_\rho)$$

$$+\sum_{p=1}^{P} b_p \int_0^\infty k_\rho\, dk_\rho\, \overline{F}_{nk}^{T}(k_\rho)$$

$$\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{F}_{np}(k_\rho)=0,\quad k=1,\ 2,\ \ldots,\ P. \tag{23b}$$

The above is a system $M+P$ linear algebraic equations involving $M+P$ unknowns from which we can solve for $a_m$ and $b_p$. Written in matrix form, it is

$$\begin{pmatrix} A_{11}^{KK} & \cdots & A_{1M}^{KK} & A_{11}^{FK} & \cdots & A_{1p}^{FK} \\ \vdots & & \vdots & \vdots & & \vdots \\ A_{M1}^{KK} & \cdots & A_{MM}^{KK} & A_{M1}^{KK} & \cdots & A_{MP}^{FK} \\ A_{11}^{KF} & \cdots & A_{1M}^{KF} & A_{11}^{FF} & \cdots & A_{1p}^{FF} \\ \vdots & & \vdots & \vdots & & \vdots \\ A_{p1}^{KF} & \cdots & A_{PM}^{KF} & A_{p1}^{FF} & \cdots & A_{pp}^{FF} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_M \\ b_1 \\ \vdots \\ b_P \end{pmatrix} = 0, \tag{24}$$

where

$$A_{ij}^{KK}=\int_0^\infty k_\rho\, dk_\rho\, \overline{\overline{K}}_{ni}^{T}(k_\rho)\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{\overline{K}}_{nj}(k_\rho), \tag{24a}$$

$$A_{ji}^{FK}=A_{ij}^{KF}=\int_0^\infty k_\rho\, dk_\rho\, \overline{F}_{ni}^{T}(k_\rho)\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{\overline{K}}_{nj}(k_\rho), \tag{24b}$$

$$A_{ij}^{FF}=\int_0^\infty k_\rho\, dk_\rho\, \overline{F}_{ni}^{T}(k_\rho)\cdot\overline{\overline{G}}(k_\rho)\cdot\overline{F}_{nj}(k_\rho). \tag{24c}$$

For nontrivial solutions of $a_m$ and $b_p$, we require that

$$\det|\overline{\overline{A}}|=0, \tag{25}$$

where $\overline{\overline{A}}$ is the matrix in (24). Equation (25) is the eigenequation for the resonant frequencies of the disk resonator. The accuracy of the resonant frequencies found can be improved by increasing $M$ and $P$ arbitrarily.

## V. PERTURBATIVE APPROACH

We have shown that in the limit when $d\to 0$, the resonant frequencies of the disk resonator approach that of a magnetic-wall cavity. Therefore, for small $d$, we can think of the disk resonator as a perturbation of the magnetic-wall cavity. It was derived in Ref. 10 using a perturbative approach that the resonant frequency shift of the magnetic-wall cavity with the magnetic wall removed is given by

$$\Delta\omega\simeq\frac{-i\iint_{\Delta S}(\overline{E}_i^{*}\times\overline{H}_f)\cdot\hat{n}\,ds}{4\langle W_T\rangle}. \tag{26}$$

In the above, $\overline{E}_i$ is the electric field in the cavity before perturbation, $\overline{H}_f$ is the magnetic field in the disk resonator after the magnetic wall has been removed, $\langle W_T\rangle$ is the time-average total energy stored in the cavity before perturbation, and $\Delta S$ is the surface of the magnetic wall of the cavity.

From Ref. 10, taking only the cosine mode for the TM wave, $E_z$ in the upper half-space above the dielectric substrate for the disk resonator is given by

$$E_z=\cos n\phi\int_0^\infty k_\rho\, dk_\rho\, e_n^c(k_\rho)$$

$$\times(1-R^{\mathrm{TM}})e^{ik_z z}J_n(k_\rho\rho). \tag{27}$$

By matching boundary conditions, $E_{1z}$ in the dielectric substrate is given by

$$E_{1z}=-\cos n\phi\,\frac{\epsilon}{\epsilon_1}\int_0^\infty k_\rho\, dk_\rho\, e_n^c(k_\rho)$$

$$\times\frac{\cos k_{1z}(z+d)}{\cos k_{1z}d}(1+R^{\mathrm{TM}})J_n(k_\rho\rho). \tag{28}$$

Similarly, with $H_z$ in the upper half-space given by

$$H_z=\sin n\phi\int_0^\infty k_\rho\, dk_\rho\, h_n^s(k_\rho)$$

$$\times(1+R^{\mathrm{TE}})e^{ik_z z}J_n(k_\rho\rho), \tag{29}$$

we derive that in the dielectric substrate,

$$H_{1z}=\sin n\phi\,\frac{\mu}{\mu_1}\int_0^\infty k_\rho\, dk_\rho\, h_n^s(k_\rho)$$

$$\times\frac{\sin k_{1z}(z+d)}{\sin k_{1z}d}(1+R^{\mathrm{TE}})J_n(k_\rho\rho). \tag{30}$$

