Group-theoretic determination of normal coordinates for molecular vibration

Hiroshi Watanabe
Department of Physics, Osaka City University, Osaka, Japan

(Received 13 March 1986; accepted for publication 16 April 1986)

It is shown that the normal coordinates for molecular vibration are determined by simply taking those linear combinations of mass-weighted displacements that are irreducible bases of the group relevant to the molecule.

I. INTRODUCTION

Since Wigner applied the symmetry principle to molecular vibrations many years ago,[1] group theoretical arguments have played important roles in discussing selection rules for infrared absorption by molecular vibration,[1] vibration–electronic interaction, e.g., the Jahn–Teller effect,[2] etc.

Quantization of the molecular vibration is carried out for normal coordinates, which are usually classified according to the irreducible representations of the symmetry group relevant to the molecular configuration. It is physically anticipated that the normal coordinates are at the same time the irreducible bases; however, no group theoretical proof has been published for this anticipation.[3] In this brief paper a proof is given with some remarks.

II. THEOREM AND PROOF

Let us consider a molecule consisting of N atoms that possesses a point symmetry. The molecule is transformed into itself by a set of point transformations $\{\hat{P}\}$; the set forms a group. Linear combinations of "mass-weighted displacements" $\xi_{i\alpha}$'s of atoms from the equilibrium position, $i = 1, 2,\dots,N$ and $\alpha = x, y, z$, are to be simply called *linear combinations* in the following.

When the $3N$-dimensional reducible representation $\{P\}$, with matrix P for $\{\hat{P}\}$ referred to the $\xi_{i\alpha}$'s, is completely reduced to irreducible representation that are all different, the following theorem is proved.

Theorem: The linear combinations that are irreducible bases for the point group relevant to the molecule are the normal coordinates.

The potential and kinetic energies, V and T, are written

$$2V = \sum_{i,j} \sum_{\alpha,\beta} \xi_{i\alpha} K_{i\alpha, j\beta} \xi_{j\beta} = \xi^t \cdot K \cdot \xi, \tag{1}$$

$$2T = \sum_{i} \sum_{\alpha} (\dot{\xi}_{i\alpha})^2 = \dot{\xi}^t \cdot \dot{\xi}, \tag{2}$$

where $K = (K_{i\alpha, j\beta})$ is the "mass-weighted force matrix," $\xi^t = (\cdots \xi_{i\alpha} \cdots)$, and the superscript t stands for transposition. Since the molecule is transformed by $\hat{P}$ into itself, one obtains

$$PKP^{-1} = K \quad \text{or} \quad PK = KP. \tag{3}$$

The $3N$-dimensional orthogonal matrix O, which completely reduces the representation $\{P\}$, transforms the relation (3) to

$$OPKO^{-1} = OKPO^{-1};$$

this is rewritten

$$R\Lambda = \Lambda R, \tag{4}$$

where

$$OPO^{-1} \equiv R \quad \text{and} \quad OKO^{-1} \equiv \Lambda.$$

Since the matrix R is completely reduced, it is written as

$$R = \begin{pmatrix} A & 0 & \cdot & \cdot & \cdot \\ 0 & B & 0 & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \tag{5}$$

where the 0's are generally rectangular null matrices, and A, B,... are square matrices. The sets $\{A\}$,$\{B\}$,... are irreducible representations, respectively. Corresponding to (5), the matrix Λ is written as

$$\Lambda = \begin{pmatrix} \Lambda_1 & X & \cdot & \cdot & \cdot \\ Y & \Lambda_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{6}$$

The matrix relation (4) is further rewritten in more detail

$$A\Lambda_1 = \Lambda_1 A, \quad B\Lambda_2 = \Lambda_2 B,\dots, \tag{7}$$

and

$$AX = XB, \quad BY = YA,\dots . \tag{8}$$

Since the matrix relations (7) hold for the irreducible representations, $\{A\}$,$\{B\}$,..., the submatrices Λ_1,Λ_2,... are constant matrices by Schur's lemma.[4]

When the reduction of $\{P\}$ contains each irreducible representation only once, i.e., the irreducible representations $\{A\}$,$\{B\}$,... are all different, the submatrices X,Y,... are null on account of (8).[4] Therefore, one obtains

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & \cdot & \cdot & \cdot \\ 0 & \Lambda_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{9}$$

The potential and kinetic energies, (1) and (2), with the replacement

$$O\xi = Q, \tag{10}$$

are rewritten as

$$2V = (Q\xi)^t \cdot \Lambda \cdot (O\xi) = Q^t \cdot \Lambda \cdot Q, \tag{11}$$

$$2T = (O\dot{\xi})^t \cdot (O\dot{\xi}) = \dot{Q}^t \cdot \dot{Q}. \tag{12}$$

Since the matrix O transforms the bases $\xi_{i\alpha}$'s into the linear combinations that are irreducible bases, the expressions (11) and (12) clearly show that the column matrix Q is that of normal coordinates.

 0022-2488/86/081931-02$02.50

III. REMARKS

When all the irreducible representations $\{A\},\{B\},\ldots$ are different, the normal coordinates are easily determined by finding those linear combinations that are irreducible bases for the point group. When two or more vibrational modes are classified according to one and the same irreducible representation, the above method fails to uniquely determine their vibrational normal coordinates.

When there are translational (and/or rotational) and vibrational motions that are classified according to the same irreducible representation, the following is to be noted: The vibrational normal coordinates can be determined by taking the linear combinations to be orthogonal to the corresponding translational (and/or rotational) normal coordinates, which are easily obtained by inspection.

ACKNOWLEDGMENT

The author heartily thanks Professor E. P. Wigner for his encouragement and suggestions.

[1] E. Wigner, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl. **1930**, 133.
[2] H. A. Jahn and E. Teller, Proc. R. Soc. London Ser. A **161**, 220 (1937); H. A. Jahn, *ibid.* **164**, 117 (1938).
[3] E. P. Wigner (private communication).
[4] E. P. Wigner, *Group Theory and its Application to Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 9, pp. 75 and 76.

The state labeling problem—A universal solution

A. J. Coleman
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

(Received 12 September 1985; accepted for publication 26 March 1986)

A new method is proposed for labeling a basis for any irreducible representation, with top weight, of a simple Lie algebra (LA). It has two advantages over Gel'fand–Zetlin patterns: (i) being exactly the same for all LA's both exceptional and classical and (ii) providing labels that are much more compact. Thus it may prove useful in discussing large representations such as those $E_8 \otimes E_8$ that occur in superstring theory. The method makes essential use of the Weyl group and is based on a theorem that associates to each weight a symmetric matrix with integer coefficients whose rank equals the multiplicity of the weight.

I. INTRODUCTION

Just as Noah had a big problem in keeping track of the animals in the ark, so in quantum mechanics we often encounter difficulty in giving names to the possible pure states of a system. In the past two decades irreducible representations of various Lie groups have been brought into play as a tool to solve this problem. The chain of unitary groups $\mathrm{U}(m) \supset \mathrm{U}(m-1)$, $n \geqslant m \geqslant 1$, was used by Gel'fand and Zetlin[1] to give unique names for the elements of a basis for any irreducible representation of $\mathrm{U}(n)$. These names, sometimes known as *GZ patterns*, are widely used. A group theoretical explanation of the GZ pattern was given by Baird and Biedenharn.[2] Later, a similar procedure was proposed by Gould[3] for representations of the orthogonal group. There is now a considerable literature devoted to variations on the basic approach used by Gel'fand and Zetlin in which a chain of subgroups, chosen as most appropriate to the particular problem under consideration, is used to name the states of the system. Once a system for labeling states is adopted it becomes of interest to calculate the matrices for generators of the group [formulas (C9)] and to obtain the Wigner–Gordan–Clebsch coefficients for the decomposition of tensor products of irreducible representations. The chief object of this paper is to propose a *uniform method of defining and labeling an orthonormal basis for the states of any irreducible representation of a simple Lie group possessing a top weight.* To this end we shall use the notation for irreducible representations of simple Lie groups introduced by Cartan in his remarkable paper on the representations of complex simple Lie groups.[4] This has the advantage over some notations currently widely used by physicists of being a common notation applicable to *all* irreducible representations of *all* simple Lie groups. As will be seen in Appendix B, it is also much more compact than the GZ patterns.

For an irreducible representation with top weight π we shall associate to each dominant weight μ occurring in the representation a matrix C^μ, with integer entries and degree s_μ. We shall prove that the *rank* of C^μ is the *multiplicity* m_μ of μ. To m_μ linearly independent columns of C^μ we shall associate vectors v_μ^i, $1 \leqslant i \leqslant m_\mu$, which span the μ weight space. There is considerable flexibility in the choice of v_μ^i. For example, using only integer arithmetic, we can arrange that they are mutually orthogonal. Normalization will require at most extracting the square root of an integer. It should not be too difficult to arrange that the v_μ^i are adapted to any chosen chain of subgroups.

By action of the Weyl group, the set $\{v_\mu^i\}$ for a dominant μ can be mapped onto a basis for the weight space of any weight conjugate to μ. Consequently any element of an orthogonal basis can be labeled by (μ, b_i^μ, w), where μ is a dominant weight, b_i^μ could be an s_μ-tuple of integers, and w is a representative of an element of W/W_μ. Here W is the Weyl group and W_μ is the stability group of μ.

It must be admitted that the above notation will seem novel and somewhat esoteric especially to satisfied GZ users! However, it should easily be amenable to manipulation by computers and the fact that it is exactly the same for all simple Lie groups will, hopefully, make it popular. In fact, once standard tables have appeared for any group, the w can be deleted from the label, the restriction of μ to be dominant can be lifted, and the label could take the form (μ, b_i^μ), where μ is a weight and $1 \leqslant i \leqslant m_\mu$.

II. IRREDUCIBLE REPRESENTATIONS OF SIMPLE LIE GROUPS

In this section we recall some basic ideas about the irreducible representations of simple Lie algebras and describe our notation. Readers who are familiar with these matters should skip to the statement of the Cartan theorem in which our notation is defined.

The facts we summarize about the representations of Lie groups, which were known to Cartan before 1913, but even in 1960 seemed esoteric to most mathematicians, are now widely known because of excellent expositions in several recent books.[5] However, the topic is so rich that no consensus has yet been achieved as to the most efficient notation. What is here proposed results from 35 years meditation by the author who had the privilege of studying with Coxeter in Toronto and Chevalley in Princeton.

In order to state the essential facts as quickly as possible, assume that our group *G is simple and compact*. Then a maximal connected Abelian subgroup will be *toroidal*, that is, a direct product of n circle groups. Such a maximal toroid, $T \subset G$, is a *Cartan subgroup* (CSG). All CSG's have the same dimension and are conjugate in G. The *dimension* of a CSG, for which we shall reserve the letter n, is called the

 0022-2488/86/081933-11$02.50

rank of G. The rank is an important invariant, playing a key role in the classification of simple Lie groups.

Killing[6] showed that a simple Lie group belongs *either* to one of four infinite classes, now denoted by A_n $(n\geqslant 1)$, B_n $(n\geqslant 2)$, C_n $(n\geqslant 3)$, and D_n $(n\geqslant 4)$ *or* is one of five exceptional groups, of rank 2, 4, 6, 7, or 8 denoted by G_2, F_4, E_6, E_7, and E_8. The four infinite classes contain[7] the so-called *classical groups*. Thus A_n contains $\mathrm{SU}(n+1)$; B_n, $\mathrm{SO}(2n+1)$; C_n, $\mathrm{Sp}(2n)$; and D_n, $\mathrm{SO}(2n)$.

Cartan[4] showed how to obtain the finite-dimensional irreducible representations of a simple Lie group. To any such group he associated n inequivalent *fundamental* representations $\Gamma_1,\Gamma_2,...,\Gamma_n$, which can be described, roughly, as the n lowest-dimensional irreducible representations by matrix groups. Then for each choice of $\pi=(p_1,p_2,...,p_n)$, where p_i is a non-negative integer, there is a unique irreducible representation of G that appears as the first representation in the reduction of the tensor product $\Gamma_1^{p_1}\times\Gamma_2^{p_2}\times\cdots\times\Gamma_n^{p_n}$ into irreducibles. Cartan thus established a 1:1 mapping between finite-dimensional irreducible representations of the simple group G of rank n and n-tuples of non-negative integers.

The reader is doubtless aware that the study of a Lie group G is greatly facilitated by means of its Lie algebra (LA) L. The LA of G is a *vector space* that can be thought of as the tangent space to G at the neutral or identity element I. This L is closed under formation of linear combinations, that is for $x,y\in L$, and $a,b\in\mathbb{R}$, $ax+by$ also belongs to L, but there is another binary operation that associates to x and y their "product," an element of L that we denote[8] by $x\circ y$. This operation satisfies two conditions, for all $x,y,z\in L$:

(i) $x\circ y+y\circ x=0$,

(ii) $x\circ(y\circ z)+y\circ(z\circ x)+z\circ(x\circ y)=0$.

Condition (i) says that the Lie product anticommutes. The so-called *Jacobi* identity (ii) is a reflection in the LA of the associative property of multiplication in the group.

Fix a CSG T of a simple Lie group (LG) G, then the LA of T, which we shall denote by $\mathscr{H}$, will be *Abelian*, that is, $h,k\in\mathscr{H}\Rightarrow h\circ k=0$. Recall that n, the rank of G, is the dimension of $\mathscr{H}$. Therefore a basis $\{h_i\}$ of $\mathscr{H}$ will consist of n linearly independent elements of $\mathscr{H}\subset L$. It can be shown that if G is compact there exists a negative definite quadratic form on L, which is invariant under the action of G, on the tangent space at the identity, induced by the map $\hat{g}$: $G\to G$ defined by $s\to gsg^{-1}=\hat{g}s$ for all $s\in G$. Further, the basis of $\mathscr{H}$ can be supplemented by $\{e_\alpha\}$, where $\alpha\in\Sigma$ takes $2m$ values such that $n+2m=r$ is the dimension of L and $\{h_i,e_\alpha\}$ is a basis of L such that for all $h\in\mathscr{H}$ and all $i,j\in[1,2,...,n]$

(i) $h_i\circ h_j=0$;

(ii) $h\circ e_\alpha=\alpha(h)e_\alpha$, for $\alpha\in\Sigma$,

where $\alpha(h)$ is a linear function of h;

(iii) $e_\alpha\circ e_\beta=N_{\alpha\beta}e_{\alpha+\beta}$.

The index set Σ is crucial, determining which particular simple LA is being described. The Coxeter–Dynkin diagram discussed below is essentially a symbol that encodes the structure of Σ. For some pairs (α,β), $\alpha+\beta\in\Sigma$ and Σ satisfies the following two conditions.

(i) $\alpha\in\Sigma\Rightarrow-\alpha\in\Sigma$.

(ii) If $\alpha,\beta\in\Sigma$ then if $\alpha+\beta\in\Sigma$, $N_{\alpha\beta}\neq 0$ whereas if $\alpha+\beta\notin\Sigma$, $N_{\alpha\beta}=0$. Thus in the latter case $e_\alpha\circ e_\beta=0$. The e_α are unique up to multiplicative factors that can be chosen so that $N_{-\alpha,-\beta}=-N_{\alpha\beta}$. When this is done the $N_{\alpha\beta}$ are integers equal numerically to 4 at most. Only for G_2 is 4 attained.

The linear forms $\alpha(h)$ defined on $\mathscr{H}$ are called the *roots* of the LA because they first appeared as the roots of an equation of degree r, which played a key role in Killing's classification of the simple LA's. Since $\mathscr{H}$ is of dimension n there are at most n linearly independent α. Since $x\cdot y$, the restriction to $\mathscr{H}$ of the above-mentioned invariant quadratic form, is nondegenerate, it is possible to find vectors r_α in $\mathscr{H}$ such that $\alpha(h)=r_\alpha\cdot h$ for all $h\in\mathscr{H}$. We can identify α with r_α and regard r_α as a member of the set Σ. Property (i) of Σ implies that together with r_α, $-r_\alpha$ belongs to Σ.

A fact of capital importance, which was known to Cartan[9] in 1927, is that it is possible in many ways to choose n linearly independent roots, which we denote by $\{r_i\}$, $1\leqslant i\leqslant n$, such that for any $r_\alpha\in\Sigma$, $r_\alpha=m_\alpha^i r_i$ with m_α^i *integers* and for a given α all nonzero m_α^i have the same sign. The roots r_α for which $m_\alpha^i\geqslant 0$ are called *positive* roots. A set of n roots with this property is called a set of *simple* roots because none of them can be expressed as a sum of two other *positive* roots. It is implicit in Cartan's paper that the number of possible ways of selecting a set of simple roots is equal to the order of the Weyl group.

The *Weyl group* W is a group of orthogonal transformations of $\mathscr{H}$ generated by the n reflections

R_i: $h\to h-2[(h\cdot r_i)/(r_i\cdot r_i)]r_i$.

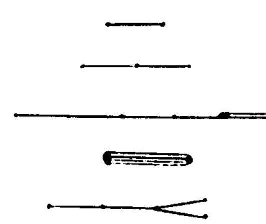
The R_i permute the roots. That is, Σ *is invariant under the action of the Weyl group*. From this it follows that W is a finite group. Although he did not have an interpretation of the Weyl group as orthogonal transformations of $\mathscr{H}$, Killing[6] made essential use of the very same group, which he regarded as a permutation group of the elements of Σ. It is fairly easy to deduce that

$2(r_\alpha\cdot r_\beta)/(r_\beta\cdot r_\beta)$

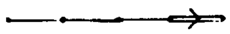
is 0, ± 1, ± 2, or ± 3. This greatly restricts Σ and led to Killing's classification of simple LA's.

Coxeter studied finite groups generated by reflections and in 1931 introduced a graph[10] with n nodes to characterize the irreducible groups. A node corresponds to one of the n reflections R_i that generate the group. Two nodes i and j are joined if R_iR_j has order greater than 2. The branch joining the i and j node is marked with the order of R_iR_j. For simple LA's the only orders for R_iR_j that can occur are 2, 3, 4, or 6. In fact, only when $n=2$ can 6 occur. Since $(R_iR_j)^2=I\Rightarrow R_iR_j=R_jR_i$, Witt[11] introduced the convention that if $(R_iR_j)^{m_{ij}}=I$, the i and j nodes are joined by $m_{ij}-2$ branches. With these conventions, for example (i) A_2, (ii) A_3, (iii) B_5, (iv) G_2, and (v) D_5 have the graphs

(i)
(ii)
(iii)
(iv)
(v)



In these diagrams instead of marking the order p of $R_i\, R_j$ we have indicated it by joining the i and j nodes by $p-2$ bonds. When two nodes are joined by an *odd* number of bonds, the corresponding R_i are conjugate in W and the r_i are of equal length. When two nodes are joined by an *even* number of nodes, which occurs once in the graph of G_2, F_4, B_n, and C_n, the corresponding r_i differ in length. A happy convention, the inventor of which I have not been able to discover, adds an inequality sign to the branch and thus



denotes B_5 indicating that $|r_5| < |r_4|$, whereas

denotes C_5. The best compilation of information about the roots and these graphs known to the author is in the Appendices to the idosyncratic book[12] of Freudenthal and de Vries.

In 1944, Dynkin[13] rediscovered the concept of a set of simple roots and, independently of Coxeter and Witt, introduced a diagram similar to Coxeter's to describe the interrelation of the r_i in such a set. Bourbaki learned of the idea from Dynkin's article and named the graph a *Dynkin diagram*. After 1949, when I drew Chevalley's attention to Coxeter's work, Bourbaki introduced a distinction between the Coxeter and the Dynkin diagrams. Since the two diagrams convey identical information I regard the distinction as spurious. Indeed, I was once vigorously attacked for not admitting that the great contribution of Dynkin was putting the inequality sign on the diagram when in fact there are no such signs in the diagrams in Dynkin's article—at least in the AMS Translation! Dynkin's contribution to the study of LA's is certainly considerable so I feel justice to history is done by referring to these very useful graphs as Coxeter–Dynkin diagrams or CD diagrams for short.

We have not yet exhausted the meaning of the CD diagrams. So far a node denotes *either* a simple root r_i *or* the corresponding R_i, which is one of n generating reflections of the Weyl group W. However, we can still associate to the nodes the n fundamental representations Γ_i referred to above.

Geologists and solid-state physicists who wander around in Brillouin zones are aware of the advantage of defining a dual or reciprocal basis. Thus corresponding to the basis $\{r_i\}$ of simple roots in $\mathscr{H}$, there is a dual basis $\{r^i\}$ such that

$$r^i \cdot r_j = \delta^i_j.$$

This has the property that if k_i are integers then $(k_i r^i)\cdot r_\alpha = (k_i\, r^i)\cdot m^j_\alpha r_j = k_j\, m^j_\alpha$ *is an integer*. So $\{r^i\}$ spans the lattice of points h in $\mathscr{H}$ such that $h\cdot r_\alpha$ is an integer for every α. These are points of $\mathscr{H}$ that Cartan recognized as corresponding to elements of the center of G under the exponential map.

For any representation Γ: $g\to\Gamma(g)$, where $\Gamma(g)$ is a linear operator on a complex linear space V, since a CSG T is Abelian it is possible to choose a basis of V consisting of simultaneous eigenvectors of T. The corresponding representation of $\mathscr{H}$, which we also denote by Γ, will have the same set of simultaneous eigenvectors. The Γ-image of an element of L will be denoted by capital letters. Thus $\Gamma(h) = H$, $\Gamma(e_\alpha) = E_\alpha$. Suppose x_λ is an eigenvector of H, then $Hx_\lambda = f(h)x_\lambda$, where it can be shown that $f(h)$ is a linear function on $\mathscr{H}$. Thus there is a vector $\lambda\in\mathscr{H}$ such that $f(h) = \lambda\cdot h$ so for all $h\in\mathscr{H}$

$$Hx_\lambda = (\lambda\cdot h)x_\lambda. \tag{2.1}$$

When (2.1) holds for $x_\lambda \neq 0$ and $\lambda \neq 0$, we say that $\lambda\cdot h$ (or, by abuse of language, λ) is a *weight* of Γ corresponding to *weight vector* x_λ. In the special case $V = L$, *weights* are called *roots*.

Since in a representation of a LA, $x\circ y \to [X,Y] = XY - YX$,

$$HE_\alpha x_\lambda = [H,E_\alpha]x_\lambda + E_\alpha H x_\lambda$$
$$= r_\alpha\cdot hE_\alpha x_\lambda + \lambda\cdot hE_\alpha x_\lambda = (\lambda + r_\alpha)\cdot hE_\alpha x_\lambda.$$

It follows that $E_\alpha x_\lambda$ has weight $\lambda + r_\alpha$. Thus $y^p = E^p_\alpha x_\lambda$ has weight $\lambda + pr_\alpha$ *or* $y^p = 0$. Since for distinct p, y^p are linearly independent, in a finite-dimensional representation $y^p \neq 0$ for only a finite set of p's. Since $E_{-\alpha}$ corresponds to $-r_\alpha$, starting from any $x_\lambda \neq 0$, there is maximal finite chain $(\lambda + sr_\alpha)$, $-q \leqslant s \leqslant p$, of weights corresponding to nonvanishing eigenvectors. From this it easily follows that

$$2[(\lambda\cdot r_\alpha)/(r_\alpha\cdot r_\alpha)] = q - p, \tag{2.2}$$

for any α. If we set

$$h_\alpha = [2/(r_\alpha\cdot r_\alpha)]r_\alpha, \tag{2.3}$$

condition (2.2) implies

$$\lambda\cdot h_\alpha \in \mathbb{Z}, \tag{2.4}$$

where $\mathbb{Z}$ denotes the integers, and in particular

$$\lambda\cdot h_i \in \mathbb{Z}, \tag{2.5}$$

for $1 \leqslant i \leqslant n$, where $\{h_i\}$ is clearly a basis for $\mathscr{H}$. Let $\{h^i\}$ be the basis of the $\mathscr{H}$ dual to $\{h_i\}$ such that

$$h^i\cdot h_j = \delta^i_j. \tag{2.6}$$

Then if $\lambda = p_i h^i$,

$$\lambda\cdot h_j = p_j \in \mathbb{Z}. \tag{2.7}$$

Conversely (2.7) implies (2.4).

We can now formulate the *Cartan theorem for representations of simple Lie algebras.*

Let Γ be a finite-dimensional *irreducible* representation of a simple LA L, which associates to $x\in L$, a linear operator $\Gamma(x) = X$ on a carrier space V such that for $a,b\in\mathbb{C}$, $x, y\in L$,

(i) $\Gamma(ax + by) = aX + bY$,

(ii) $\Gamma(x\circ y) = [X,Y] = XY - YX$.

Then there is a *top vector* x_π such that

(i) $Hx_\pi = (\pi\cdot h)x_\pi$, for all $h\in\mathscr{H}$,

(ii) $E_i x_\pi = 0$, for $1 \leqslant i \leqslant n$,

and n simple *positive* roots, r_i,

(iii) $\pi = p_i h^i$, with p_i non-negative integers. (2.8)

If we denote $\Gamma(e_{-\alpha})$ by F_α, then V is spanned by

$$F_{i_1}F_{i_2},\ldots,F_{i_k}x_\pi, \tag{2.9}$$

where $1 \leqslant i_j \leqslant n$ (E_α and F_α are called *raising* and *lowering* operators, respectively). A vector x_μ of the form (2.9) has weight

$$\mu = \pi - k^i r_i, \tag{2.10}$$

where k^i are non-negative integers. We shall call $k = \Sigma k^i$ the *depth* of x_μ. For fixed μ, the set of all vectors (2.9), for the $s_\mu = k!/\Pi k^i!$ permutations of the subscripts, span the μ weight space V_μ. The carrier space V is a direct sum of the V_μ over all weights μ which actually occur in V.

Remark 1: The above, formulated slightly differently, is the main content of Cartan's paper.[4]

Remark 2: For fixed μ the s_μ vectors (2.9) are, in general, not linearly independent. Indeed some may be zero. The dimension of V_μ is called the *multiplicity* of the weight μ and denoted by m_μ.

Remark 3: It follows from (2.10) that the *difference* between any two weights of Γ_π is an integral combination of the *roots* of L. This is another indication of the importance of Σ for the structure and properties of L and G.

Remark 4: The top weight π of an irreducible representation is not necessarily an integral combination of roots. Indeed it is precisely because half-integer combinations occur for representations of the orthogonal groups that fundamental particles exhibit "spin." Thus Cartan's paper, in which spinors were defined for the first time, may justly be regarded as the most basic mathematical paper for quantum chemistry.

Remark 5: Weyl[14] obtained an explicit form for the *character* of an irreducible representation of a simple LA from which it is easy to deduce the *dimension* of the carrier space V of the irreducible representation Γ_π. Define

$$\delta = \frac{1}{2}\sum_{\alpha>0} r_\alpha = \sum_i h^i. \tag{2.11}$$

Then the dimension d_π of V is given by

$$d_\pi = \frac{\Pi(\pi+\delta)\cdot r_\alpha}{\Pi\delta\cdot r_\alpha}, \tag{2.12}$$

where the products are over all positive roots.

For example, for A_1 or su(2), $h^1 = \frac{1}{2}r^1$, $\pi = mh^1$, $0 \leqslant m$. The product in (2.12) has one factor for r_1 and $\delta = \frac{1}{2}r_1$. So

$$d_\pi = [(\tfrac{1}{2}m + \tfrac{1}{2})r^1\cdot r_1]/(\tfrac{1}{2}r^1\cdot r_1) = m+1,$$

corresponding to the well-known fact that all positive integers occur as dimensions of representations of su(2).

Remark 6: If in succession we set p_i in (2.8) (iii) equal to δ_{ij} we obtain the *fundamental* irreducible representations Γ_j corresponding, respectively, to h^j. In the reduction of the tensor product $\Gamma_1^{p_1}\otimes\Gamma_2^{p_2}\otimes\cdots\otimes\Gamma_n^{p_n}$ into irreducibles, the irreducible Γ_π, with top weight $\pi = p_i h^i$, occurs exactly once. The top weight of any *other* irreducible occurring in this tensor product is below π. This was Cartan's original observation.

III. THE MATRIX C^μ

If $m_\mu = 1$ for all weights μ that occur in a representation, then μ can be used as a label for a basis vector in the one-dimensional space V_μ. However, the state-labeling problem is nontrivial if $m_\mu > 1$ and is that of naming a basis of V_μ for those μ that actually occur as a weight in the irreducible representation.

Our approach involves associating to each weight μ of Γ_π a matrix C^μ of a degree s_μ such that m_μ, the multiplicity of μ, equals the rank of C^μ. That is, $m_\mu = \mathrm{rank}(C^\mu)$. Thus we can already infer that $m_\mu \leqslant s_\mu = k!/\Pi k^i!$.

If $\mu = \pi - 2r_1 - r_2$ and x_π is a top vector for Γ_π, then V_μ is spanned by

$$F_1F_1F_2x_\pi,\ F_1F_2F_1x_\pi, \text{ and } F_2F_1F_1x_\pi,$$

corresponding to the $s_\mu = 3!/1!2!$ permutations of $\{1,1,2\}$. It is also helpful to associate these three vectors with "paths" from π to μ, e.g., $F_1F_1F_2x_\pi$ corresponds to the path

$$\pi\to\pi - r_2\to\pi - r_2 - r_1\to\pi - r_2 - 2r_1,$$

each step of which goes "deeper" into the potential weight spaces of Γ_π. We say "potential" because, for example, if $\pi = 2h^1 + 3h^3$, then, as we shall see, $\pi - r_2$ would not occur as a weight in Γ_π and $F_2x_\pi = 0$ so that $F_1F_1F_2x_\pi = 0$ and we could immediately conclude that $m_\mu < 3 = s_\mu$.

If $\pi - \mu = k^i r_i$, $k^i\in\mathbb{Z}$, $k^i\geqslant 0$, $\Sigma k^i = k$, then by a *path from π to μ* we shall mean a *permutation of the set of* k^1, 1's, k^2, 2's,...,k^n, n's. Thus if in the preceding example ($\mu = \pi - 2r_1 - r_2$), we say that A is a path from π to μ, then A is one of the three sequences (1,1,2), (1,2,1), (2,1,1).

If $A = (1,1,2)$, then F_A will denote the lowering operator $F_A = F_1F_1F_2$. But E_A will denote a raising operator with the order of the sequence A *reversed*. Thus $E_A = E_2E_1E_1$ for our present example. For all paths A and B from π to μ, $E_AF_Bx_\lambda$ will be a multiple, possibly zero, of x_π. We define C^μ_{AB} by

$$E_AF_Bx_\pi = C^\mu_{AB}x_\pi. \tag{3.1}$$

Main Theorem: The matrix $C^\mu = (C^\mu_{AB})$ has the following properties.

(i) The rank of C^μ equals m_μ, the dimension of the weight space V_μ, i.e., the multiplicity of μ.

(ii) The elements C^μ_{AB} are integers.

(iii) $C^\mu_{AB} = C^\mu_{BA}$, that is, C^μ is symmetric.

Proof: The proof of (i) is straightforward involving only basic linear algebra. We shall denote by P^π_μ the set of possible paths from π to μ. Thus, the cardinality of P^π_μ, $|P^\pi_\mu| = s_\mu$. For x_π the top vector on V, the vectors $v_B = F_Bx_\pi$, for $B\in P^\pi_\mu$, span V_μ. If $m_\mu < s_\mu$, there exist linear dependencies among v_B. Suppose

$$b^Bv_B = b^BF_Bx_\pi = 0, \tag{3.2}$$

where not all the scalars b^B vanish. Therefore

$$b^BE_AF_Bx_\pi = b^BC^\mu_{AB}x_\pi = 0.$$

Now $x_\pi \neq 0$, therefore

$$C^\mu_{AB}b^B = 0, \tag{3.3}$$

for all A. Thus (b^B) is in the null space of C^μ. This is the case for each relation (3.2) among $\{v_B\}$. This implies that $\mathrm{rank}(C^\mu)\leqslant m_\mu$.

But conversely, suppose (3.3) holds for all $A\in P^\pi_\mu$, then (3.2) must hold. Because if not, then

$$v = b^BF_Bx_\pi \neq 0 \quad\text{and}\quad v\in V_\mu.$$

If $E_iv = 0$ for all i then v would be a nonzero top vector but in an irreducible representation there is only *one* top vector, which in our case is x_π. Thus there is at least one i for which $E_i v \neq 0$. By induction there must exist a path A such that $E_A v \neq 0$. But $E_A v = b^B C^\mu_{AB} x_\pi = 0$ by (3.3). This proves (i).

The proof of (ii) [and (iii)] could get a bit messy if we try to set it out in complete detail. The argument will seem familiar to anyone who has used Wick's theorem for expanding Green's functions. Its essential aspects become clear from particular cases. We first recall that if $i \neq j$ then $F_i E_j = E_j F_i$. This follows from the fact that for simple positive roots r_i and r_j, $r_i - r_j \notin \Sigma$. For if $r_i - r_j$ is a positive root, then $r_i = (r_i - r_j) + r_j$ would be the sum of two positive roots, which is impossible for a simple root. Whereas if $r_i - r_j$ is negative, then $r_j - r_i$ would be positive and $r_j = (r_j - r_i) + r_i$. Therefore $e_i \circ f_j = 0$, which implies $[E_i, F_j] = 0$ or $E_i F_j = F_j E_i$.

Recall also that $\pi = p_i h^i$ with p_i non-negative integers. Further $[E_i F_i] = H_i$, and $H_i x_\pi = (\pi \cdot h_i) x_\pi$. *First* consider μ of depth 1, that is, $k = \Sigma k^i = 1$. Suppose $P^\pi_\mu = \{(i)\}$. Take $A = B = (i)$,

$$
\begin{aligned}
E_i F_i x_\pi &= [E_i F_i] x_\pi + F_i E_i x_\pi \\
&= H_i x_\pi + 0 \\
&= (\pi \cdot h_i) x_\pi = p_i x_\pi .
\end{aligned}
$$

Thus the weight $\pi - r_i$ of depth $k = 1$ actually occurs in Γ_π if and only if $p_i > 0$. Thus $C^\mu = (p_i)$, which is integral and symmetric!

Now, suppose $A = B = (1,2)$. Calculate C^μ_{AB}:

$$
\begin{aligned}
C^\mu_{AB} x_\pi = E_A F_B x_\pi &= E_2 E_1 F_1 F_2 x_\pi \\
&= E_2 [E_1 F_1] F_2 x_\pi + E_2 F_1 E_1 F_2 x_\pi .
\end{aligned} \tag{3.4}
$$

But $E_1 F_2 = F_2 E_1$ and $E_1 x_\pi = 0$ since x_π is the top vector, so the second term vanishes. Further $[E_1 F_1] = H_1$, and $F_2 x_\pi$ has weight $\pi - r_2$. Thus the first term in (3.4) is

$$
(\pi - r_2) \cdot h_1 E_2 F_2 x_\pi = (p_1 - a_{12}) p_2 x_\pi . \tag{3.5}
$$

Thus $C^\mu_{AA} = (p_1 - a_{12}) p_2$. Similarly if $B = (2,1)$, $C^\mu_{BB} = (p_2 - a_{21}) p_1$. This is an integer since p_1, p_2, and a_{12} are integers. Here $a_{12} = h_1 \cdot r_2$. In general the *Cartan matrix* $a_{ij} = h_i \cdot r_j$ is integral. Now, let $A = (1,2)$, $B = (2,1)$,

$$
\begin{aligned}
C^\mu_{AB} x_\pi = E_2 E_1 F_2 F_1 x_\pi &= E_2 F_2 E_1 F_1 x_\pi \\
&= p_1 E_2 F_2 x_\pi = p_1 p_2 x_\pi ,
\end{aligned}
$$

$C^\mu_{AB} = p_1 p_2$, which is an integer.

Clearly C^μ_{BA} is obtained from C^μ_{AB} by interchanging 1 and 2 so $C^\mu_{AB} = C^\mu_{BA}$ as asserted in (iii).