We know $\overline{E}_i$ in (26) from the field in the magnetic-wall cavity,

$$\overline{E}_i=-\hat{z}\,\frac{A\beta_{nm}}{i\omega\epsilon_1 a}\cos n\phi\,J_n\left(\beta_{nm}\frac{\rho}{a}\right). \tag{31}$$

Noting that $\overline{E}_i$ only has a $\hat{z}$-component, we can rewrite (26) as

$$\Delta\omega\simeq\frac{i\int_{-d}^{0}\int_0^{2\pi}E_{iz}^{*}H_{f\phi}|_{\rho=a}\,a\,d\phi\,dz}{4\langle W_T\rangle}, \tag{26a}$$

where $H_{f\phi}$ can be derived from (28) and (30)[16] giving

$$H_{f\phi}=\cos n\phi\left(\frac{\mu n}{\mu_1\rho}\int_0^\infty\frac{h_n^s(k_\rho)k_{1z}\cos k_{1z}(z+d)}{k_\rho\sin k_{1z}d}\right.$$

$$\left.\times(1+R^{\mathrm{TE}})J_n(k_\rho\rho)\,dk_\rho-i\omega\epsilon\int_0^\infty e_n^c(k_\rho)\right.$$

$$\times\frac{\cos k_{1z}(z+d)}{\cos k_{1z}d}(1+R^{\mathrm{TM}})J_n'(k_\rho\rho)\,dk_\rho\Bigg). \tag{32}$$

When $d\to 0$, $e_n^c(k_\rho)$ and $h_n^s(k_\rho)$ can be approximated as

$$e_n^c(k_\rho)\simeq\frac{Ak_\rho}{2i\omega\epsilon}\frac{\beta_{nm}}{(\beta_{nm}/a)^2-k_\rho^2}\times J_n(\beta_{nm})J_n'(k_\rho a), \tag{33a}$$

$$h_n^s(k_\rho)\simeq\frac{-iAna}{2k_z\beta_{nm}}J_n(\beta_{nm})J_n(k_\rho a). \tag{33b}$$

Consequently,

$$\int_{-d}^{0}\int_0^{2\pi}E_{iz}^{*}H_{f\phi}\Big|_{\rho=a}a\,d\phi\,dz \simeq-\frac{\pi|A|^2\beta_{nm}^2}{2i\omega\epsilon_1}J_n^2(\beta_{nm})\times\Bigg(\frac{i\mu n^2}{\mu_1\beta_{nm}^2}\int_0^\infty\frac{J_n^2(k_\rho a)}{k_\rho k_z}(1+R^{\mathrm{TE}})\,dk_\rho + \int_0^\infty k_\rho\frac{[J_n'(k_\rho a)]^2}{[(\beta_{nm}/a)^2-k_\rho^2]}\times\frac{\tan k_{1z}d}{k_{1z}}(1+R^{\mathrm{TM}})\,dk_\rho\Bigg),\quad n>0. \tag{34}$$

Using the fact that

$$\langle W_T\rangle=\frac{|A|^2}{4\omega\epsilon_1}\pi d(\beta_{nm}^2-n^2)J_n^2(\beta_{nm}),\quad n>0, \tag{35}$$

we obtain

$$\frac{\Delta\omega}{\omega}=-\frac{\beta_{nm}^2}{2d(\beta_{nm}^2-n^2)}\Bigg(\frac{i\mu n^2}{\mu_1\beta_{nm}^2}\times\int_0^\infty\frac{J_n^2(k_\rho a)}{k_z k_\rho}(1+R^{\mathrm{TE}})\,dk_\rho + \int_0^\infty k_\rho\frac{[J_n'(k_\rho a)]^2}{[(\beta_{nm}/a)^2-k_\rho^2]}\frac{\tan k_{1z}d}{k_{1z}}(1+R^{\mathrm{TM}})\,dk_\rho\Bigg). \tag{36}$$

In the above, we take $k_z=\sqrt{\omega_{nm}^2\mu\epsilon-k_\rho^2}$, $k_{1z}=\sqrt{(\omega_{nm}^2)\mu_1\epsilon_1-k_\rho^2}$, where $\omega_{nm}=\beta_{nm}/a\sqrt{\mu_1\epsilon_1}$. Furthermore, if we are only interested in the imaginary part of the frequency shift, we only need to integrate from 0 to $k$

$$\mathrm{Im}(\Delta\omega/\omega)=-\frac{\beta_{nm}^2}{2d(\beta_{nm}^2-n^2)}\mathrm{Im}\Bigg(\frac{i\mu n^2}{\mu_1\beta_{nm}^2}\times\int_0^k\frac{J_n(k_\rho a)}{k_z k_\rho}(1+R^{\mathrm{TE}})\,dk_\rho + \int_0^k k_\rho+\frac{[J_n'(k_\rho a)]^2}{[(\beta_{nm}/a)^2-k_\rho^2]}\frac{\tan k_{1z}d}{k_{1z}}(1+R^{\mathrm{TM}})\,dk_\rho\Bigg). \tag{37}$$

This is because the imaginary frequency shift is due to radiation loss and the fields due to $k_\rho$ from $k$ to infinity do not contribute to the radiation fields and the quantity inside the large parentheses becomes pure real.

## VI. RADIATION FIELD

We can derive the radiation field in the upper half-space of the disk resonator from (27) and (29). First, we write (27) as a Hankel integral

$$E_z=\frac{\cos n\phi}{2}\int_{-\infty}^{\infty}k_\rho e_n^c(k_\rho)\times(1-R^{\mathrm{TM}})e^{ik_z z}H_n^{(1)}(k_\rho\rho)\,dk_\rho. \tag{38}$$

When $\rho$ and $z$ are large, the integrand in (38) is rapidly oscillating. Using the large argument approximation to $H_n^{(1)}(k_\rho\rho)$, we find a stationary point at $k_\rho=k\sin\theta$ where $\theta=\tan^{-1}(\rho/z)$ (Ref. 16, p. 218). Thus, the leading order approximation to (38) can be obtained by approximating the slowly varying part of the integrand with its value at the stationary point, and integrating the rapidly varying part exactly. With the approximation to $e_n^c(k_\rho)$ given by (33a), we have

$$E_z^{(0)}\sim\frac{A\cos n\phi k^2\sin\theta\cos\theta\beta_{nm}}{4i\omega\epsilon[(\beta_{nm}/a)^2-k^2\sin^2\theta]}\times J_n(\beta_{nm})J_n'(ka\sin\theta)[1-R^{\mathrm{TM}}(\theta)]\times\frac{H_n^{(1)}(k\rho\sin\theta)}{H_0^{(1)}(k\rho\sin\theta)}\int_{-\infty}^{\infty}\frac{k_\rho}{k_z}e^{ik_z z}\times H_0^{(1)}(k_\rho\rho)\,dk_\rho,\quad \rho,z\to\infty. \tag{39}$$

Noting that the above integral can be evaluated exactly, we have

$$E_z^{(0)}\sim-\frac{A\cos n\phi k^2\sin\theta\cos\theta\beta_{nm}}{2\omega\epsilon[(\beta_{nm}/a)^2-k^2\sin\theta]}\times J_n(\beta_{nm})J_n'(ka\sin\theta)[1-R^{\mathrm{TM}}(\theta)]\times(-i)^n\frac{e^{ikr}}{r},\quad r\to\infty, \tag{40}$$

where $r=\sqrt{z^2+\rho^2}$. In the above, we have replaced $H_n^{(1)}(x)/H_0^{(1)}(x)\sim(-i)^n$ when $x\to\infty$. Since $E_z^{(0)}=-E_\theta^{(0)}\sin\theta$, we obtain $E_\theta^{(0)}$ as

$$E_\theta^{(0)}\sim\frac{A\cos n\phi k^2\cos^2\theta\beta_{nm}}{2\omega\epsilon[(\beta_{nm}/a)^2-k^2\sin^2\theta]}J_n(\beta_{nm})\times J_n'(ka\sin\theta)[1-R^{\mathrm{TM}}(\theta)](-i)^n\frac{e^{ikr}}{r}. \tag{41}$$

Similarly, from (29), we can approximate $H_z^{(0)}$ in the far-field as

$$H_z^{(0)}\sim-\frac{An a\sin n\phi}{2\beta_{nm}}J_n(\beta_{nm})J_n\times(ka\sin\theta)[1+R^{\mathrm{TE}}(\theta)](-i)^n e^{ikr}/r. \tag{42}$$

Since $E_\phi^{(0)}=-\sqrt{\mu/\epsilon}\,H_\theta^{(0)}$, we have

$$E_\phi^{(0)}\sim\sqrt{\mu/\epsilon}\,\frac{An a\sin n\phi}{2\beta_{nm}\sin\theta}J_n(\beta_{nm})J_n(ka\sin\theta)\times[1+R^{\mathrm{TE}}(\theta)](-i)^n\frac{e^{ikr}}{r}. \tag{43}$$

We note that the radiation field given in (41) and (43) is exactly that derived in Ref. 17 under a different context.