As a *final example*, let $A = (1,1,2)$, $B = (1,2,1)$, and calculate $C^\mu_{AB} = C^\mu_{BA}$:

$$
\begin{aligned}
C^\mu_{AB} x_\pi &= E_2 E_1 E_1 F_1 F_2 F_1 x_\pi = (\pi - r_1 - r_2) \cdot h_1 E_2 E_1 F_2 F_1 x_\pi + E_2 E_1 F_1 E_1 F_2 F_1 x_\pi \\
&= (p_1 - a_{11} - a_{12}) p_1 p_2 x_\pi + p_1 E_2 E_1 F_1 F_2 x_\pi = (p_1 - a_{11} - a_{12}) p_1 p_2 x_\pi + p_1 (p_1 - a_{12}) p_2 x_\pi ,
\end{aligned} \tag{3.6}
$$

$$
\therefore C^\mu_{AB} = p_1 p_2 (2p_1 - 2 - 2a_{12})
$$

—an integer! Now

$$
\begin{aligned}
C^\mu_{BA} x_\pi &= E_1 E_2 E_1 F_1 F_1 F_2 x_\pi = (\pi - r_1 - r_2) \cdot h_1 E_1 E_2 F_1 F_2 x_\pi + E_1 E_2 F_1 E_1 F_1 F_2 x_\pi \\
&= (p_1 - a_{11} - a_{12}) p_1 p_2 x_\pi + (\pi - r_2) \cdot h_1 E_1 E_2 F_1 F_2 x_\pi = (p_1 - a_{11} - a_{12}) p_1 p_2 x_\pi + (p_1 - a_{12}) p_2 p_1 x_\pi ,
\end{aligned}
$$

$$
\therefore C^\mu_{AB} = C^\mu_{BA} . \tag{3.7}
$$

Notice that in the second terms on the right-hand side (rhs) of (3.6) and (3.7), though the order in which the factors appear in the calculation is different, the factors are the same.

The calculation is like peeling off levels of an onion except we start from the inside, that is, with the kth raising operator and work it over to the right until it annihilates x_π. A contribution to the final expression comes only from the encounter of an E_i with an F_i.

To prove (iii) we merely notice that interchanging A and B leaves the set of such encounters unchanged. Ordinary mortals may find this argument unconvincing but Wick's theorem practitioners will accept it in a flash!

There is a neater proof of (iii), which is not quite as elementary since it assumes that a compact group has a unitary representation in which $E_i^\dagger = F_i$. Then if *we take x_π of unit norm*

$$
\begin{aligned}
C^\mu_{AB} &= \langle x_\pi | E_A F_B x_\pi \rangle \\
&= \langle F_A x_\pi | F_B x_\pi \rangle ,
\end{aligned}
$$

so C^μ_{AB} is Hermitian, but since it is real it is symmetric. Also in this interpretation C^μ is the so-called Grammian or Gram matrix of the set of vectors $F_A x_\pi$. It is well known that the dimension of the space spanned by a set of vectors is equal to the rank of their Grammian. In this way we have another proof of (i) and can even conclude that for unitary representations C^μ is positive semidefinite.

Remark 1: It is clear from our discussion of (ii) that the calculation of C^μ is straightforward and easily could be programmed for computer once the Cartan matrix (a_{ij}) is given and this is strictly equivalent to the choice of simple LA. For given $k^i r_i = \pi - \mu$, C^μ can be given in terms of π and (a_{ij}) in a form *universally valid for all simple* LA's.

Remark 2: The calculation of C^μ does not require that Γ_π be finite dimensional but only that it has a unique top vector. If an infinite-dimensional representation has a top vector of weight π, we can still express π in the form $p_i h^i$ but now the p_i will not be integers so part (ii) of the theorem will not necessarily be true. However, (i) and (iii) will still obtain.

IV. THE LABELING OF STATES

The elements of $\mathscr{H}$ consisting of linear combinations of $\{r^i\}$ or $\{h^i\}$ with non-negative real coefficients constitute a closed convex cone, usually called a *Weyl chamber*, which we shall denote by WC. Here R_i is a reflection in a face of a WC. As remarked above, the n reflections R_i generate a finite group of orthogonal transformations of $\mathscr{H}$, which permute the roots Σ. The vector r^i (or the vector h^i) is a basis for one of the edges of the WC. The simple root r_i is perpendicular to the face of the WC opposite the edge r^i. Choosing the sense of all the r_i to point inwards into the WC established the distinction between the positive Σ^+ and the negative roots Σ^-, $\Sigma = \Sigma^+ \cup \Sigma^-$. Of basic importance is that a WC is a fundamental region for the action of the Weyl group W on $\mathscr{H}$. That is, every vector in $\mathscr{H}$ is equivalent under W to precisely one vector in the WC. Any element of W sends the WC into another simple convex cone, which could have been used to define a different set of simple roots. This justifies our previous remark that the number of possible sets of simple roots equals $|W|$, the order of W.

Any vector *interior* to WC (i.e., $h = u_i h^i$ with $u_i > 0$ for all i) has $|W|$ distinct images under the action of W on $\mathscr{H}$. But, for example h^1, which is not interior to WC, is fixed under $R_2, R_3, \ldots, R_n$ and therefore, under the subgroup W_1, generated by these reflections. The orbit of h^1 will therefore contain $[W{:}W_1] = |W|/|W_1|$ vectors.

It follows from all this that any weight of Γ_π is equivalent to a unique $\mu = m_i h^i$ with all $m_i \geqslant 0$. If W_μ is the stability group of μ then the *equivalence class* of weights for which μ is an unambiguous label contains $[W{:}W_\mu]$ weights. For $w \in W$ let $w\mu$ denote the image of μ by w, then w: $V_\mu \to V_{w\mu}$. Thus if $\{x^i_\mu\}$ is a basis of V_μ, $\{wx^i_\mu\}$ is a basis of $V_{w\mu}$ and we can denote wx^i_μ by $x^i_{w\mu}$. Thus if we had labels for a basis $\{x^i_\mu\}$ of V_μ, bases for the other weight spaces in the orbit of V_μ can be labeled by a set $\{w_\alpha\}$ of representatives of the left cosets of W with respect to W_μ.

However, by comparing (3.2) and (3.3) we see that linearly independent vectors b^B_i such that $C^\mu_{AB} b^B_i \neq 0$, with $1 < i \leqslant m_\mu$, define a basis

$$v^i_\mu = b^B_i F_B x_\pi \tag{4.1}$$

of V_μ.

Clearly if $m_\mu > 1$ there will be great flexibility for the choice of (b^B_i) and at this stage it is too early for the author to confidently propose one or more "canonical" choices since what will prove to be most appropriate will probably appear only after considerable experimentation with this new notation. However, the following observations may prove relevant. Since C^μ is a symmetric matrix, with integer entries, it will have real eigenvalues and m_μ linearly independent eigenvectors with real coefficients. These would give rise to a basis for V_μ and if the eigenvalues of C^μ happened to be distinct, which in general would be probable, each real eigenvalue would provide a label for the corresponding v^i_μ.

Alternatively, row reduction of C^μ_{AB} will quickly lead to a basis of V_μ in which each v^i_μ is a single $F_B x_\pi$ for an appropriate choice of B.

When the representation of G is unitary, $F^\dagger_B = E_B$, and there is a scalar product defined on V so that

$$\begin{aligned}\langle v^i_\mu | v^j_\mu \rangle &= b^A_i b^B_j \langle F_A x_\pi | F_B x_\pi \rangle \\ &= b^A_i b^B_j \langle x_\pi | E_A F_B x_\pi \rangle \\ &= b^A_i b^B_j C^\mu_{AB} \langle x_\pi | x_\pi \rangle \\ \therefore v^i_\mu \perp v^j_\mu &\Leftrightarrow b_i \perp b_j \text{ with respect to } C^\mu.\end{aligned}$$

It will therefore be possible to obtain a basis of mutually *orthogonal* vectors $\{v^i_\mu\}$, with purely rational and therefore integral coefficients (by multiplying by the LCM of the denominator). To *normalize* the basis will, in general, require the extraction of square roots of integers.

Some of these possibilities are illustrated in the Appendices.

V. CONCLUDING REMARKS

Hopefully the reader is convinced that the scheme proposed in this paper provides a solution of the state labeling problem that is *universal* in the sense of being a uniform approach for all simple LG's. It circumvents the necessity of calculating parentage coefficients, which complicates the common solution of labeling by recourse to subgroup chains. In some situations, of course, a subgroup chain can throw significant light on the physical problem.

For the sake of consistency it will be desirable to make canonical choice of the W_μ coset representatives w. The most intelligible and perspicuous choice will depend on the peculiarities of the individual Weyl groups but should not present an insurmountable difficulty since these groups have been studied[15] in exquisite detail by Cartan, Coxeter, and others.

A critical test for the proposed notation will be the ease with which it enables the Wigner–Clebsch–Gordan coefficients to be calculated and described.

The author plans to address these practical questions in subsequent papers along lines suggested in the following Appendices.

ACKNOWLEDGMENT

The author is most appreciative of support of the present research by the Natural Sciences and Engineering Research Council of Canada through Grant No. A-2990.

APPENDIX A: THE MATRIX C^μ

Case 1 ($s_\mu = 1$): This occurs when there is only one path from π to μ, which happens if and only if $\pi - \mu = kr_i$. Set $C^\mu_{AA} = C_k$. We easily see that if $E_i F^k_i x_\pi = a_k F^{k-1}_i x_\pi$, then $a_k = p_i - 2(k-1) + a_{k-1}$ and $a_1 = p_i$ whence, by induction,

$$C_k = (k!)^2 \binom{p_i}{k}, \tag{A1}$$

since $C_k = \Pi^k_{j=1} a_j$. Thus the longest straight r_i-chain starting from x_π has length p_i. In particular, if $p_i = 0$, all paths starting with r_i give rise to zero elements of C^μ.

Case 2 ($s_\mu = 2$): This occurs if and only if $\pi - \mu = r_i + r_j$ with $i \neq j$. With no loss of generality, assume $i = 1$, $j = 2$:

$$\begin{aligned}C^\mu_{12,12} x_\pi &= E_2 E_1 F_1 F_2 x_\pi \\ &= (\pi - r_2) \cdot h_1 E_2 F_2 x_\pi = (p_1 - a_{12}) p_2 x_\pi.\end{aligned}$$

Recall that the *Cartan matrix* of a simple LA (a_{ij}) is defined by

$$a_{ij} = h_i \cdot r_j = (2r_i \cdot r_j)/(r_i \cdot r_i). \tag{A2}$$

As before, $C^{\mu}_{12,21} = C^{\mu}_{21,12} = p_1 p_2$. Hence C^{μ} is given by

$$\begin{array}{cc} & \begin{array}{cc} (12) & (21) \end{array} \\ \begin{array}{c} (12) \\ (21) \end{array} & \begin{bmatrix} (p_1 - a_{12})p_2 & p_1 p_2 \\ p_1 p_2 & (p_2 - a_{21})p_1 \end{bmatrix} \end{array}. \tag{A3}$$

Since the Cartan matrices for A_2,B_2,G_2, are, respectively,

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix},$$

the C^{μ} for arbitrary π when $\pi - \mu = r_1 + r_2$ for A_2,B_2,G_2, are, respectively,

$$\begin{bmatrix} (p_1+1)p_2 & p_1 p_2 \\ p_1 p_2 & (p_2+1)p_1 \end{bmatrix}, \quad \begin{bmatrix} (p_1+1)p_2 & p_1 p_2 \\ p_1 p_2 & (p_2+2)p_1 \end{bmatrix}, \quad \begin{bmatrix} (p_1+1)p_2 & p_1 p_2 \\ p_1 p_2 & (p_2+3)p_1 \end{bmatrix}. \tag{A4}$$

We recall that the dimension of the irreducible representation (p_1,p_2) for A_2 is

$$d_{\pi} = \tfrac{1}{2}(p_1+1)(p_2+1)(p_1+p_2+2), \tag{A5}$$

for B_2 is

$$d_{\pi} = \tfrac{1}{6}(p_1+1)(p_2+1)(p_1+p_2+2)(2p_1+p_2+3), \tag{A6}$$

and for G_2 is

$$d_{\pi} = \tfrac{1}{120}(p_1+1)(p_2+1)(p_1+p_2+2) \times (p_1+2p_2+3)(p_1+3p_2+4)(2p_1+3p_2+5). \tag{A7}$$

These result immediately from Weyl's formula (2.12).

We note that for A_2, both (1,0) and (0,1) give representations of dimension 3 corresponding to SU(3) and to the representation SU(3) $\wedge$ SU(3) on antisymmetric bivectors. The adjoint representation of dimension 8 is given by $\pi = (1,1)$.

For B_2, $d_{(1,0)} = 5$ is the defining representation SO(5), whereas $d_{(0,2)} = 10$ is the adjoint representation.

For G_2, $d_{(1,0)} = 7$, the lowest-dimensional representation of G_2 and $d_{(0,1)} = 14$ is the adjoint representation.

Remark: In case 2 with $s_{\mu} = 2$, as presented above, we assumed that G had rank 2. In fact that was not an essential assumption. If the rank of G is 3 or more the calculation would be exactly the same; however, the G_2 type could not occur. We would be lead to a C^{μ} identical to that for A_2 or B_2 above.

Case 3 ($s_{\mu} = 3$): This occurs only when $\pi - \mu = 2r_i + r_j$, $i \neq j$. Again we assume that $i = 1$ and $j = 2$, without loss of generality. There are three paths from μ to π, (112), (121), and (211), which we have set down in lexicographical or numerical order, assuming that 1 precedes 2, 2 precedes 3, etc. Thus the clumsy symbol $C^{\mu}_{112,121}$ can now be abbreviated to $C_{1,2}$. We have also dropped the superscript μ since for a given irreducible $\pi = (p_i)$, the path by itself determines μ. The $C_{i,j}$, $1 \leqslant i,j \leqslant 3$, can be calculated as before. We give details for $C_{1,2}$ in order to point out explicitly that the C matrices can be calculated by recursion on k:

$$C_{1,2} x_{\pi} = E_2 E_1 E_1 F_1 F_2 F_1 x_{\pi} \tag{A8}$$

$$\begin{aligned} &= (E_2 E_1 H_1 F_2 F_1 + E_2 F_1 F_1 E_1 F_2 F_1) x_{\pi} \\ &= (\pi - r_1 - r_2) \cdot h_1 E_2 E_1 F_2 F_1 x_{\pi} \\ &\quad + \pi \cdot h_1 E_2 E_1 F_1 F_2 x_{\pi} \end{aligned} \tag{A9}$$

$$= (p_1 - 2 - a_{12}) C_{12,21} x_{\pi} + p_1 C_{12,12} x_{\pi}, \tag{A10}$$

where $C_{12,21}$ and $C_{12,12}$ are the coefficients that were calculated in case 2.

Equations (A8)–(A10) illustrate the basic method of evaluating C_{AB}. One starts by moving the innermost E_i through the F's by means of the commutation relation $[E_i,F_j] = H_i \delta_{ij}$. Thus in (A8), for which $k = 3$, we moved E_1 to the right, leading to (A9), which involves a sum of terms for which $k = 2$. We substitute, collect terms, and obtain

$$C_{1,2} = 2p_1 p_2 (p_1 - 1 - a_{12}). \tag{A11}$$

Thus there is a simple algorithm that easily can be programmed for a computer to churn out expressions for the $C_{i,j}$ in terms of the integers (p_j) defining the top weight and the Cartan matrix (a_{ij}), which specifies the group under consideration.

With $A_1 = (112)$, $A_2 = (121)$, and $A_3 = (211)$ and setting $a_{21} = -a$, $a_{12} = -b$, we find that $C_{i,j}$ is

$$\begin{bmatrix} 2p_2(p_1+b)(p_1+b-1) & 2p_1 p_2(p_1+b-1) & 2p_1 p_2(p_1-1) \\ 2p_1 p_2(p_1+b-1) & p_1[2p_1 p_2 + bp_1 - (2-b)(p_2+a)] & 2p_1(p_1-1)(p_2+a) \\ 2p_1 p_2(p_1-1) & 2p_1(p_1-1)(p_2+a) & 2p_1(p_1-1)(p_2+2a) \end{bmatrix}. \tag{A12}$$

Of the two numbers a and b, at least one of them is 1 or, if $r_1 \perp r_2$, they are both zero. By a suitable ordering of the simple roots we can arrange that $b = 1$, $a \neq 0$, or $a = b = 0$.

If $a = b = 0$, the above matrix becomes

$$2p_1 p_2 (p_1 - 1) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{A13}$$

of rank 1 if and only if $p_1 p_2 (p_1 - 1) \neq 0$.

With $b=1$, (A12) becomes

$$\begin{bmatrix} 2p_1p_2(p_1+1) & 2p_1^2p_2 & 2p_1p_2(p_1-1) \\ 2p_1^2p_2 & p_1[2p_1p_2-p_2+a(p_1-1)] & 2p_1(p_1-1)(p_2+a) \\ 2p_1p_2(p_1-1) & 2p_1(p_1-1)(p_2+a) & 2p_1(p_1-1)(p_2+2a) \end{bmatrix}. \tag{A14}$$

When $p_1=0$, both (A12) and (A14) vanish so $m_\mu=0$, whereas if $p_1=1$ the third row and third column vanish and

$$\begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix} = p_2 \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \tag{A15}$$

of rank 1 or 0 according as $p_2 \neq 0$ or $p_2=0$.

APPENDIX B: ORTHONORMAL BASES

To illustrate the ideas of this paper we use the 15-dimensional representation $\pi=(p_1,p_2)=(2,1)$ of A_2 or SU(3). This representation should not be confused with the inequivalent representations (1,2) and (4,0), which, according to (A5), also have dimension 15.

Using the Cartan matrix of A_2 in (A4) we find that

$$h^1=\tfrac{2}{3}r_1+\tfrac{1}{3}r_2, \quad h^2=\tfrac{1}{3}r_1+\tfrac{2}{3}r_2, \tag{B1}$$

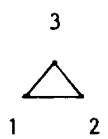
and that

$$r_1=2h^1-h^2, \quad r_2=-h^1+2h^2. \tag{B2}$$

Formulas (B2) illustrate the fact that in the $\{h^i\}$ basis all weights have integral coefficients.

For $\pi=(2,1)$, the only *dominant* weights (i.e., those with $p_j \geqslant 0$) that occur in the representation are $(2,1)=\pi$, $(0,2)=\pi-r_1$, and $(1,0)=\pi-r_1-r_2$.