## VII. A NOTE ON NUMERICAL COMPUTATION

In evaluating (24a) for (25), we encounter an integral which has a pole near the real axis. The proximity of this pole to the path of integration makes the evaluation of the integral difficult. Letting $G^{TM}(k_\rho) = k_z(1 - R^{TM})$, and noting that $G^{TM}(-k_\rho) = G^{TM}(k_\rho)$, we can rewrite the particular integral as

$$I_{ij} = \frac{\beta_{ni}\beta_{nj}}{2\omega\epsilon} J_n(\beta_{ni}) J_n(\beta_{nj}) \Bigg( \int_0^\infty dk_\rho\, k_\rho \left[ G^{TM}(k_\rho) - \frac{H(k_g)}{k_\rho^2 - k_g^2} \right] \times [J_n'(k_\rho a)]^2 \left\{ \left[ k_\rho^2 - \left(\frac{\beta_{ni}}{a}\right)^2 \right] \left[ k_\rho^2 - \left(\frac{\beta_{nj}}{a}\right)^2 \right] \right\}^{-1} + H(k_g) \int_0^\infty dk_\rho\, k_\rho \times [J_n'(k_\rho a)]^2 \left\{ [k_\rho^2 - k_g^2] \left[ k_\rho^2 - \left(\frac{\beta_{ni}}{a}\right)^2 \right] \left[ k_\rho^2 - \left(\frac{\beta_{nj}}{a}\right)^2 \right] \right\}^{-1} \Bigg), \tag{44}$$

where

$$H(k_g) = \lim_{k_\rho \to k_g} (k_\rho^2 - k_g^2) G^{TM}(k_\rho), \tag{44a}$$

and $\pm k_g$ are the locations of the poles. By so doing, the first singularity is removed from the integrand and the integral can be integrated efficiently using Gaussian quadrature. The second integral can be integrated exactly using contour integration[18], giving

$$\int_0^\infty dk_\rho\, k_\rho [J_n'(k_\rho a)]^2 \left\{ [k_\rho^2 - k_g^2] \left[ k_\rho^2 - \left(\frac{\beta_{ni}}{a}\right)^2 \right] \left[ k_\rho^2 - \left(\frac{\beta_{nj}}{a}\right)^2 \right] \right\}^{-1} = \pi i J_n'(k_g a) H_n^{(1)\prime}(k_g a) \left\{ 2 \left[ k_g^2 - \left(\frac{\beta_{ni}}{a}\right)^2 \right] \left[ k_g^2 - \left(\frac{\beta_{nj}}{a}\right)^2 \right] \right\}^{-1} - \delta_{ij} \times a^2 \left[ \left(\frac{n}{\beta_{ni}}\right)^2 - 1 \right] \left\{ 2\beta_{ni}^2 \left[ \left(\frac{\beta_{ni}}{a}\right)^2 - k_g^2 \right] \right\}^{-1} + n \left\{ 2 \left[ k_g \frac{\beta_{ni}\beta_{nj}}{a} \right]^2 \right\}^{-1} \tag{45}$$

Also, in evaluating the integrals in (24), there are branch points at $k_\rho = \omega\sqrt{\mu\epsilon}$ and $k_\rho = \omega\sqrt{\mu_1\epsilon_1}$. Since $\omega$ can be complex, these branch points can be below the real axis. Thus the path of integration in (24) has to be deformed sufficiently below the real axis to avoid the branch points. To search for the zeros of (25), Muller's method was used.

The subtraction of singularity method can also be applied to (36) for the efficient evaluation of the integral.

## VIII. RESULTS AND DISCUSSIONS

Since the $n = 1$ mode of the circular microstrip disk has been used widely in microstrip antenna applications, we shall study its resonant frequency closely here.

In Fig. 2, we show the resonant wave number shift vs $d/a$ using Galerkin's method discussed in Sec. IV. The ordered pair of numbers assigned to each curve represent the numbers for $M$ and $P$, respectively, used in (18). We note that the imaginary part of the resonant frequency shift converges rapidly as we increase $(M, P)$ from (1, 0) to (2, 2). The real part of the resonant frequency shift converges less slowly. This is because the real resonant frequency shift is due to the reactive power leakage to the near field. Since the near field is singular, good approximation to it can only be obtained with many basis functions. We also note that the curves due to various choices of $(M, P)$ do not deviate from each other drastically due to the stationary characteristic of Galerkin's method. We note that Galerkin's method converges for $(M, P) = (2, 2)$.