For A_n the Weyl group is isomorphic to S_{n+1}, the symmetric group on $n+1$ objects. Just as S_3 is generated by two transpositions (12) and (23) whose product $(1\,2)(2\,3)=(1\,2\,3)$ has order 3, so W for A_2 is generated by two reflections R_1 and R_2 such that $(R_1R_2)^3=I$. In the extended CD diagram



,

the nodes 1 and 2 denote the reflections R_1 and R_2 while the node 3 corresponds to $R_3=R_1R_2R_3=R_2R_1R_2$, which is also a reflection such that $(R_1R_3)^3=(R_2R_3)^3=I$. The mapping from W into S_3 defined by $R_1\to(12)$, $R_2\to(23)$ sends $R_3\to(13)$, $R_1R_2\to(123)$, and $R_2R_1\to(132)$, where, in a common notation, (123) denotes the cycle of period 3 which replaces 1 by 2, 2 by 3, and 3 by 1. To simplify our notation, set $w_0=I$, $w_1=R_1$, $w_2=R_2$, $w_3=R_3$, $w_4=R_1R_2$, and $w_5=R_2R_1$. Of course, R_1 and R_2 denote the previously defined reflections along r_1 and r_2, respectively.

Since r_i is perpendicular to h^j if $j\neq i$, $R_ih^j\neq h^j$ only for $j=i$ and, with no summation on i,

$$R_ih^i=h^i-2[(h^i\cdot r_i)/(r_i\cdot r_i)]r_i=h^i-r_i.$$

It follows from (B2) that for A_2

$$\begin{aligned} &w_0: h^1\to h^1, \quad h^2\to h^2, \\ &w_1: h^1\to -h^1+h^2, \quad h^2\to h^2, \\ &w_2: h^1\to h^1, \quad h^2\to h^1-h^2, \\ &w_3: h^1\to -h^2, \quad h^2\to -h^1, \\ &w_4: h^1\to -h^1+h^2, \quad h^2\to -h^1, \\ &w_5: h^1\to -h^2, \quad h^2\to h^1-h^2. \end{aligned} \tag{B3}$$

Since $\pi=(2,1)=2h^1+h^2$ does not lie in a face of the Weyl chamber its stability group consists of the identity only. Therefore its orbit under W consists of *six* elements, which by (B3) are

$$\begin{aligned} &(2,1),\ (-2,3),\ (3,-1), \\ &(-1,-2),\ (-3,2),\ (1,-3). \end{aligned} \tag{B4}$$

On the other hand, $\pi-r_1=(0,2)$ is fixed by w_0 and w_1 and hence its orbit has only three elements:

$$(0,2),\ (2,-2),\ (-2,0). \tag{B5}$$

Similarly, $\pi-r_1-r_2=(1,0)$ is fixed by w_0 and w_2 and its orbit consists of

$$(1,0),\ (-1,1),\ (0,-1). \tag{B6}$$

For $\mu=(1,0)$, it follows from (A4) that

$$C^\mu=\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}, \tag{B7}$$

which has rank 2. Thus each of the weights (B6) has multiplicity 2. Since $6+3+2\times 3=15$, (B4)–(B6) account for the 15 dimensions of the (2,1) representation of A_2 with $m_{21}=m_{02}=1$ and $m_{10}=2$.

It follows from (B7) that $u_1=F_1F_2x_\pi$ and $u_2=F_2F_1x_\pi$ are linearly independent and span V_{10}. From (B7) we immediately read off $\langle u_1|u_1\rangle=3$, $\langle u_2|u_2\rangle=4$, $\langle u_1|u_2\rangle=2$. Thus u_1 and u_2 are not orthogonal. For some purposes they may form the most convenient basis; however, it would be possible in a continuous infinity of ways to obtain an equivalent orthonormal pair of vectors.

For example, in the present case, u_2 and $2u_1-u_2$ are orthogonal. Defining $2v_{10}^1=u_2$ and $2\sqrt{2}\,v_{10}^2=2u_1-u_2$ it follows that $\{v_{10}^i\}$ is an orthonormal basis of V_{10}. Then if $v_{21}=x_\pi$ and $\sqrt{2}v_{02}=F_1x_\pi$, a complete orthonormal basis for the representation space V is

$$wv_{21},\ \tilde{w}v_{02},\ \tilde{w}v_{10}^i, \tag{B8}$$

where

$$w \in W \quad \text{and} \quad \tilde{w} \in \{w_0, w_4, w_5\}. \tag{B9}$$

It is perhaps worth noting that the representation theory of S_k puts limitations on possible values of m_μ. Denote the n-vector (k^i) by $\underset{\sim}{k}$ and, as on p. 36 of the author's treatment[16] of the representations of S_k, denote by $[\underset{\sim}{k};\uparrow]$ the representation of S_k induced from the identity representation of ΠS_{k^i}. This representation of dimension s_μ is easily reduced into its irreducible constituents by means of the Littlewood–Richardson rule. Thus since, for example,

$$aa \times bb \uparrow S_4 = \begin{matrix} aabb + & aab + & aa \\ & b & bb \end{matrix}$$

so

$$[2,2;\uparrow] = [4] + [3,1] + [2^2] \tag{B10}$$

with dimensions

$$6 = 1 + 3 + 2.$$

In the representation of A_2 discussed in this appendix, $\mu = (0,-1) = (2,1) - 2r_1 - 2r_2$ has $m_\mu = 2$ and $s_\mu = 6$. A possible path from π to μ is (1212). We therefore expect the six $F_A x_\pi$ to span the representation $[2,2;\uparrow]$ of S_4, so C^μ should be equivalent to a diagonal block matrix with blocks of degree 1, 3, and 2. Since $m_\mu = 2$, we can conclude that the first two blocks are identically zero.

When one of the irreducibles of $[\underset{\sim}{k};\uparrow]$ is actually present possibly Young's method of defining an orthogonal irreducible representation of S_k can be invoked to help us find an orthogonal basis for V_μ.

The Young diagrams of the irreducible representations of S_k that arise can never have more than n rows. Possibly they will play some role, at least for SU(n), in any attempt to correlate the notation proposed in the present paper with the GZ patterns.

APPENDIX C: MATRICES FOR THE GENERATORS OF THE LIE ALGEBRA

In this appendix in order to illustrate the power of our notation we give the matrix for E_1 in the representation $(2,1)$ discussed in Appendix B. Our techniques for obtaining the matrix elements are still somewhat *ad hoc* so they will only be briefly indicated since we expect to improve them and give a full presentation in subsequent papers.

A simple LA is generated by $2n$ elements e_i and f_i such that $e_i \circ f_i = h_i \in \mathscr{H}$ with e_i, $1 \leqslant i \leqslant n$, corresponding to a set of simple roots. Under a representation Γ such that $\Gamma(e_i) = E_i$, etc., $[E_i, F_i] = H_i$. It is sometimes helpful to use the following notations: (i) $f_\alpha = e_{-\alpha} = e_{\bar{\alpha}}$, when the root $\alpha > 0$; and (ii) if, for example $\alpha = 2r_1 + 3r_2$, we could denote e_α by e_{23}, e_1 by e_{10}, and e_2 by e_{01}. Since $e_1 \circ e_2$ has weight $r_1 + r_2$ it is possible to define e_{11} by $e_{11} = e_1 \circ e_2$, which would map into $\Gamma(e_{11}) = E_{11} = \Gamma(e_1 \circ e_2) = [\Gamma(e_1), \Gamma(e_2)] = [E_1, E_2]$.

In a unitary representation, $F_i = E_i^\dagger$, so in order to completely specify Γ we need know only the matrices for E_i, $1 \leqslant i \leqslant n$. Thus a unitary representation of A_2 is completely determined by E_1 and E_2.

The Weyl group W plays a key role in the approach of the present paper. This is a finite group acting on $\mathscr{H}$ obtained by restriction to the tangent plane of T at the identity of the group of inner automorphisms of G that leave the maximal toroid T invariant. Thus any $w \in W$ is covered by a $\hat{w} \in G$ such that $\hat{w} T \hat{w}^{-1} = T$. Since T is Abelian, for any element $t \in T$, $\hat{w} t$ will have the same action on T and $\mathscr{H}$ as $\hat{w}$. In fact, $\{\hat{w} t \mid t \in T\}$ is precisely the set of elements of G which "cover" w. The generators e_i and f_i of L are eigenvectors for the action of T on L so that $\hat{w} t$ sends e_α to $\hat{w} e_\alpha \hat{w}^{-1}$ multiplied by a phase factor. This may be the source of the famous "phase problem" in quantum mechanics.

For a reflection $R_i \in W$ a group element $\hat{R}_i$ will have the property $\hat{R}_i^2 \in T$. The n degrees of freedom in the choice of t in $\hat{R}_i t$ can be used to ensure[17] that $\hat{R}_i^4$ is the identity element of G. We shall always assume that this has been done.

If the representation Γ maps $\hat{w}$ onto $\widehat{W}$ we shall define $w(H)$ to mean

$$\Gamma(w(h)) = \widehat{W} H \widehat{W}^{-1}. \tag{C1}$$

So if $x_\mu \in V_\mu$

$$\begin{aligned} H \widehat{W} x_\mu &= \widehat{W} \widehat{W}^{-1} H \widehat{W} x_\mu = \widehat{W} w^{-1}(H) x_\mu \\ &= (w^{-1}(h) \cdot \mu) \widehat{W} x_\mu = (h \cdot w\mu) \widehat{W} x_\mu. \end{aligned} \tag{C2}$$

Thus

$$\widehat{W} x_\mu \in V_{w\mu}. \tag{C3}$$

Once a particular cover for w has been chosen, an action by the Weyl group on G and on the representation space V is defined so we can dispense with the notations $\hat{w}$ and $\widehat{W}$. Thus for $x \in V$ we shall interpret wx as $\widehat{W} x$.

Applying all this to the representation $(2,1)$ of A_2 we can define the action of W on the generators of L and on the weight spaces V_μ of Γ by means of the following table in which a column gives the image of the first column under action by the indicated element of the Weyl group:

w_0	w_1	w_2	w_3	w_4	w_5
e_{10}	f_{10}	e_{11}	f_{01}	e_{01}	f_{11}
e_{01}	e_{11}	f_{01}	f_{10}	f_{01}	e_{10}
e_{11}	e_{01}	e_{10}	f_{11}	f_{10}	f_{01}
21	$\bar{2}3$	$3\bar{1}$	$\bar{1}2$	$\bar{3}2$	$1\bar{3}$
02	02	$2\bar{2}$	$\bar{2}0$	$\bar{2}0$	$2\bar{2}$
10	$\bar{1}1$	10	$0\bar{1}$	$\bar{1}1$	$0\bar{1}$

(C4)

The meaning of the last three rows is conveyed by the example $w_3 V_{21} = V_{\bar{1}2}$.

For the carrier space V of the $(2,1)$ representation of A_2 we take a basis consisting of 15 vectors v_i, obtained from the unit top vector $x_\pi = x_{21}$ as follows. We are using the previously defined notation w_i for the elements of W.

The first *six*, in the orbit of V_{21}, are

$$v_i = w_i x_{21}, \quad \text{so} \quad v_0 = x_{21}, \quad 0 \leqslant i \leqslant 5. \tag{C5}$$

The next *three*, in the orbit of V_{02}, are defined by

$$\sqrt{2} v_6 = F_1 v_0, \quad v_7 = w_2 v_6, \quad v_8 = w_3 v_6. \tag{C6}$$

The final *six* in the orbit of V_{10} are defined by

$$\sqrt{3} v_9 = u_1 = F_1 F_2 v_0, \quad v_{10} = w_1 v_9, \quad v_{11} = w_3 v_9, \tag{C7}$$

$$2v_{12} = u_2 = F_2F_1v_0, \quad v_{13} = w_1v_{12}, \quad v_{14} = w_3v_{12}. \quad \text{(C8)}$$

These v_i are all of unit length and the only failure of orthogonality is between the three pairs $v_9 - v_{12}$, $v_{10} - v_{13}$, and $v_{11} - v_{14}$.

In this basis, with $\epsilon_i \in \{\pm 1, \pm i\}$ two phase factors, the action of E_1 is defined as follows:

$$\begin{aligned}
&E_1v_0 = 0, \quad E_1v_1 = \epsilon_1\sqrt{2}v_6, \quad E_1v_2 = 0,\\
&E_1v_3 = \epsilon_2v_5, \quad E_1v_4 = \epsilon_2\sqrt{3}v_{10}, \quad E_1v_5 = 0,\\
&E_1v_6 = \sqrt{2}v_0, \quad E_1v_7 = 0, \quad E_1v_8 = \sqrt{2}v_{14},\\
&E_1v_9 = \epsilon_2\sqrt{3}v_2, \quad E_1v_{10} = \epsilon_12v_9,\\
&E_1v_{11} = -\sqrt{\tfrac{2}{3}}\epsilon_1\epsilon_2v_7, \quad E_1v_{12} = \epsilon_2v_2,\\
&E_1v_{13} = \epsilon_1(\sqrt{3}v_9 - v_{12}), \quad E_1v_{14} = -\sqrt{2}\epsilon_1\epsilon_2v_7.
\end{aligned} \quad \text{(C9)}$$

Since $w_1V_{02} = V_{02}$ is one dimensional and is spanned by v_6 it follows that $w_1v_6 = \epsilon_1v_6$ with $\epsilon_1 = \pm 1$ or $\pm i$. Using (C4) and the multiplication table for S_3, we easily conclude

$$w_1v_6 = \epsilon_1v_6, \quad w_5v_6 = \epsilon_1v_7, \quad w_4v_6 = \epsilon_1v_8. \quad \text{(C10)}$$

We also note that $F_2v_0 \in F_{3\bar{1}}$, which is spanned by v_2. Therefore $F_2v_0 = av_2$. But

$$\begin{aligned}
\langle F_2v_0|F_2v_0\rangle &= \langle v_0|E_2F_2v_0\rangle\\
&= \langle v_0|(H_2 + F_2E_2)v_0\rangle = \langle v_0|v_0\rangle = 1.
\end{aligned}$$

Thus $|a|^2 = 1$. Set $a = \epsilon_2$ and we have

$$F_2v_0 = \epsilon_2v_2, \quad v_2 = \bar{\epsilon}_2F_2v_0, \quad \text{(C11)}$$

with $|\epsilon_2| = 1$. In fact we shall see below that $\epsilon_2^4 = 1$. The only indetermination in (C9) arises from the two ϵ_i. So there are left sixteen alternative choices for phase factors for E_1, presumably associated with distinct but equivalent representations. We effectively narrowed the choice of phase factors when we decided on the particular action of W on L defined by the table (C4).

We now exhibit some typical calculations to illustrate how the formulas (C9) were obtained. We shall make constant use of the table (C4) giving the action of W on e_α for $\alpha > 0$, and the corresponding table that can be deduced immediately from (C4) for f_α, $\alpha > 0$. For example, $E_1w_4 = w_4F_{11}$, since $E_1w_4 = w_4w_4^{-1}E_1w_4 = w_4w_5(E_1) = w_4F_{11}$. Here we used the fact that $w_5w_4 = w_0$ and that by (C4), $w_5(e_1) = f_{11} = f_1 \circ f_2$.

The zeros in (C9) for E_1v_i when $i \in \{0,2,5,7\}$ result from the fact that for these i, E_1v_i has a weight that does not occur in the irreducible representation (2,1).

Take another example. By (C6), $\sqrt{2}v_6 = F_1v_0$. Now $E_1F_1v_0 = (H_1 + F_1E_1)v_0 = (h_1 \cdot \pi)v_0 = 2v_0$. Thus

$$E_1v_6 = \sqrt{2}v_0. \quad \text{(C12)}$$

Further,

$$\begin{aligned}
\sqrt{2}E_1v_8 &= E_1w_3F_1v_0 = w_3F_2F_1v_0\\
&= w_3u_2 = 2w_3v_{12} = 2v_{14}.
\end{aligned}$$

Thus

$$E_1v_8 = \sqrt{2}v_{14}. \quad \text{(C13)}$$

In order to exhibit the action of E_1, or indeed of any generator, on wV_{10} for arbitrary w, it is helpful—probably essential—to first determine the action of w_2 on V_{10}. Recall that V_{10} is spanned by u_1 and u_2. As follows from (C4), w_2 generates the stability group of V_{10}. Since $E_2F_2v_0 = H_2v_0 = v_0$, by (C11) we have

$$\begin{aligned}
w_2u_1 &= w_2F_1F_2v_0 = F_{11}E_2w_2v_0 = \bar{\epsilon}_2F_{11}E_2F_2v_0\\
&= \bar{\epsilon}_2(F_1F_2 - F_2F_1)v_0 = \bar{\epsilon}_2(u_1 - u_2).
\end{aligned}$$

Therefore

$$w_2u_1 = \bar{\epsilon}_2(u_1 - u_2). \quad \text{(C14)}$$

Similarly, we can show that

$$w_2u_2 = -\bar{\epsilon}_2u_2. \quad \text{(C15)}$$

If we use (C14) and (C15) to apply w_2 twice we find that

$$w_2^2u_1 = \bar{\epsilon}_2^2u_1, \quad w_2^2u_2 = \bar{\epsilon}_2^2u_2.$$

Since w_2^4 must be the identity, it follows, as previously noted, that

$$\epsilon_2 = \pm 1 \text{ or } \pm i. \quad \text{(C16)}$$

As a final example consider E_1v_{11}. We first note that $E_1v_{11} \in V_{2\bar{2}}$ and is therefore a multiple of the unit vector v_7. Noting (C10) and (C14), consider

$$\begin{aligned}
&\langle E_1w_3u_1|v_7\rangle\\
&= \langle w_3F_2u_1|w_2v_6\rangle = \langle F_2u_1|w_5v_6\rangle\\
&= \epsilon_1\langle u_1|E_2w_2v_6\rangle = (1/\sqrt{2})\epsilon_1\langle w_2u_1|F_2F_1v_0\rangle\\
&= (1\sqrt{2})\epsilon_1\epsilon_2\langle u_1 - u_2|u_2\rangle = (1/\sqrt{2})\epsilon_1\epsilon_2(2-4)\\
&= -\sqrt{2}\epsilon_1\epsilon_2.
\end{aligned}$$

Thus, since $w_3u_1 = \sqrt{3}v_{11}$,

$$E_1v_{11} = -\sqrt{\tfrac{2}{3}}\epsilon_1\epsilon_2v_7. \quad \text{(C17)}$$

In the above discussion it was essential to know the action of the stability group $\{w_0, w_1\}$ on V_{02} and $\{w_0, w_2\}$ on V_{10}. In general, if $m_\mu > 1$, Σ_μ, the stability group of V_μ, will be nontrivial. Preparatory to finding the matrix of any generator it will be necessary to explicitly describe the action of Σ_μ on V_μ.

[1] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).

[2] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[3] M. D. Gould, J. Math. Phys. **22**, 2376 (1981).

[4] E. Cartan, Bull. Soc. Math. France **41**, 53 (1913); *Collected Works* (Gauthier–Villars, Paris, 1952), Vol. I, p. 355.

[5] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1963); D. P. Zhelobenko, *Compact Lie Groups and their Representations* (Am. Math. Soc. Translations, Providence, RI, 1973); Dynkin, in Ref. 13; Witt, in Ref. 11; R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).

[6] W. Killing, Math. Ann. **31**, 252 (1888); **33**, 1 (1889); **34**, 57 (1889); **36**, 161 (1890).

[7] H. Weyl, *The Classical Groups—Their Invariants and Representations* (Princeton U.P., Princeton, NJ, 1938).

[8] The symbol $[x,y]$ is often used for the Lie product $x \circ y$ but we prefer to reserve the square brackets for the commutator of two matrices. Thus if γ: $x \to X$ is a representation of L by matrices or linear operators, $\gamma(x \circ y) = [\gamma(x), \gamma(y)] = [X, Y] = XY - YX$.

[9] E. Cartan, Ann. Mat. **4**, 209 (1927).

[10] H. S. M. Coxeter, J. London Math. Soc. **6**, 132 (1931).

[11] E. Witt, Ahh. Math. Sem. Univ. Hamburg **14**, 289 (1941).

[12]H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).
[13]E. B. Dynkin, *The Structure of Semi-Simple Algebras*, Usp. Mat. Nauk **2**, 59 (1947); Am. Math. Soc. Trans. Ser. 1 **9**, 328 (1962).
[14]H. Weyl. Math. Zeit. **23**, 271 (1925).
[15]E. Cartan, Am. J. Math. **18**, 1 (1896); H. S. M. Coxeter, *Regular Polytopes* (Methuen, London, 1948); Witt. see Ref. 11; N. Bourbaki, *Groupes et Algèbres de Lie* (Hermann, Paris, 1968), Chap. 6.
[16]A. J. Coleman, *Induced Representations with Applications to S_n and $GL(n)$, Queen's Papers in Pure and Applied Mathematics*, Vol. 4 (Campus Bookstore, Queen's University, 1966), cf. Sec. 18.
[17]R. V. Moody and J. Patera, J. Math. Phys. **25**, 2838 (1984).

Multiplicity-free Wigner coefficients for semisimple Lie groups. I. The U(n) pattern calculus

M. D. Gould
School of Chemistry, University of Western Australia, Nedlands, Western Australia 6009

(Received 14 November 1984; accepted for publication 16 April 1986)

This is the first paper in a series of two dedicated to a new algebraic determination of the multiplicity-free reduced Wigner coefficients for the Lie groups U(n) and O(n). The approach employed enables a direct (nonrecursive) derivation of reduced Wigner coefficients. The absolute value squared of the reduced Wigner coefficients is expressed as a rational polynomial function (numerator polynomial divided by denominator polynomial) whose Weyl group symmetries are seen to fall out naturally in our approach from the transformation properties of polynomial functions determined by Casimir invariants. A unified treatment of the composition laws of reduced Wigner coefficients and the evaluation of their associated denominator polynomials is presented, which applies to both U(n) and O(n). An explicit formula for the numerator polynomials of U(n) is also derived. The numerator polynomials for the orthogonal groups will be given in the second paper of the series.

I. INTRODUCTION

In the literature two approaches to a general study of the classical groups have emerged. First there is the algebraic infinitesimal approach, which exploits only the generators and their commutation relations. This approach has its origin in the pioneering researches of Casimir,[1] Casimir and Van der Waerden,[2] and Racah.[3] Second, there is the integral approach as expounded in the classic works of Weyl.[4] The methods of Weyl have proved a powerful tool in group theoretical applications to physics and have been applied, in conjunction with the Jordan–Schwinger[5] boson calculus, by several authors.[6]

From the point of view of physical applications the principal problems to be solved are the complete determination of the states of an irreducible representation and the explicit determination of Wigner (or Clebsch–Gordan) coefficients. A major step in this direction was made in 1950 by Gel'fand and Tsetlin,[7] who constructed, with a full set of labels, a complete set of basis vectors for the irreducible representations of the unitary and orthogonal groups. The matrix elements of the group generators were also given initially by Gel'fand and Tsetlin[7] and later by Baird and Biedenharn,[8] who made an important contribution by providing a detailed proof, based on Weyl's well-known branching law,[4] of the Gel'fand–Tsetlin results. The results of Ref. 8 are also of interest because they reveal the structure of the matrix elements (i.e., a product of a reduced matrix element and a Wigner coefficient). The fundamental Wigner coefficients (and reduced Wigner coefficients) for the Lie group U(n) were thus given for the first time. This work has recently been extended to the orthogonal group[9] using different methods (see also Refs. 10 and 11). The more general problem of evaluating all multiplicity-free Wigner coefficients has been solved in the case of U(n) by Biedenharn and coworkers.[12–14] Although much work has been done on O(n) by several authors,[15,16] the complete program followed by Biedenharn *et al.*[12–14] for U(n) has never been carried out for O(n).

This paper is the first in a series of two in which we present a new derivation of the multiplicity-free Wigner coefficients for both the unitary and orthogonal groups. This work grew out of a previous series of papers[9–11] in which it was demonstrated that the fundamental U(n): U($n-1$) [resp. O(n): O($n-1$)] reduced Wigner coefficients (RWC's) may be obtained from the eigenvalues of certain U($n-1$) [resp. O($n-1$)] Casimir invariants. In this series of papers we extend these results to give an algebraic determination of the multiplicity-free U(n): U($n-1$) and O(n): O($n-1$) RWC's. The (multiplicity-free) squared RWC's are determined in our work as eigenvalues of certain Casimir invariants and hence must determine (rational) polynomial functions in the representation labels of the group U(n) [resp. O(n)] and its subgroup U($n-1$) [resp. O($n-1$)]. The Weyl group symmetries of these RWC's are evident from the outset in our approach from the simple transformation properties of polynomial functions.

Although our approach is intimately related to that of Biedenharn *et al.*[12–14] on U(n) there are some essential differences. The pattern calculus of these latter authors, which is inherently integral in nature, affords a recursive determination of the (multiplicity-free) RWC's whereby a general (multiplicity-free) RWC may be composed from certain elementary RWC's. This building up method of constructing RWC's by manipulating and multiplying elementary RWC's (hence the term pattern calculus) has the advantage that it may be extended, at least in principle, to the general case of evaluating all RWC's for the unitary groups (indeed the pattern calculus was designed with this application in mind). In general this latter problem involves a multiplicity problem [in both the U(n) and U($n-1$) groups], which are well known to be difficult to handle, and has been made the subject of a comprehensive series of articles by Biedenharn, Louck, and collaborators.[17]

By contrast the algebraic methods of this paper allow a direct (i.e., nonrecursive) determination of the (multiplicity-free) RWC's. Moreover our approach applies to more general groups and in particular enables a treatment of the

 0022-2488/86/081944-20$02.50

orthogonal group in exact analogy with the unitary group. However, we do not consider here the problem of phases, but these have been obtained in the case of $U(n)$ by Biedenharn *et al.*[12] and a beginning of a phase calculus for $O(n)$ has been given in Ref. 9, where suitable phases for the fundamental (i.e., vector) RWC's for $O(n)$ were given. Also we do not consider the tensor product multiplicity problem at all in this series of papers and our approach seems unlikely to yield any information in this direction (for a new, projection-based, approach to the tensor operator problem see Ref. 18). Nevertheless our approach does extend to the problem of evaluating squared-multiplicity-averaged $U(n)$: $U(n-1)$ [resp. $O(n):O(n-1)$] RWC's, which are given as the sum (over tensor product multiplicities) of the squares of the RWC's. Moreover our approach extends to more general (noncanonical) subgroup imbeddings $G \supset G_0$ to yield multiplicity-averaged G: G_0 RWC's which are multiplicity averaged with respect to the multiplicity labels of the (noncanonical) subgroup G_0. In particular our methods enable an extension to the symplectic group to yield multiplicity-averaged $sp(2n)$: $sp(2n-2) \times sp(2)$ RWC's. We remark that our approach to evaluating (multiplicity-averaged) G: G_0 RWC's is directly related to the problem of evaluating multiplicity-averaged eigenvalues for certain G_0-invariants constructed in the universal enveloping algebra of G. Thus our methods enable, in principle, a systematic determination of the multiplicity-averaged eigenvalues of missing labeling invariants for the $G \supset G_0$ state labeling problem. The above-mentioned problems are currently under investigation. It is felt that such considerations are likely to have physical significance, particularly for statistical treatments of systems admitting hierarchies of symmetries (cf. Ref. 19).

In this paper we restrict ourselves primarily to the unitary groups and leave it to the second paper of the series[20] to determine the multiplicity-free RWC's of the orthogonal group. Since our approach is algebraic and based on eigenvalues of Casimir invariants the emphasis in our work is on (rational) polynomial functions rather than the Wigner operators of Refs. 12–14 and 17: we shall obtain our squared RWC's as a rational polynomial function (numerator polynomial divided by denominator polynomial). We shall present a direct algebraic determination of the (phase-free) $U(n):U(n-1)$ RWC's of Refs. 12, which includes all multiplicity-free RWC's except for certain RWC's for the symmetric tensor representations, which have already been obtained in Refs. 14 and 21 using the pattern calculus. Nevertheless it can be shown that our methods may also be extended to this latter class of RWC's and will be dealt with in a subsequent publication.

From the point of view of physical applications we remark that the unitary group has been applied in elementary particle physics[22,23] and has proved to be an invaluable tool for handling many-body problems in nuclear and molecular physics.[22–29] More recently the unitary group has proved indispensable in the unitary calculus approach to many electron problems[24–29] enabling large-scale configuration interaction calculations to be performed on molecules[26,29] that would be intractable using other methods. There is currently a need, in this latter application, to obtain a matrix element calculus for determining matrix elements of products of one, two, three, and four generators appropriate to the calculation of one-, two-, three-, and four-body operators, respectively. Clearly the pattern calculus for $U(n)$ would be of invaluable assistance in treating this problem. With regard to the orthogonal group, there is evidence to suggest[22] the use of a pattern calculus based on $O(n)$, rather than $U(n)$, particularly for treating short-range interactions in nuclear and atomic physics (i.e., interactions primarily described by an ordinary pairing force).

The paper is set up as follows. In Sec. II we establish our notation and basic conventions and in particular we consider the transformation properties (under the Weyl group) of polynomial functions that will be later applied in determining the symmetries of RWC's. (We remark that our notation is developed with more general subgroup imbeddings in mind.) In Sec. III we show how a large class of RWC's for $U(n)$ and $O(n)$ may be obtained by considering eigenvalues of certain Casimir invariants. In Sec. IV we apply these results to determine the Weyl group symmetries and composition laws of RWC's for $O(n)$ and $U(n)$. In Sec. V we present a unified approach to the evaluation of the denominator polynomials that applies to both $U(n)$ and $O(n)$. In Sec. VI we restrict ourselves to $U(n)$ and some of its basic properties. In Sec. VII we determine the $U(n)$ numerator polynomials and hence the (multiplicity-free) RWC's for $U(n)$. In Sec. VIII we consider some interesting examples and how they may be applied to determine the eigenvalues of certain $U(n-1)$-Casimir invariants in the universal enveloping algebra of $U(n)$. Finally, in Sec. IX, we consider the pattern calculus laws of Biedenharn *et al.*[12] cast into our framework, keeping in mind extensions to more general groups.

II. PRELIMINARIES

Following Gel'fand the Tsetlin[7] a basis for the finite-dimensional irreducible representations of a (compact) semisimple Lie group G may be constructed by considering a suitable chain of subgroups

$$G = G_n \supset G_{n-1} \supset \cdots \supset G_1, \tag{1}$$

whose Casimir invariants yield a set of commuting (Hermitian) operators whose eigenvalues may serve to label the basis states. For the so-called canonical[8,30] subgroup chains [i.e., $G_i = U(i)$, $i = 1,...,n$, or $G_i = O(i+1)$, $i = 1,...,n$] this method (cf. Weyl's branching law[4]) in fact yields a complete labeling scheme for the basis states. For general subgroup chains, however, this method of labeling is incomplete and it is necessary to supplement the Casimir invariants of the subgroup chain one is considering with additional labeling operators.

The advantage of working with a basis for the irreducible representations of $G = G_n$, which is symmetry adapted to the subgroup chain (1), is that the (multiplicity-free) Wigner coefficients of the group factor into a Wigner coefficient of the subgroup G_{n-1} times a G_{n-1}-invariant part called the G_n: G_{n-1} reduced Wigner coefficient (RWC). Thus every (multiplicity-free) Wigner coefficient for the group G may be expressed as a product of RWC's for the subgroup chain (1) and the problem of evaluating Wigner coefficients is reduced to that of evaluating the RWC's, which are the main concern of this paper. We are thus naturally led to investigate subalgebra imbeddings $L_0 \subset L$, where L is a semisimple Lie algebra and L_0 a semisimple subalgebra.

Following Humphreys,[31] let L be a complex semisimple Lie algebra of rank l, let U be the universal enveloping algebra of L, and let Z be the center of U. Select a Cartan subalgebra H of L, with dual space H^*, and let $\Phi \subset H^*$ denote the set of roots of L relative to H. Let $\Phi^+ \subset \Phi$ denote the system of positive roots, $\Delta \subseteq \Phi^+$ the system of simple roots, and take δ to be half the sum of the positive roots. Let $(\ ,\ )$ denote the inner product induced on H^* by the Killing form and for $\lambda \in H^*$, $\alpha \in \Phi$, set

$$\langle \lambda,\alpha \rangle = 2(\lambda,\alpha)/(\alpha,\alpha)\,.$$

Finally let $\Lambda^+ \subset H^*$ be the set of dominant integral linear functions on H (i.e., those $\lambda \in H^*$ such that $\langle \lambda,\alpha \rangle \in \mathbb{Z}^+$ for all $\alpha \in \Delta$) and let W denote the Weyl group.

For any $\nu \in H^*$, let t_ν denote the translation map on H^* defined by

$$t_\nu(\lambda) = \lambda + \nu, \quad \text{for all } \lambda \in H^*\,.$$

The translated Weyl group $\tilde{W}$ is defined as the conjugate $t_{-\delta} W t_\delta$ of W in the group of invertible affine transformations of H^*. Thus every element of $\tilde{W}$ is of the form

$$\tilde{\sigma} = t_{-\delta} \sigma t_\delta \quad \sigma \in W\,.$$

Therefore $\tilde{W}$ acts on H^* according to

$$\tilde{\sigma}(\lambda) = \sigma(\lambda + \delta) - \delta, \quad \lambda \in H^*\,.$$

Now let R denote the ring of polynomial functions on H^*. It is often convenient[32,33] to identify R with the universal enveloping algebra $U(H)$ of the Cartan subalgebra H. The Weyl group W acts on R, where, if $\sigma \in W$ and $f \in R$, $\lambda \in H^*$, then

$$(\sigma f)(\lambda) = f(\sigma^{-1}(\lambda))\,.$$

Similarly the translated Weyl group $\tilde{W}$ acts on R according to

$$(\tilde{\sigma} f)(\lambda) = f(\tilde{\sigma}^{-1}(\lambda)) = f(\sigma^{-1}(\lambda + \delta) - \delta)\,. \quad (2)$$

It is a well-known result, due to Harish Chandra,[31,32] that the center Z of U is generated as an algebra by l algebraically independent invariants $z_1,\ldots,z_l$. In polynomial algebra notation we write

$$Z = \mathbb{C}[z_1 \cdots z_l]\,. \quad (3)$$

We define an infinitesimal character χ as an algebra homomorphism χ: $Z \to \mathbb{C}$: it is uniquely determined by the numbers $\chi(z_i)$ $(i = 1,\ldots,l)$, which may be arbitrary complex numbers. If v_+^λ is a maximal weight vector of weight $\lambda \in H^*$, then v_+^λ determines an algebra homomorphism

$$\chi_\lambda\colon Z \to \mathbb{C}\,,$$

where $\chi_\lambda(z)$ is the eigenvalue of $z \in Z$ on v_+^λ. If $V(\lambda)$ is a finite-dimensional irreducible U-module with highest weight $\lambda \in \Lambda^+$ and $z \in Z$, Schur's lemma implies that z takes a constant value on $V(\lambda)$. Since $V(\lambda)$ has a highest weight vector (which is unique, up to scalar multiples) of weight λ this eigenvalue is clearly given by $\chi_\lambda(z)$. We say that $V(\lambda)$ admits the infinitesimal character χ_λ.

The infinitesimal characters χ_λ play a fundamental role in character analysis since it is a theorem of Harish Chandra[31,32] that every infinitesimal character χ over Z is of the form $\chi = \chi_\lambda$ for some $\lambda \in H^*$. The following result on infinitesimal characters is due to Harish Chandra.

Theorem (1): $\chi_\lambda = \chi_\mu$ if and only if λ and μ are $\tilde{W}$ conjugate; i.e.,

$$\lambda = \sigma(\mu + \delta) - \delta, \quad \text{for some } \sigma \in W\,.$$

From the remarks above we may associate every element $z \in Z$ with a polynomial function $f_z \in R$ defined by

$$f_z(\lambda) = \chi_\lambda(z), \quad \lambda \in H^*\,.$$

In view of Eq. (2) and theorem (1) we see that the polynomial function f_z is fixed by all elements of $\tilde{W}$ (cf. Ref. 33): viz.

$$(\tilde{\sigma} f_z)(\lambda) = f_z(\lambda), \quad \lambda \in H^*, \quad \tilde{\sigma} \in \tilde{W}\,.$$

It is a theorem of Harish Chandra that the mapping

$$z \to f_z, \quad z \in Z\,,$$

gives the algebra isomorphism of Z onto the ring $\tilde{R}$ of $\tilde{W}$-invariant polynomial functions on H^* (called the Harish Chandra isomorphism).