In Fig. 3, we compare Galerkin's method with the perturbative approach, and the quasistatic approach. In the quasistatic approach, the disk resonator is thought of as an $LC$ circuit resonator. The shift in resonant frequency is due to the fringing field in $C$.[1,6,8] Thus, the change in resonant frequency is given by

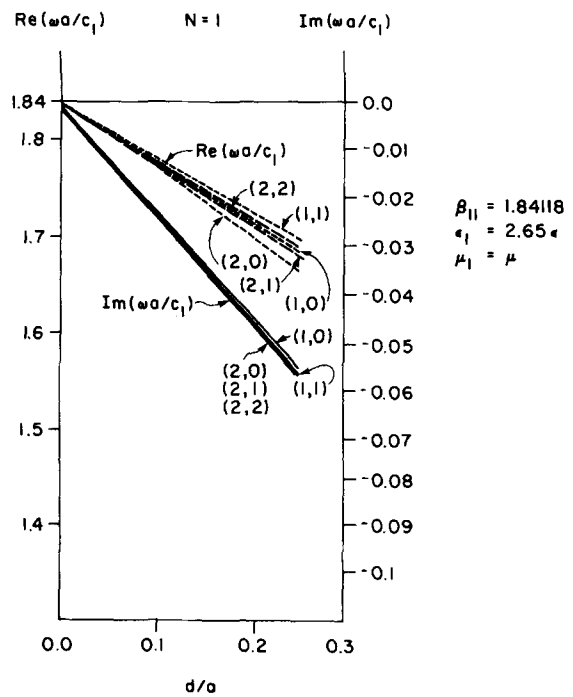



FIG. 2. Plot showing computed resonant frequency shifts of the lowest $N = 1$ mode for different values of $(M,P)$ in Galerkin's method.

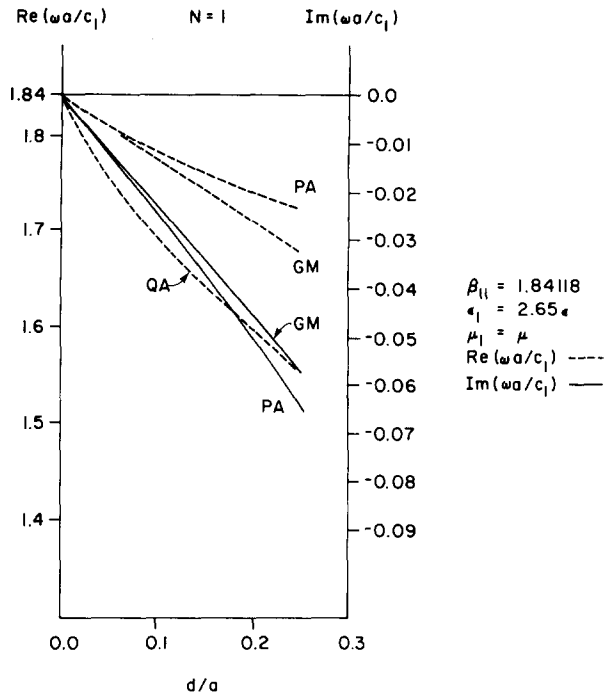



FIG. 3. The comparison of Galerkin's method (GM), perturbative approach (PA), and the quasistatic approach (QA) for the resonant frequency shift when $\epsilon_1 = 2.65\epsilon$.

$$\omega/\omega_0 = \sqrt{C_0/C}\,, \tag{46}$$

where $C_0$ is the capacitance without fringing field effect.

We note that when $d/a < 0.1$, the agreement of the perturbative approach with Galerkin's method is excellent,

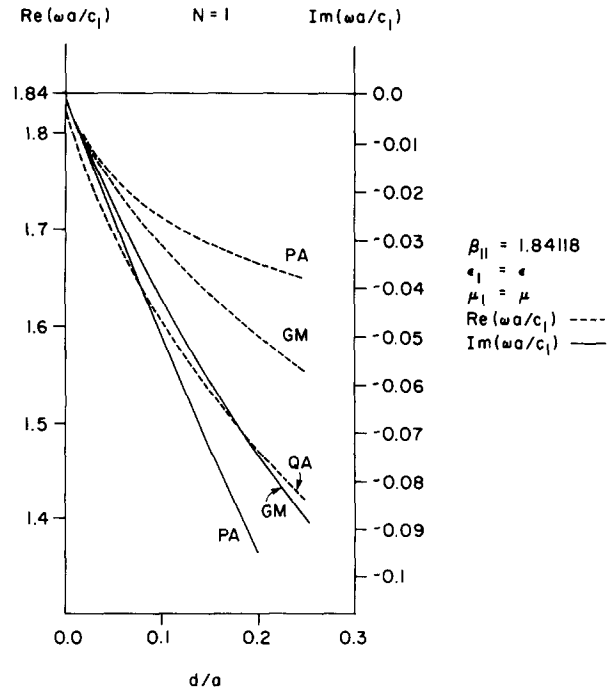



FIG. 4. The comparison of Galerkin's method (GM), perturbative approach (PA), and the quasistantic approach (QA) for the resonant frequency shift when $\epsilon_1 = \epsilon$.

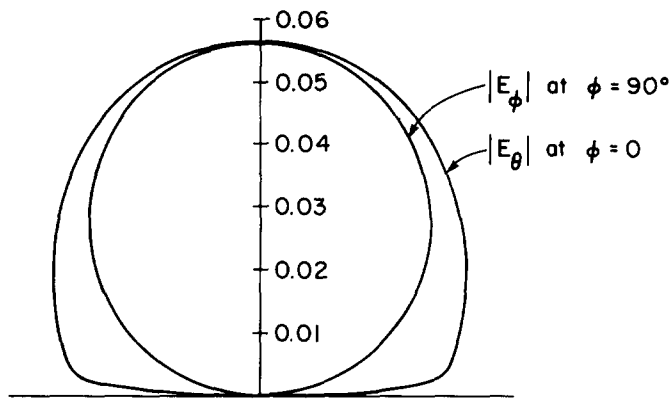



FIG. 5. Radiation pattern of the lowest $N = 1$ mode.

while there exist substantial discrepancies with the quasitatic approach. Thus, for small $d/a$, the perturbative approach reigns superior since it does not involve numerical search of zeros as in Galerkin's method.

In Fig. 4, we show the resonant frequency shift for $\epsilon_1/\epsilon = 1.0$. In this case, we find an increase in frequency shift. Also, the perturbative approach starts to deviate from Galerkin's method for $d/a > 0.05$. This is because the fringing field effect becomes important rapidly for increasing $d/a$ due to the less trapping of electric flux in the dielectric substrate for smaller $\epsilon_1/\epsilon$. We see that the discrepancy between the quasistatic approach and Galerkin's method remains substantial. Note that the quasistatic approach is not asymptotic to Galerkin's method like the perturbative approach when $d/a \to 0$.

In Fig. 5, we plot the radiation pattern from the radiation field. We note that the radiation field vanishes at the horizon ($\theta = 90°$) due to the presence of the dielectric–air boundary. Radiation field derived using free-space Green's space does not vanish at the horizon.[2,3]

From the above, we can conclude that for accurate numerical computation of the frequency shift, Galerkin's method remains superior. However, when $d/a$ is small and $\epsilon_1/\epsilon$ large, the perturbative approach is more efficient. The quasistatic approach is not as good since it gives the same fractional shift in resonant frequencies for different modes.

## APPENDIX

### Vector Hankel Transform

The vector Hankel transform, transforms a vector function from one space to a vector function in another space. First, we shall postulate the form of such a transform.

$$\begin{bmatrix} f_1(\rho) \\ f_2(\rho) \end{bmatrix} = \int_0^\infty k_\rho\, dk_\rho \times \begin{bmatrix} J_n'(k_\rho \rho) & \dfrac{n}{k_\rho \rho} J_n(k_\rho \rho) \\ \dfrac{n}{k_\rho \rho} J_n(k_\rho \rho) & J_n'(k_\rho \rho) \end{bmatrix} \cdot \begin{bmatrix} F_1(k_\rho) \\ F_2(k_\rho) \end{bmatrix}, \tag{A1a}$$

$$\begin{bmatrix} F_1(k_\rho) \\ F_2(k_\rho) \end{bmatrix} = \int_0^\infty \rho\, d\rho \times \begin{bmatrix} J_n'(k_\rho \rho) & \frac{n}{k_\rho \rho} J_n(k_\rho \rho) \\ \frac{n}{k_\rho \rho} J_n(k_\rho \rho) & J_n'(k_\rho \rho) \end{bmatrix} \cdot \begin{bmatrix} f_1(\rho) \\ f_2(\rho) \end{bmatrix}. \tag{A1b}$$

We can write (A1a) asnd (A1b) more concisely as

$$\bar{f}(\rho) = \int_0^\infty k_\rho\, dk_\rho\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{F}(k_\rho), \tag{A2a}$$

$$\bar{F}(k_\rho) = \int_0^\infty \rho\, d\rho\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{f}(\rho), \tag{A2b}$$

where $\bar{\bar{H}}_n(k_\rho \rho)$ are matrices, $\bar{F}(k_\rho)$ and $\bar{f}(\rho)$ are vectors in (A1a) and (A1b). If (A1a) and (A1b) are true, it implies that

$$\bar{f}(\rho) = \int_0^\infty k_\rho\, dk_\rho \int_0^\infty \rho'\, d\rho'\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{\bar{H}}_n(k_\rho \rho') \cdot \bar{f}(\rho'). \tag{A3}$$