Consider, for example, the universal Casimir element C_L of L defined by

$$C_L = \sum_r x^r x_r = \sum_r x_r x^r\,, \quad (4)$$

where $\{x_r\}$ denotes a basis for L with corresponding dual basis $\{x^r\}$ with respect to the Killing form on L. The eigenvalue of C_L on a maximal weight state of weight $\lambda \in H^*$ is given by the well-known formula

$$\chi_\lambda(C_L) = (\lambda,\lambda + 2\delta)\,. \quad (5)$$

Thus C_L may be identified with the $\tilde{W}$-invariant polynomial function $f_L(\lambda) = (\lambda,\lambda + 2\delta)$ under Harish Chandra's isomorphism.

Now let L_0 be a complex semisimple Lie subalgebra of L of rank $l_0(\leqslant l)$. With regard to the Lie algebra L_0 we follow the notation above except that we add a subscript $_0$ to everything to indicate we are considering the semisimple subalgebra L_0 rather than L. The notation we adopt is obvious in the present context. Thus we denote the universal enveloping algebra of L_0 and its center by U_0 and Z_0, respectively: we assume, without loss of generality, that $U_0 \subset U$. Similarly we denote the Cartan subalgebra of L_0 by H_0 and the corresponding root system by Φ_0, etc. We denote the set of dominant integral weights of L_0 by $\Lambda_0^+ \subset H_0^*$ and the Weyl group (resp. translated Weyl group) by W_0 (resp. $\tilde{W}_0$). In analogy with Eq. (3) we may write

$$Z_0 = \mathbb{C}[z_1',\ldots,z_{l_0}']\,, \quad (6)$$

where $z_1',\ldots,z_{l_0}'$ are algebraically independent.

We similarly define the ring of polynomial functions R_0 on the Cartan subalgebra dual H_0^* and infinitesimal characters χ_{λ_0}, $\lambda_0 \in H_0^*$, on the center Z_0 of U_0. Throughout this paper we shall always denote weights of L by Greek letters $\lambda \in H^*$ and weights of L_0 by subscripted Greek letters $\lambda_0 \in H_0^*$.

Finally we denote a finite-dimensional irreducible U- (resp. U_0-) module with highest weight $\lambda\in\Lambda^+$ (resp. $\lambda_0\in\Lambda_0^+$) by $V(\lambda)$ [resp. $V(\lambda_0)$] and we let π_λ (resp. π_{λ_0}) denote the representation afforded by $V(\lambda)$ [resp. $V(\lambda_0)$].

We shall be interested, in the following, with the ring $\mathscr{R}$ of polynomial functions on $H^*\times H_0^*$. We denote elements of $H^*\times H_0^*$ by $(\lambda|\lambda_0)$, where $\lambda\in H^*$, $\lambda_0\in H_0^*$. Now let $f\in\mathscr{R}$ be any polynomial function on $H^*\times H_0^*$. Then we may define the action of the Weyl groups W and W_0 on f according to

$$\begin{aligned}\sigma f(\lambda|\lambda_0) &= f(\sigma^{-1}(\lambda)|\lambda_0), \quad \sigma\in W,\\ \sigma_0 f(\lambda|\lambda_0) &= f(\lambda|\sigma_0^{-1}(\lambda_0)), \quad \sigma_0\in W_0.\end{aligned} \tag{7}$$

Equation (7) defines the action of the group $W\times W_0$ on $\mathscr{R}$. Similarly the action of the translated group $\tilde{W}\times\tilde{W}_0$ on the polynomial function $f\in\mathscr{R}$ is given by

$$\begin{aligned}\tilde\sigma f(\lambda|\lambda_0) &= f(\tilde\sigma^{-1}(\lambda)|\lambda_0), \quad \tilde\sigma\in\tilde W,\\ \tilde\sigma_0 f(\lambda|\lambda_0) &= f(\lambda|\tilde\sigma_0^{-1}(\lambda_0)), \quad \tilde\sigma_0\in\tilde W_0.\end{aligned} \tag{8}$$

An important algebra for us in the following is (the integral domain)

$$\mathscr{Z} = ZZ_0 = Z_0Z, \tag{9}$$

which clearly centralizes U_0 in U: i.e.,

$$[u_0,c]=0, \quad \text{for all } u_0\in U_0, \quad c\in\mathscr{Z}.$$

In view of Eqs. (3) and (6) we may write

$$\mathscr{Z} = \mathbb{C}[z_1,\dots,z_l\,;z_1',\dots,z_{l_0}'].$$

We shall be concerned in this paper with *canonical* imbeddings $L\supset L_0$, where L is the Lie algebra of $\mathrm{U}(n+1)$ [or $\mathrm{O}(n+1)$] and L_0 is the Lie algebra of $\mathrm{U}(n)$ [resp. $\mathrm{O}(n)$]. Hence, unless otherwise stated, throughout the remainder of this paper we assume that $L\supset L_0$ is one of the above-mentioned canonical imbeddings. Under these assumptions one may deduce that the algebra $\mathscr{Z}$ of Eq. (9) is the centralizer of U_0 in U; viz.

$$\mathscr{Z} = \{c\in U\,|\,[x,c]=0, \text{ for all } x\in L_0\}.$$

This result implies that in the reduction of a finite-dimensional irreducible U-module into irreducible modules over U_0 all U_0- modules occur with at most unit multiplicity. (We note that if the imbedding $L\supset L_0$ is noncanonical then in general the centralizer of L_0 in U properly contains the algebra $\mathscr{Z}$.)

We define an infinitesimal character $\chi_{(\lambda|\lambda_0)}$ on $\mathscr{Z}$, $(\lambda|\lambda_0)\in H^*\times H_0^*$, uniquely according to

$$\begin{aligned}\chi_{(\lambda|\lambda_0)}(z) &= \chi_\lambda(z), \quad z\in Z,\\ \chi_{(\lambda|\lambda_0)}(z_0) &= \chi_{\lambda_0}(z_0), \quad z_0\in Z_0,\end{aligned}$$

which we extend to an algebra homomorphism to all of $\mathscr{Z}$. We clearly have $\chi_{(\lambda|\lambda_0)}=\chi_{(\mu|\mu_0)}$ if and only if $\chi_\lambda=\chi_\mu$ and $\chi_{\lambda_0}=\chi_{\mu_0}$. We thus obtain the following (obvious) generalization of Theorem (1).

Theorem (2): $\chi_{(\lambda|\lambda_0)}=\chi_{(\mu|\mu_0)}$ if and only if $(\lambda|\lambda_0)$ and $(\mu|\mu_0)$ are conjugate under $\tilde W\times\tilde W_0$; i.e.,

$$\begin{aligned}\lambda &= \sigma(\mu+\delta)-\delta, \quad \text{for some } \sigma\in W,\\ \lambda_0 &= \sigma_0(\mu_0+\delta_0)-\delta_0, \quad \text{for some } \sigma_0\in W_0.\end{aligned}$$

In view of the above remarks we may associate with every centralizer element $c\in\mathscr{Z}$ the $\tilde W\times\tilde W_0$-invariant polynomial function on $H^*\times H_0^*$ defined by

$$f_c(\lambda|\lambda_0)=\chi_{(\lambda|\lambda_0)}(c), \quad \text{for all } (\lambda|\lambda_0)\in H^*\times H_0^*.$$

It is easily deduced, in view of Harish Chandra's isomorphism, that the mapping $c\to f_c$ determines an algebra isomorphism of $\mathscr{Z}$ onto the ring $\tilde{\mathscr{R}}$ of $\tilde W\times\tilde W_0$-invariant polynomial functions on $H^*\times H_0^*$, which we also call the Harish Chandra isomorphism.

We call a weight $(\lambda|\lambda_0)\in\Lambda^+\times\Lambda_0^+$ *lexical* if the irreducible U_0-module $V(\lambda_0)$ occurs in the irreducible U-module $V(\lambda)$. We denote the set of lexical weights by $\mathscr{L}\subset\Lambda^+\times\Lambda_0^+$. For $\lambda\in\Lambda^+$ we let $[\lambda]\subset\Lambda_0^+$ denote the set of weights

$$[\lambda]=\{\lambda_0\in\Lambda_0^+\,|\,(\lambda|\lambda_0)\in\mathscr{L}\}.$$

With this convention we may write the decomposition of a finite dimensional irreducible U-module $V(\lambda)$ with highest weight $\lambda\in\Lambda^+$ into irreducible U_0-modules according to

$$V(\lambda)=\bigoplus_{\lambda_0\in[\lambda]}V(\lambda|\lambda_0), \tag{10}$$

where $V(\lambda|\lambda_0)$ denotes an irreducible U_0-module with highest weight $\lambda_0\in\Lambda_0^+$ which is contained in the irreducible U-module $V(\lambda)$. We remark that the set of lexical weights $\mathscr{L}$, and hence the decomposition law (10), is easily deduced for the canonical imbeddings we are considering from the well-known Gel'fand–Weyl betweenness conditions.[7–9]

We note that the centralizer elements $c\in\mathscr{Z}$ take a constant value on the space $V(\lambda|\lambda_0)$ occurring in the decomposition (10), this eigenvalue being given by $\chi_{(\lambda|\lambda_0)}(c)$. We say that $V(\lambda|\lambda_0)$ admits the infinitesimal character $\chi_{(\lambda|\lambda_0)}$.

Following Refs. 34 and 35 there is a natural method for constructing elements of $\mathscr{Z}$. We let A_λ denote the algebra

$$A_\lambda=[\mathrm{End}\,V(\lambda)]\otimes U,$$

and consider the map

$$\partial_\lambda:\ U\to A_\lambda,$$

defined for $x\in L$ according to

$$\partial_\lambda(x)=\pi_\lambda(x)\otimes 1+1\otimes x,$$

which we extend to an algebra homomorphism to all of U. Thus, for example, the image of the universal Casimir element C_L of L under the mapping ∂_λ is given by [see Eq. (4)]

$$\partial_\lambda(C_L)=\pi_\lambda(C_L)\otimes 1+1\otimes C_L+2\sum_r\pi_\lambda(x^r)\otimes x_r. \tag{11}$$

In the following we find it convenient to consider the operator

$$b_\lambda=-\tfrac{1}{2}[\partial_\lambda(C_L)-\pi_\lambda(C_L)\otimes 1-1\otimes C_L], \tag{12a}$$

which, in view of Eq. (11), may be alternatively written

$$b_\lambda=-\sum_r\pi_\lambda(x^r)\otimes x_r. \tag{12b}$$

We may thus regard b_λ as a $D[\lambda]\times D[\lambda]$ matrix ($D[\lambda]=\dim V(\lambda)$) with entries from L. More generally we may regard the matrix powers b_λ^m as constituting $D[\lambda]\times D[\lambda]$ matrices with entries from U (cf. Ref. 33).

We now consider the trace map

$$\tau_\lambda:\ A_\lambda\to U,$$

defined by

$$\tau_\lambda : \sum_i \rho_i \otimes u_i \to \sum_i \mathrm{tr}[\rho_i] u_i , \tag{13}$$

where $\rho_i \in \mathrm{End}\, V(\lambda)$, $u_i \in U$. It follows from Ref. 34 [see Theorem (2)] that the operators

$$I_m(\lambda) = \tau_\lambda [b_\lambda^m] \tag{14}$$

are Casimir invariants of L; i.e., $I_m(\lambda) \in Z$. In passing it is interesting to note that in the case where π_λ is the fundamental contragredient vector representation, the invariants of Eq. (14) give the well-known Gel'fand invariants of the orthogonal and unitary groups.

We note that, with regard to the Lie algebra L_0, we may consider the L_0-analog of the mapping ∂_λ. Thus we let $V(\lambda_0)$ denote a finite-dimensional irreducible U_0-module with highest weight $\lambda_0 \in \Lambda_0^+$ and set $A_{\lambda_0} = [\mathrm{End}\, V(\lambda_0)] \otimes U_0$. In analogy with Eqs. (11)–(14) we may then consider the maps $\partial_{\lambda_0} : U_0 \to A_{\lambda_0}$ and $\tau_{\lambda_0} : A_{\lambda_0} \to U_0$. We similarly introduce the operator

$$a_{\lambda_0} = -\tfrac{1}{2}[\partial_{\lambda_0}(C_{L_0}) - \pi_{\lambda_0}(C_{L_0}) \otimes 1 - 1 \otimes C_{L_0}] ,$$

where C_{L_0} denotes the universal Casimir element of L_0, which may be regarded as a $D[\lambda_0] \times D[\lambda_0]$ matrix ($D[\lambda_0] = \dim V(\lambda_0)$) with entries from L_0. We then obtain the Casimir invariants

$$I_m(\lambda_0) = \tau_{\lambda_0}[a_{\lambda_0}^m] \in Z_0 .$$

More generally if $(\lambda|\lambda_0) \in \mathscr{L}$ is a lexical weight then we may consider the compound $D[\lambda] \times D[\lambda_0]$ matrices

$$b_\lambda^m a_{\lambda_0}^n , \tag{15}$$

whose entries belong to U. Taking the partial trace with respect to the space $V(\lambda_0)$ we obtain the centralizer elements

$$I_{m,n}(\lambda|\lambda_0) = \tau_{\lambda_0}(b_\lambda^m a_{\lambda_0}^n) \in \mathscr{Z} . \tag{16}$$

In view of Theorem (2) we see that the invariants of Eq. (16) are to determine $\tilde{W} \times \tilde{W}_0$-invariant polynomial functions on $H^* \times H_0^*$.

To be more explicit, let $\{v_i^{\lambda_0}\}_{i=1}^d$ [$d = \dim V(\lambda_0)$] constitute the usual Gel'fand–Tsetlin basis for the space $V(\lambda_0)$. We denote the corresponding basis for the subspace $V(\lambda|\lambda_0)$ of $V(\lambda)$ [cf. Eq. (10)] by $\{v_i^{(\lambda|\lambda_0)}\}_{i=1}^d$. Then the invariants of Eq. (16) may be written

$$I_{m,n}(\lambda|\lambda_0) = \sum_{i,j=1}^d [b_\lambda^m]_{ij} [a_{\lambda_0}^n]_{ji} ,$$

where

$$[b_\lambda^m]_{ij} = \langle v_i^{(\lambda|\lambda_0)} | b_\lambda^m | v_j^{(\lambda|\lambda_0)} \rangle \in U ,$$

$$[a_{\lambda_0}^n]_{ij} = \langle v_i^{\lambda_0} | a_{\lambda_0}^n | v_j^{\lambda_0} \rangle \in U_0 .$$

We remark, in view of the properties of trace, that the invariants of Eq. (16) are independent of the (orthonormal) basis chosen for $V(\lambda_0)$.

III. REDUCED WIGNER COEFFICIENTS

Let us consider the reduction of the tensor product module $V = V(\lambda) \otimes V(\mu)$, where $V(\lambda)$ and $V(\mu)$ denote finite-dimensional irreducible U-modules with highest weights λ and μ, respectively. There are two natural decompositions of V into irreducible U_0-submodules. We may first decompose V into irreducible U-modules to give the Clebsch–Gordan (CG) reduction

$$V(\lambda) \otimes V(\mu) = \bigoplus_{\rho \in [\lambda \otimes \mu]} \bigoplus_{\alpha=1}^{m_\rho} V(\rho)_\alpha , \tag{17}$$

where $[\lambda \otimes \mu]$ denotes the set of distinct highest weights occurring in the decomposition of $V(\lambda) \otimes V(\mu)$ and α is a multiplicity label used to distinguish the equivalent U-modules $V(\rho)$ (m_ρ in all) occurring in the decomposition (17). Next we may decompose each space $V(\rho)$ into irreducible U_0-submodules [cf. Eq. (10)] to give the following decomposition into irreducible U_0-submodules:

$$V(\lambda) \otimes V(\mu) = \bigoplus_{\rho \in [\lambda \otimes \mu]} \bigoplus_{\alpha=1}^{m_\rho} \bigoplus_{\nu_0 \in [\rho]} V(\rho|\nu_0)_\alpha . \tag{18}$$

On the other hand, we may decompose each space $V(\lambda)$ and $V(\mu)$ into irreducible U_0-submodules to give

$$V(\lambda) \otimes V(\mu) = \bigoplus_{\substack{\lambda_0 \in [\lambda] \\ \mu_0 \in [\mu]}} V(\lambda|\lambda_0) \otimes V(\mu|\mu_0) .$$

Now we may write the CG decomposition of the tensor product module $V(\lambda_0) \otimes V(\mu_0)$ according to

$$V(\lambda_0) \otimes V(\mu_0) = \bigoplus_{\nu_0 \in [\lambda_0 \otimes \mu_0]} \bigoplus_{\beta=1}^{m_{\nu_0}} V(\nu_0)_\beta ,$$

where β is a multiplicity label. Thus we obtain the following reduction of the space $V(\lambda) \otimes V(\mu)$ into irreducible U_0-submodules (the notation being obvious from the context):

$$V(\lambda) \otimes V(\mu) = \bigoplus_{\substack{\lambda_0 \in [\lambda] \\ \mu_0 \in [\mu]}} \bigoplus_{\nu_0 \in [\lambda_0 \otimes \mu_0]} \bigoplus_{\beta=1}^{m_{\nu_0}} V(\lambda_0 \otimes \mu_0|\nu_0)_\beta . \tag{19}$$

Throughout we assume that the decompositions (18) and (19) are orthogonal. It is our aim here to demonstrate that the overlap coefficients between the two decompositions (18) and (19) give the required $L: L_0$ RWC's.