We can prove (A3) by exchanging the order of integration. Writing it explicitly, we have

$$\int_0^\infty \rho'\, d\rho' \int_0^\infty k_\rho\, dk_\rho\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{\bar{H}}_n(k_\rho \rho') \cdot \bar{f}(\rho') = \int_0^\infty \rho'\, d\rho' \int_0^\infty k_\rho\, dk_\rho \times \begin{bmatrix} J_n'(k_\rho \rho) J_n'(k_\rho \rho') + \frac{n^2}{k_\rho^2 \rho\rho'} J_n(k_\rho \rho) J_n(k_\rho \rho') & \frac{n}{k_\rho \rho'} J_n(k_\rho \rho') J_n'(k_\rho \rho) + \frac{n}{k_\rho \rho} J_n(k_\rho \rho) J_n'(k_\rho \rho') \\ \frac{n}{k_\rho \rho'} J_n(k_\rho \rho') J_n'(k_\rho \rho) + \frac{n}{k_\rho \rho} J_n(k_\rho \rho) J_n'(k_\rho \rho') & J_n'(k_\rho \rho) J_n'(k_\rho \rho') + \frac{n^2}{k_\rho^2 \rho\rho'} J_n(k_\rho \rho) J_n(k_\rho \rho') \end{bmatrix} \cdot f(\rho'), \tag{A4}$$

Noting that

$$k_\rho J_n'(k_\rho \rho) J_n'(k_\rho \rho') + \frac{n^2}{k_\rho \rho\rho'} J_n(k_\rho \rho) J_n(k_\rho \rho') = k_\rho J_{n+1}(k_\rho \rho) J_{n+1}(k_\rho \rho') + \frac{n}{\rho\rho'} \frac{d}{dk_\rho} J_n(k_\rho \rho) J_n(k_\rho \rho'). \tag{A5}$$

The diagonal elements in (A4) can be evaluated since the first term in (A5) evaluates to a delta function (Ref. 19) while the second term evaluates to zero. Similarly, the off diagonal elements can be written as

$$\frac{n}{\rho'} J_n(k_\rho \rho') J_n'(k_\rho \rho) + \frac{n}{\rho} J_n(k_\rho \rho) J_n'(k_\rho \rho') = \frac{n}{\rho\rho'} \frac{d}{dk_\rho} J_n(k_\rho \rho') J_n(k_\rho' \rho), \tag{A6}$$

which integrates to give zero. Consequently, (A4) becomes

$$\int_0^\infty \rho'\, d\rho' \int_0^\infty k_\rho'\, dk_\rho'\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{\bar{H}}_n(k_\rho \rho') \cdot \bar{f}(\rho') = \int_0^\infty \rho'\, d\rho' \begin{bmatrix} \frac{\delta(\rho - \rho')}{\rho'} & 0 \\ 0 & \frac{\delta(\rho - \rho')}{\rho'} \end{bmatrix} \cdot \bar{f}(\rho') = \bar{f}(\rho), \tag{A7}$$

proving the assertion (A3).

Properties of vector Hankel transform akin to that of scalar Hankel transform can be derived. We shall list some of the properties as follows:

*Property* 1: (Symmetry)

$$\bar{\bar{H}}_n^T(k_\rho \rho) = \bar{\bar{H}}_n(k_\rho \rho). \tag{A8}$$

*Property* 2: (Delta-function representation)

$$\text{(a)} \int_0^\infty \rho\, d\rho\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{\bar{H}}_n(k_\rho' \rho) = \bar{\bar{I}} \frac{\delta(k_\rho - k_\rho')}{k_\rho'}, \tag{A9a}$$

$$\text{(b)} \int_0^\infty k_\rho\, dk_\rho\, \bar{\bar{H}}_n(k_\rho \rho) \cdot \bar{\bar{H}}_n(k_\rho \rho') = \bar{\bar{I}} \frac{\delta(\rho - \rho')}{\rho'}. \tag{A9b}$$

*Property* 3: (Reflection)

$$\text{(a)}\ \bar{F}(-k_\rho) = (-1)^{n+1} \bar{F}(k_\rho), \tag{A10a}$$

$$\text{(b)}\ \bar{f}(-\rho) = (-1)^{n+1} \bar{f}(\rho). \tag{A10b}$$

*Property* 4: (Parseval's Theorem)

$$\int_0^\infty \rho\, d\rho\, \bar{f}^T(\rho) \cdot \bar{g}(\rho)$$

$$= \int_0^\infty k_\rho \, dk_\rho \, \bar{F}^T(k_\rho) \cdot \bar{G}(k_\rho), \tag{A11}$$

where $\bar{F}(k_\rho)$ and $\bar{G}(k_\rho)$ are vector Hankel transform of $\bar{f}(\rho)$ and $\bar{g}(\rho)$, respectively.

*Proof*: By taking the transpose of Eq. (A2a) and making use of Property 1, we can show that

$$\bar{f}^T(\rho) = \int_0^\infty k_\rho \, dk_\rho \, \bar{F}^T(k_\rho) \cdot \bar{\bar{H}}_n(k_\rho \rho). \tag{A12}$$

Thus, the left-hand side of (A11) can be written as

$$\int_0^\infty \rho \, d\rho \bar{f}^T(\rho) \cdot \bar{g}(\rho)$$

$$= \int_0^\infty \rho \, d\rho \int_0^\infty k_\rho \, dk_\rho \, \bar{F}^T(k_\rho) \cdot \bar{\bar{H}}_n(k_\rho \rho).$$

$$\times \int_0^\infty k'_\rho dk'_\rho \bar{\bar{H}}_n(k'_\rho \rho) \cdot \bar{G}(k'_\rho). \tag{A13}$$

Exchanging the integrals and performing the $\rho$ integration first, and making use of Property 2, we have

$$\int_0^\infty \rho \, d\rho \bar{f}^T(\rho) \cdot \bar{g}(\rho)$$

$$= \int_0^\infty k_\rho \, dk_\rho \, \bar{F}^T(k_\rho)$$

$$\times \int_0^\infty k'_\rho \, dk'_\rho \frac{\delta(k_\rho - k'_\rho)}{k'_\rho} \bar{G}(k'_\rho)$$

$$= \int_0^\infty k_\rho \, dk_\rho \, \bar{F}^T(k_\rho) \cdot \bar{G}(k_\rho), \tag{A14}$$

proving Property 4.


[1] L. C. Shen, S. A. Long, M. R. Allerding, and M. D. Walton, "Resonant frequency of a circular disc printed circuit antenna," IEEE Trans. Antenna Propag. **AP-25**, 595–591 (1977).

[2] S. A. Long, L. C. Shen, and P. B. Morel, "Theory of the printed circuit antenna," Proc. IEE (London), **125**, 925–928 (1978).

[3] Y. T. Lo, D. Solomon, and W. F. Richards, "Theory and experiment on microstrip antennas," IEEE Trans. Antenna Propag. **AP-27**, 137–45 (1979).

[4] A. G. Derneryd, "Analysis of the microstrip disk antenna element," IEEE Trans. Antenna Propag. **AP-27**, 660–664 (1979).

[5] J. Watkins, "Circular resonant structures in microstrip," Electron. Lett. **5**, 524–525 (1978).

[6] S. Mao, S. Jones, and G. D. Vandelin, "Millimeter-wave integrated circuits," IEEE Trans. Microwave Theory Tech. **MTT-16**, 455–461 (1968).

[7] I. Wolff and N. Knoppik, "Rectangular and circular microstrip disk capacitors and resonators," IEEE Trans. Microwave Theory Tech." **MTT-22**, 857–864 (1974).

[8] S. R. Borkar and R. F. M. Yang, "Capacitance of a circular disk for applications in microwave integrated circuits," IEEE Trans. Microwave Theory Tech. **MTT-23**, 588 (1975).

[9] T. Itoh, "Analysis of microstrip resonators," IEEE Trans. Microwave Theory Tech. **MTT-22**, 946 (1974).

[10] W. C. Chew and J. A. Kong, "Resonance of the axial-symmetric modes in microstrip disk resonators," J. Math. Phys. **21**, 582 (1980).

[11] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U. P., Oxford, 1937), p. 334.

[12] C. J. Tranter, "On some dual integral equations occurring in potential problems with axial symmetry," J. Mech. Appl. Math, 411 (1950).

[13] I. N. Sneddon, *Mixed Boundary Value Problems in Potential Theory* (Wiley, New York, 1966).

[14] T. Itoh and R. Mittra, "A new method for calculating the capacitance of a circular disk for microwave integrated circuits," IEEE Trans. Microwave Theory Tech. **MTT-21**, 431–432 (1973).

[15] W. C. Chew and J. A. Kong, "Effects of fringing fields on capacitance of circular microstrip disk," IEEE Trans. Microwave Theory Tech. **MTT-28**, 98–104 (1980).

[16] J. A. Kong, *Theory of Electromagnetic Theory* (Wiley, New York, 1975), p. 215.

[17] W. C. Chew, J. A. Kong, and L. C. Shen, "Radiation characteristics of a circular microstrip antenna," J. Appl. Phys. **7**, 3907 (1980).

[18] A. Sommerfeld, *Partial Differential Equations* (Academic, New York, 1949).

[19] I. N. Sneddon, *Fourier Transforms* (McGraw-Hill, New York, 1951).