Let us denote the orthogonal projection onto the submodule $V(\rho|\nu_0)_\alpha$ of Eq. (18) by $P_\alpha[\rho|\nu_0]$. Similarly we denote the orthogonal projection onto the submodule $V(\lambda_0 \otimes \mu_0|\nu_0)_\beta$ of Eq. (19) by $P_\beta[\lambda_0 \otimes \mu_0|\nu_0]$. We write the basis states for the irreducible U_0-module $V(\nu_0)$ in the form $\left|\begin{smallmatrix}\nu_0 \\ (\xi)\end{smallmatrix}\right\rangle$, where (ξ) denotes a suitable set of labels used to distinguish the basis states which we assume are orthonormal. We denote the basis states of the space $V(\lambda_0 \otimes \mu_0|\nu_0)_\beta$ by

$$\left|\begin{matrix}\lambda_0 \otimes \mu_0 \\ \nu_0, \beta \\ (\xi)\end{matrix}\right\rangle \tag{20}$$

and the basis states for the space $V(\rho|\nu_0)_\alpha$ by

$$\left|\begin{matrix}\rho, \alpha \\ \nu_0 \\ (\xi)\end{matrix}\right\rangle . \tag{21}$$

We remark that the states (20) [resp. (21)] constitute an O.N.B. for the irreducible module $V(\lambda) \otimes V(\mu)$ of the algebra $L \oplus L$, which is symmetry adapted to the chain $L \oplus L \supset L_0 \oplus L_0 \supset L_0$ (resp. $L \oplus L \supset L \supset L_0$).

The overlap coefficients for the states (20) and (21) may be written

$$\left\langle \begin{matrix}\rho,\alpha \\ \nu_0' \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda_0 \otimes \mu_0 \\ \nu_0,\beta \\ (\xi)\end{matrix} \right\rangle = \delta_{\nu_0',\nu_0}\delta_{(\xi'),(\xi)} \left\langle \begin{matrix}\rho,\alpha \\ \\ \nu_0,\beta\end{matrix} \middle\| \begin{matrix}\lambda \\ \\ \lambda_0\end{matrix} ; \begin{matrix}\mu \\ \\ \mu_0\end{matrix} \right\rangle, \tag{22}$$

where

$$\left\langle \begin{matrix}\rho,\alpha \\ \nu_0,\beta\end{matrix} \middle\| \begin{matrix}\lambda \\ \lambda_0\end{matrix} ; \begin{matrix}\mu \\ \mu_0\end{matrix} \right\rangle$$

is a constant, independent of the labels (ξ) for the subalgebra L_0, which we call the L: L_0 reduced Wigner coefficient (RWC). The proof of Eq. (22) is closely connected to Schur's lemma. To clarify the situation we note that both of the projection operators $P_\beta[\lambda_0 \otimes \mu_0|\nu_0]$ and $P_\alpha[\rho|\nu_0]$ intertwine (i.e., commute with) the action of L_0 (since they are projections onto irreducible U_0-submodules). It is thus clear, in view of Schur's lemma, that we may write

$$P_\beta[\lambda_0 \otimes \mu_0|\nu_0]P_\alpha[\rho|\nu_0]P_\beta[\lambda_0 \otimes \mu_0|\nu_0] = \gamma P_\beta[\lambda_0 \otimes \mu_0|\nu_0], \tag{23}$$

$$P_\alpha[\rho|\nu_0]P_\beta[\lambda_0 \otimes \mu_0|\nu_0]P_\alpha[\rho|\nu_0] = \gamma P_\alpha[\rho|\nu_0],$$

for some scalar γ. In the notation of Eq. (22) it is easily seen that the constant γ is given by

$$\gamma = \left| \left\langle \begin{matrix}\rho,\alpha \\ \nu_0,\beta\end{matrix} \middle\| \begin{matrix}\lambda \\ \lambda_0\end{matrix} ; \begin{matrix}\mu \\ \mu_0\end{matrix} \right\rangle \right|^2 .$$

Thus Eq. (23) may be regarded as an operator generalization of Eq. (22).

To see how this ties up with the evaluation of Wigner coefficients for L we note that we may choose a basis for the irreducible representations $V(\lambda)$ and $V(\mu)$, which is symmetry adapted to the imbedding $L \supset L_0$. Since we are assuming a canonical imbedding we may write the (orthonormal) basis states in the Gel'fand–Tsetlin form

$$\left| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} \right\rangle, \quad \left| \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle,$$

where $\lambda_0 \in [\lambda]$ (resp. $\mu_0 \in [\mu]$) and where (ξ) [resp. (η)] denotes a set of labels to distinguish the basis states of $V(\lambda_0)$ [resp. $V(\mu_0)$].

Now we have, for our Wigner coefficients,

$$\left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\lambda \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle = \sum_\beta \left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| P_\beta[\lambda_0 \otimes \mu_0|\nu_0] \middle| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle,$$

where

$$\left| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle$$

is shorthand notation for the tensor product state

$$\left| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} \right\rangle \otimes \left| \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle .$$

Introducing a complete set of states for the space $V(\lambda_0 \otimes \mu_0|\nu_0)_\beta$ we may thus write

$$\begin{aligned}
\left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle &= \sum_{\beta,(\eta')} \left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda_0 \otimes \mu_0 \\ \nu_0,\beta \\ (\eta')\end{matrix} \right\rangle \left\langle \begin{matrix}\lambda_0 \otimes \mu_0 \\ \nu_0,\beta \\ (\eta')\end{matrix} \middle| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle \\
&= \sum_{\beta,(\eta')} \left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda_0 \otimes \mu_0 \\ \nu_0,\beta \\ (\eta')\end{matrix} \right\rangle \left\langle \begin{matrix}\nu_0,\beta \\ \\ (\eta')\end{matrix} \middle| \begin{matrix}\lambda_0 \\ \\ (\xi)\end{matrix} ; \begin{matrix}\mu_0 \\ \\ (\eta)\end{matrix} \right\rangle \\
&= \sum_\beta \left\langle \begin{matrix}\rho,\alpha \\ \\ \nu_0,\beta\end{matrix} \middle\| \begin{matrix}\lambda \\ \\ \lambda_0\end{matrix} ; \begin{matrix}\mu_0 \\ \\ \mu_0\end{matrix} \right\rangle \left\langle \begin{matrix}\nu_0,\beta \\ \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda_0 \\ \\ (\xi)\end{matrix} ; \begin{matrix}\mu_0 \\ \\ (\eta)\end{matrix} \right\rangle,
\end{aligned} \tag{24}$$

where the last equality follows from Eq. (22). Thus we have shown that a Wigner coefficient for L reduces to a sum (over L_0-multiplicity) of terms each of which factor into a RWC, which is L_0-invariant, times a Wigner coefficient for L_0. In the special case where $V(\nu_0)$ occurs exactly once in $V(\lambda_0) \otimes V(\mu_0)$ (i.e., no L_0-multiplicity), Eq. (24) reduces to a single product

$$\left\langle \begin{matrix}\rho,\alpha \\ \nu_0 \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda \\ \lambda_0 \\ (\xi)\end{matrix} ; \begin{matrix}\mu \\ \mu_0 \\ (\eta)\end{matrix} \right\rangle = \left\langle \begin{matrix}\rho,\alpha \\ \\ \nu_0\end{matrix} \middle\| \begin{matrix}\lambda \\ \\ \nu_0\end{matrix} ; \begin{matrix}\mu \\ \\ \mu_0\end{matrix} \right\rangle \left\langle \begin{matrix}\nu_0 \\ \\ (\xi')\end{matrix} \middle| \begin{matrix}\lambda_0 \\ \\ (\xi)\end{matrix} ; \begin{matrix}\mu_0 \\ \\ (\eta)\end{matrix} \right\rangle, \tag{25}$$

where the L_0 Wigner coefficient on the right-hand side (rhs) is multiplicity-free.

It is important to note that the RWC of Eq. (22) will depend on both the L and L_0 multiplicity labels α and β. In such a case the RWC's will depend on the specific (multiplicity) labeling scheme chosen and hence cannot be evaluated using the properties of the Lie algebras L and L_0 alone. Nevertheless there is a class of RWC's (herein referred to as *optimal*) that may be evaluated simply as a rational polynomial function in the representation labels of L and L_0.

It is well known[18,31] that the irreducible U-modules occurring in the tensor product space $V(\lambda) \otimes V(\mu)$ have highest weights of the form $\mu + \Delta$, where Δ is a weight in $V(\lambda)$. Moreover the multiplicity $m(\mu + \Delta : \lambda \otimes \mu)$ of $V(\mu + \Delta)$ in $V(\lambda) \otimes V(\mu)$ is less than or equal to the multiplicity $m_\lambda(\Delta)$ of the weight Δ in $V(\lambda)$. An analogous statement holds for the CG reduction of the tensor product $V(\lambda_0) \otimes V(\mu_0)$ of two irreducible U_0-modules. In particular if the weight Δ (resp. Δ_0) is Weyl group conjugate to the maximal weight λ (resp. λ_0) then the irreducible module $V(\mu + \Delta)$ [resp. $V(\mu_0 + \Delta_0)$] occurs with at most unit multiplicity. In this case there is no L (resp. L_0) tensor product multiplicity problem to be considered. This leads us to consider the special RWC's of the form

$$\left\langle \begin{matrix} \mu + \Delta \\ \\ \mu_0 + \Delta_0 \end{matrix} \middle\| \begin{matrix} \lambda & & \mu \\ & ; & \\ \lambda_0 & & \mu_0 \end{matrix} \right\rangle ,$$

where Δ (resp. Δ_0) is Weyl group conjugate to the highest weight λ (resp. λ_0). For simplicity we call the special RWC's of the above form *optimal*. These are the RWC's for which the pattern calculus rules of Biedenharn and Louck[12] apply. It is our aim in this paper to present a direct (algebraic) method for the evaluation of the optimal RWC's that applies to the canonical imbeddings $\mathrm{U}(n+1) \supset \mathrm{U}(n)$ and $\mathrm{O}(n+1) \supset \mathrm{O}(n)$.

We henceforth denote the set of distinct weights in $V(\lambda)$ [resp. $V(\lambda_0)$] by $S(\lambda)$ [resp. $S(\lambda_0)$] and we let $\mathrm{Sym}(\lambda)$ [resp. $\mathrm{Sym}(\lambda_0)$] denote the set of weights Weyl group conjugate to λ (resp. λ_0). We now let b_λ be the matrix of Eq. (12). Acting on the irreducible U-module $V(\mu)$ the matrix b_λ may be viewed as an operator on the tensor product space $V(\lambda) \otimes V(\mu)$:

$$b_\lambda = -\tfrac{1}{2}[\pi_{\lambda \otimes \mu}(C_L) - \pi_\lambda(C_L) \otimes 1 - 1 \otimes \pi_\mu(C_L)] .$$

It follows that the matrix b_λ satisfies the following polynomial identity on the space $V(\mu)$ (cf. Ref. 33)

$$\prod_{\nu \in S(\lambda)} (b_\lambda - \beta_{\lambda,\nu}(\mu)) = 0 ,$$

where $\beta_{\lambda,\nu}$ denotes the linear polynomial function on H^* defined by

$$\begin{aligned} \beta_{\lambda,\nu}(\mu) &= -\tfrac{1}{2}[\chi_{\mu+\nu}(C_L) - \chi_\lambda(C_L) - \chi_\mu(C_L)] \\ &= \tfrac{1}{2}(\lambda,\lambda + 2\delta) - \tfrac{1}{2}(\nu,\nu + 2(\mu + \delta)), \quad \mu \in H^* , \end{aligned} \tag{26}$$

where we have used Eq. (5).

We note that the translated Weyl group $\tilde{W}$ acts on the polynomial functions $\beta_{\lambda,\nu}$ according to

$$\tilde{\sigma}\beta_{\lambda,\nu}(\mu) = \beta_{\lambda,\nu}[\tilde{\sigma}^{-1}(\mu)] = \beta_{\lambda,\sigma(\nu)}(\mu), \quad \mu \in H^* ,$$

i.e.,

$$\tilde{\sigma}\beta_{\lambda,\nu} = \beta_{\lambda,\sigma(\nu)}, \quad \sigma \in W . \tag{27}$$

Thus $\tilde{W}$ permutes the polynomial functions $\beta_{\lambda,\nu}$, $\nu \in S(\lambda)$. Following Ref. 32 we find it convenient to regard the roots $\beta_{\lambda,\nu}$ in a representation-independent way as operators that belong to an algebraic extension of the center Z of U whose eigenvalues on an irreducible U-module with highest weight $\mu \in \Lambda^+$ are given by Eq. (26).

In an analogous way we deduce that the matrix a_{λ_0} [see remarks preceding Eq. (15)] satisfies the following polynomial identity acting on the irreducible U_0-module $V(\mu_0)$:

$$\prod_{\nu_0 \in S(\lambda_0)} (a_{\lambda_0} - \alpha_{\lambda_0,\nu_0}(\mu_0)) = 0 ,$$

where α_{λ_0,ν_0} denotes the linear polynomial function on H_0^* defined by

$$\alpha_{\lambda_0,\nu_0}(\mu_0) = \tfrac{1}{2}(\lambda_0,\lambda_0 + 2\delta_0) - \tfrac{1}{2}(\nu_0,\nu_0 + 2(\mu_0 + \delta_0)) . \tag{28}$$

In analogy with Eq. (27) we see that the translated Weyl group $\tilde{W}_0$ acts on the polynomial functions α_{λ_0,ν_0} according to

$$\tilde{\sigma}_0 \alpha_{\lambda_0,\nu_0} = \alpha_{\lambda_0,\sigma_0(\nu_0)}, \quad \sigma_0 \in W_0 . \tag{29}$$

We now introduce the operators

$$T\begin{bmatrix} \lambda \\ \Delta \end{bmatrix} = \prod_{\substack{\nu \in S(\lambda) \\ \neq \Delta}} (b_\lambda - \beta_{\lambda,\nu}), \quad \Delta \in \mathrm{Sym}(\lambda) ,$$

$$T\begin{bmatrix} \lambda_0 \\ \Delta_0 \end{bmatrix} = \prod_{\substack{\nu_0 \in S(\lambda_0) \\ \neq \Delta_0}} (a_{\lambda_0} - \alpha_{\lambda_0,\nu_0}), \quad \Delta_0 \in \mathrm{Sym}(\lambda_0) .$$

We also find it convenient to consider the compound operators

$$T\begin{bmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{bmatrix} = T\begin{bmatrix} \lambda \\ \Delta \end{bmatrix} T\begin{bmatrix} \lambda_0 \\ \Delta_0 \end{bmatrix} , \tag{30}$$

which may be viewed as a $D[\lambda] \times D[\lambda_0]$ matrix with entries given by

$$\left\langle \begin{matrix} \lambda \\ \mu_0 \\ (\xi') \end{matrix} \middle| T\begin{bmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{bmatrix} \middle| \begin{matrix} \lambda_0 \\ (\xi) \end{matrix} \right\rangle = \sum_{(\eta)} \left\langle \begin{matrix} \lambda \\ \mu_0 \\ (\xi') \end{matrix} \middle| T\begin{bmatrix} \lambda \\ \Delta \end{bmatrix} \middle| \begin{matrix} \lambda \\ \lambda_0 \\ (\eta) \end{matrix} \right\rangle \left\langle \begin{matrix} \lambda_0 \\ (\eta) \end{matrix} \middle| T\begin{bmatrix} \lambda_0 \\ \Delta_0 \end{bmatrix} \middle| \begin{matrix} \lambda_0 \\ (\xi) \end{matrix} \right\rangle \in U$$

in a basis for $V(\lambda)$ symmetry adapted to the imbedding $L \supset L_0$ [cf. Eq. (21)]. The operators (30) may be expressed as a linear combination of the compound matrices

$$b_\lambda^m a_{\lambda_0}^n$$

of Eq. (15). Taking the λ_0-trace of this latter operator we obtain the invariants $I_{m,n}(\lambda|\lambda_0)$ of Eq. (16). These invariants take a constant value on the irreducible U_0-modules $V(\mu|\mu_0) \subseteq V(\mu)$, $\mu_0 \in [\mu]$, denoted by

$$\chi_{(\mu|\mu_0)}[I_{m,n}(\lambda|\lambda_0)] ,$$

which determines a $\tilde{W} \times \tilde{W}_0$-invariant polynomial function on $H^* \times H_0^*$.

Taking the λ_0-trace of Eq. (30) we obtain the U_0-invariants

$$\tau\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix} = \tau_{\lambda_0}\left(T\begin{bmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{bmatrix}\right) ,$$

which determines a polynomial function on $H^* \times H_0^*$ defined by

$$\tau\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}(\mu|\mu_0) \equiv \chi_{(\mu|\mu_0)}\left[\tau\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}\right] , \quad (\mu|\mu_0) \in H^* \times H_0^* . \tag{31}$$

Using the action of the translated Weyl group on the polynomial functions $\beta_{\lambda,\nu}$ and α_{λ_0,ν_0} together with the fact that the invariants of Eq. (16) determine $\tilde{W}\times\tilde{W}_0$-invariant polynomial functions, we deduce that the action of the translated Weyl group on the polynomial functions (31) is given by

$$\tilde{\sigma}\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\tau\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \Delta_0\end{pmatrix},\quad \sigma\in W,$$
$$\tilde{\sigma}_0\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \sigma_0(\Delta_0)\end{pmatrix},\quad \sigma_0\in W_0. \tag{32}$$

We similarly have the polynomial functions on $H^*\times H_0^*$ defined by

$$\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=\xi\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}(\mu)\xi\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}(\mu_0),$$
$$(\mu|\mu_0)\in H^*\times H_0^*, \tag{33}$$

where $\xi(\begin{smallmatrix}\lambda\\ \Delta\end{smallmatrix})$ and $\xi(\begin{smallmatrix}\lambda_0\\ \Delta_0\end{smallmatrix})$ denote the polynomial functions on H^* and H_0^* defined by

$$\xi\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}=\prod_{\substack{\nu\in S(\lambda)\\ \neq\Delta}}(\beta_{\lambda,\Delta}-\beta_{\lambda,\nu}),$$
$$\xi\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}=\prod_{\substack{\nu_0\in S(\lambda_0)\\ \neq\Delta_0}}(\alpha_{\lambda_0,\Delta_0}-\alpha_{\lambda_0,\nu_0}),$$

respectively. In view of Eqs. (27) and (29) we deduce that the translated Weyl group $\tilde{W}\times\tilde{W}_0$ acts on the polynomial functions $\xi(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ according to

$$\tilde{\sigma}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\xi\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \Delta_0\end{pmatrix},\quad \sigma\in W,$$
$$\tilde{\sigma}_0\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \sigma_0(\Delta_0)\end{pmatrix},\quad \sigma_0\in W_0. \tag{34}$$

Now suppose $\Delta\in\mathrm{Sym}(\lambda)$, $\Delta_0\in\mathrm{Sym}(\lambda_0)$. We may then consider the functions on $\Lambda^+\times\Lambda_0^+$ defined by

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=\left|\left\langle\begin{matrix}\mu+\Delta\\ \\ \mu_0+\Delta_0\end{matrix}\right|\left|\begin{matrix}\lambda & & \mu\\ & ; & \\ \lambda_0 & & \mu_0\end{matrix}\right\rangle\right|^2,$$
$$(\mu|\mu_0)\in\Lambda^+\times\Lambda_0^+, \tag{35}$$

which determines the squared L: L_0 RWC's. Note that if $(\mu|\mu_0)$ is lexical then the rhs of Eq. (35) vanishes unless the weight $(\mu+\Delta|\mu_0+\Delta_0)$ is lexical. We now show that the functions $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ are related to the polynomial functions of Eqs. (31) and (33) by the relation

$$\frac{D[\mu_0+\Delta_0]}{D[\mu_0]}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} \tag{36}$$

which holds on $\Lambda^+\times\Lambda_0^+$ (where, as usual, $D[\mu_0]$ denotes the well-known Weyl dimension function[30]). Equation (36) defines the function $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ as a rational polynomial function on $H^*\times H_0^*$ (or at least a Zariski-dense subset of $H^*\times H_0^*$) and hence enables an extension of the definition of optimal RWC's to nondominant weights $(\mu|\mu_0)\in H^*\times H_0^*$. Such considerations are likely to be of importance in discussing noncompact real forms of L and L_0.

To prove Eq. (36) we note that the eigenvalue of the U_0-invariant $\tau(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ on the U_0-module $V(\mu|\mu_0)\subseteq V(\mu)$, $\mu_0\in[\mu]$, is given by

$$\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=\frac{1}{D[\mu_0]}\sum_{(\eta)}\left\langle\begin{matrix}\mu\\ \mu_0\\ (\eta)\end{matrix}\right|\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}\left|\begin{matrix}\mu\\ \mu_0\\ (\eta)\end{matrix}\right\rangle$$
$$=\frac{1}{D[\mu_0]}\sum_{(\eta),(\xi)}\left\langle\begin{matrix}\mu & & \lambda\\ \mu_0 & ; & \lambda_0\\ (\eta) & & (\xi)\end{matrix}\right|T\begin{bmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{bmatrix}\left|\begin{matrix}\lambda & & \mu\\ \lambda_0 & ; & \mu_0\\ (\xi) & & (\eta)\end{matrix}\right\rangle$$
$$=\frac{1}{D[\mu_0]}\tau_{\lambda_0\otimes\mu_0}\left(T\begin{bmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{bmatrix}\right), \tag{37}$$

where $\tau_{\lambda_0\otimes\mu_0}$ denotes the partial trace with respect to the subspace $V(\lambda|\lambda_0)\otimes V(\mu|\mu_0)$ of $V(\lambda)\otimes V(\mu)$. We now note that the operator $T[\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix}]$ is zero on the spaces $V(\rho|\nu_0)$ of Eq. (18) unless $\rho=\mu+\Delta$ and $\nu_0=\mu_0+\Delta_0$ in which case $T[\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix}]$ takes the constant value $\xi(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ $(\mu|\mu_0)$. If we let $P[\mu+\Delta|\mu_0+\Delta_0]$ denote the orthogonal projection onto the subspace $V(\mu+\Delta|\mu_0+\Delta_0)$ of $V(\lambda)\otimes V(\mu)$ (where we have dropped the multiplicity label α in view of the fact that Δ is W-conjugate to the highest weight λ) we may write, in view of Eq. (37),

$$\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=\frac{1}{D[\mu_0]}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)\tau_{\lambda_0\otimes\mu_0}(P[\mu+\Delta|\mu_0+\Delta_0])$$
$$=\frac{1}{D[\mu_0]}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)\tau_{\lambda_0\otimes\mu_0}(P[\lambda_0\otimes\mu_0|\mu_0+\Delta_0]P[\mu+\Delta|\mu_0+\Delta_0]P[\lambda_0\otimes\mu_0|\mu_0+\Delta_0]),$$

where $P[\lambda_0\otimes\mu_0|\mu_0+\Delta_0]$ denotes the orthogonal projector onto the subspace $V(\lambda_0\otimes\mu_0|\mu_0+\Delta_0)$ of Eq. (19) noting that in this case there is no multiplicity label β (since Δ_0 is W_0-conjugate to the highest weight λ_0).

Thus using Eq. (23) we obtain

$$\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0) = \frac{1}{D[\mu_0]}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)\left|\left\langle\begin{matrix}\mu+\Delta\\ \\ \mu_0+\Delta_0\end{matrix}\right|\left|\begin{matrix}\lambda & & \mu\\ & ; & \\ \lambda_0 & & \mu_0\end{matrix}\right\rangle\right|^2 \tau_{\lambda_0\otimes\mu_0}(P[\lambda_0\otimes\mu_0|\mu_0+\Delta_0])$$

$$= \frac{D[\mu_0+\Delta_0]}{D[\mu_0]}\xi\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)\left|\left\langle\begin{matrix}\mu+\Delta\\ \\ \mu_0+\Delta_0\end{matrix}\right|\left|\begin{matrix}\lambda & & \mu\\ & ; & \\ \lambda_0 & & \mu_0\end{matrix}\right\rangle\right|^2,$$

which is Eq. (36) as required.

Equation (36) demonstrates that we may evaluate the optimal RWC's of Eq. (35) once the spectrum of the operators $\tau(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ is determined. After canceling out common factors between the polynomial functions $\tau(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ and $\xi(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$, Eq. (36) then affords a useful expression for the squared RWC's of Eq. (35) as a rational polynomial function (numerator polynomial divided by denominator polynomial) in the representation labels of L and L_0.

For ease of notation we shall henceforth refer to the rational polynomial functions $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ as reduced Wigner functions (RWF's). The remainder of this paper is devoted to the explicit evaluation and general properties of RWF's. It shall be implicitly assumed, unless otherwise stated, that all RWF's $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ are optimal: i.e, $\Delta\in\mathrm{Sym}(\lambda)$, $\Delta_0\in\mathrm{Sym}(\lambda_0)$.

IV. SYMMETRIES AND COMPOSITION OF RWF'S

Before going on to the explicit evaluation of the optimal RWF's we consider here certain symmetry properties of the RWF's, which may be obtained from general considerations.

We begin by noting that Eq. (36) implies certain Weyl group symmetries for the RWF's of Eq. (35). Using the symmetry relations (32) and (34) together with Eq. (36) we see that the translated Weyl group $\tilde{W}\times\tilde{W}_0$ acts on the RWF's of Eq. (35) according to

$$\begin{aligned}\tilde{\sigma}\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \Delta_0\end{pmatrix}, \quad \sigma\in W,\\ \tilde{\sigma}_0\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \sigma_0(\Delta_0)\end{pmatrix}, \quad \sigma_0\in W_0,\end{aligned} \tag{38}$$

where we have used the transformation law[33,34]

$$\tilde{\sigma}_0\frac{D[\mu_0+\Delta_0]}{D[\mu_0]} = \frac{D[\mu_0+\sigma_0(\Delta_0)]}{D[\mu_0]}.$$

In the case of the canonical imbedding $U(n+1)\supset U(n)$ the symmetry relations of Eq. (38) reduce to the generalized Weyl group symmetries on reduced Wigner operators obtained previously by Biedenharn and Louck.[12] The important thing from our point of view is that the relations (38) fall out naturally in our approach from the simple transformation properties of polynomial functions.

Equation (38) demonstrates that it suffices to evaluate the single optimal RWF

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}, \tag{39}$$

the rest following from Weyl group symmetry:

$$\begin{aligned}&\rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\lambda) & \sigma_0(\lambda_0)\end{pmatrix}(\mu|\mu_0)\\ &= \tilde{\sigma}\tilde{\sigma}_0\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\mu|\mu_0)\\ &= \rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\tilde{\sigma}^{-1}(\mu)|\tilde{\sigma}_0^{-1}(\mu_0)).\end{aligned} \tag{40}$$

We call the special RWF's of Eq. (39) *semimaximal*. In the case of the canonical imbedding $U(n+1)\supset U(n)$ the rational polynomial function (39) is equivalent to the semimaximal reduced Wigner operator of Biedenharn and Louck[12] (see in particular Sec. VII). In the language of Ref. 12 this corresponds to the matrix elements of the most general boson monomial.

It should be emphasized that in any application it is desirable to find the optimum way of expressing a RWC. Although it turns out that we may obtain an explicit expression for the optimal RWC's directly it may be desirable, from the point of view of actual calculations, to express a RWC as a product of certain elementary RWC's. This leads us to investigate the composition of two RWF's in the hope that a general pattern calculus will emerge for manipulating and multiplying RWF's.

Following Biedenharn and Louck[12] we define the composition of two optimal RWF's according to

$$\begin{aligned}&\left[\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix}\right](\mu|\mu_0)\\ &= \rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu+\Delta'|\mu_0+\Delta_0')\,\rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix}(\mu|\mu_0).\end{aligned} \tag{41}$$

It is easily demonstrated that the operation of composition, as defined above, is not generally commutative. We say that two optimal RWF's

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} \quad\text{and}\quad \rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix} \tag{42}$$

commute if and only if

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix} = \rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix}\circ\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}.$$

Following Ref. 12 we say that the two RWF's of Eq. (42) are *tied* (or have the same *tie structure*) if there exists Weyl group elements $\sigma\in W$, $\sigma_0\in W_0$, such that

$$\begin{aligned}&\sigma(\lambda)=\Delta, \quad \sigma(\nu)=\Delta',\\ &\sigma_0(\lambda_0)=\Delta_0, \quad \sigma_0(\nu_0)=\Delta_0'.\end{aligned}$$

In such a case we write

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}\sim\rho\begin{pmatrix}\nu & \nu_0\\ \Delta' & \Delta_0'\end{pmatrix}.$$

A trivial consequence of this definition is that we always have

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\lambda) & \sigma_0(\lambda_0)\end{pmatrix}\sim\rho\begin{pmatrix}\nu & \nu_0\\ \sigma(\nu) & \sigma_0(\nu_0)\end{pmatrix},\quad (\lambda\,|\lambda_0),(\nu|\nu_0)\in\mathscr{L}\,.$$

The significance of RWF's of the same tie structure is that they always commute. To see this it suffices to consider the composition of two semimaximal RWF's $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{smallmatrix})$ and $\rho(\begin{smallmatrix}\nu & \nu_0\\ \nu & \nu_0\end{smallmatrix})$. We have, for $(\mu|\mu_0)\in\Lambda^+\times\Lambda_0^+$ [cf. Eq. (35)]

$$\begin{aligned}\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)&=\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\mu+\nu|\mu_0+\nu_0)\,\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)\\ &=\left|\left\langle\begin{matrix}\mu+\nu+\lambda\\ \\ \mu_0+\nu_0+\lambda_0\end{matrix}\right\|\left.\begin{matrix}\lambda & & \mu+\nu\\ & ; & \\ \lambda_0 & & \mu_0+\nu_0\end{matrix}\right\rangle\right.\\ &\quad\times\left.\left\langle\begin{matrix}\mu+\nu\\ \\ \mu_0+\nu_0\end{matrix}\right\|\left.\begin{matrix}\nu & & \mu\\ & ; & \\ \nu_0 & & \mu_0\end{matrix}\right\rangle\right|^2.\end{aligned}\tag{43}$$

Now let $v_+^{(\lambda\,|\lambda_0)}$ denote the unique (normalized) maximal weight state of L_0 of weight λ_0, which is contained in the irreducible U-module $V(\lambda)$. Such a state is commonly referred to as *semimaximal.* We now consider the Wigner coefficient

$$\langle v_+^{(\mu+\nu|\mu_0+\nu_0)}\,|\,v_+^{(\nu|\nu_0)}\otimes v_+^{(\mu|\mu_0)}\rangle\,.$$

Applying Eq. (25) we have immediately

$$\langle v_+^{(\mu+\nu|\mu_0+\nu_0)}\,|\,v_+^{(\nu|\nu_0)}\otimes v_+^{(\mu|\mu_0)}\rangle=\left\langle\begin{matrix}\mu+\nu\\ \\ \mu_0+\nu_0\end{matrix}\right\|\left.\begin{matrix}\nu & & \mu\\ & ; & \\ \nu_0 & & \mu_0\end{matrix}\right\rangle\langle v_+^{\mu_0+\nu_0}\,|\,v_+^{\nu_0}\otimes v_+^{\mu_0}\rangle\,,\tag{44}$$

where $v_+^{\mu_0}$, etc. refers to the unique (normalized) maximal weight state of the irreducible U_0-module $V(\mu_0)$. Clearly $v_+^{\nu_0}\otimes v_+^{\mu_0}$ is a normalized maximal weight state of weight $\mu_0+\nu_0$, whence we see that the maximal L_0-Wigner coefficient on the rhs of Eq. (44) has absolute value unity. Thus we may write for our semimaximal RWF's

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\mu|\mu_0)=|\langle v_+^{(\mu+\lambda\,|\mu_0+\lambda_0)}\,|\,v_+^{(\lambda\,|\lambda_0)}\otimes v_+^{(\mu|\mu_0)}\rangle|^2.\tag{45}$$

This equation implies immediately the symmetry relation

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\mu|\mu_0)=\rho\begin{pmatrix}\mu & \mu_0\\ \mu & \mu_0\end{pmatrix}(\lambda\,|\lambda_0)\,.\tag{46}$$

In passing it is interesting to note the following special case of Eq. (45):

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(0|0)=|\langle v_+^{(\lambda\,|\lambda_0)}\,|\,v_+^{(\lambda\,|\lambda_0)}\otimes v_+^{(0|0)}\rangle|^2=1\,,\tag{47}$$

where $v_+^{(0|0)}$ denotes the unique (normalized) basis vector for the trivial one-dimensional representation of L. Equation (47) will later be applied in determining the normalization constants of our RWF's.

Substituting Eq. (45) into Eq. (43) we obtain

$$\begin{aligned}\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)&=|\langle v_+^{(\mu+\nu|\mu_0+\nu_0)}\,|\,v_+^{(\nu|\nu_0)}\otimes v_+^{(\mu|\mu_0)}\rangle\\ &\quad\times\langle v_+^{(\mu+\nu+\lambda\,|\mu_0+\nu_0+\lambda_0)}\,|\,v_+^{(\lambda\,|\lambda_0)}\otimes v_+^{(\mu+\nu|\mu_0+\nu_0)}\rangle|^2.\end{aligned}$$

In view of the maximal nature of all the shifts concerned it is then easily deduced that

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)=|\langle v_+^{(\lambda+\nu+\mu|\lambda_0+\nu_0+\mu_0)}\,|\,v_+^{(\lambda\,|\lambda_0)}\otimes v_+^{(\nu|\nu_0)}\otimes v_+^{(\mu|\mu_0)}\rangle|^2.\tag{48}$$

By rearranging the terms in the triple tensor product state

$$v_+^{(\lambda\,|\lambda_0)}\otimes v_+^{(\nu|\nu_0)}\otimes v_+^{(\mu|\mu_0)}\,,$$

the following symmetry rule is seen to hold:

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)=\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}\circ\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(\mu|\mu_0)\,.$$

This equation is to hold for all $(\mu|\mu_0)\in\Lambda^+\times\Lambda_0^+$ and hence, from the properties of polynomial functions (cf. Ref. 35, Appendix D), we may write

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}=\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}\circ\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix},\tag{49a}$$

which must hold on all of $H^*\times H_0^*$ (or at least the Zariski dense subset on which the RWF's are well defined). Applying Weyl group symmetry to Eq. (49a), we obtain immediately the result

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\lambda) & \sigma_0(\lambda_0)\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \sigma(\nu) & \sigma_0(\nu_0)\end{pmatrix}=\rho\begin{pmatrix}\nu & \nu_0\\ \sigma(\nu) & \sigma_0(\nu_0)\end{pmatrix}\circ\rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\lambda) & \sigma_0(\lambda_0)\end{pmatrix}\tag{49b}$$

$(\sigma\times\sigma_0\in W\times W_0)$, which proves our assertion that all optimal RWF's of the same tie structure commute. Evidently one may establish the stronger result that two optimal RWF's commute if and only if they have the same tie structure. This result has been proved for the canonical imbedding $\mathrm{U}(n+1)\supset\mathrm{U}(n)$ by Biedenharn and Louck.[12]

We remark that in general the composition of two optimal RWF's need not yield a RWF (even if the RWF's commute). In fact it is not hard to demonstrate the relation [cf. Eq. (48)]

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0)=C^{\lambda\ \ \nu}_{\lambda_0\ \nu_0}\,\rho\begin{pmatrix}\lambda+\nu & \lambda_0+\nu_0\\ \lambda+\nu & \lambda_0+\nu_0\end{pmatrix}(\mu|\mu_0)\,,$$

where the constant $C^{\lambda\ \ \nu}_{\lambda_0\ \nu_0}$ is given by

$$C^{\lambda\ \ \nu}_{\lambda_0\ \nu_0} = \left| \left\langle \begin{matrix} \lambda+\nu \\ \\ \lambda_0+\nu_0 \end{matrix} \middle| \middle| \begin{matrix} \lambda & & \nu \\ & ; & \\ \lambda_0 & & \nu_0 \end{matrix} \right\rangle \right|^2$$
$$= \rho\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}(\nu|\nu_0) .$$

The above equation is to hold for all $(\mu|\mu_0)\in\Lambda^+\times\Lambda_0^+$ and hence we may write (via Zariski continuity)

$$\rho\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}\circ\rho\begin{pmatrix} \nu & \nu_0 \\ \nu & \nu_0 \end{pmatrix} = C^{\lambda\ \ \nu}_{\lambda_0\ \nu_0}\,\rho\begin{pmatrix} \lambda+\nu & \lambda_0+\nu_0 \\ \lambda+\nu & \lambda_0+\nu_0 \end{pmatrix}. \tag{50}$$

This result, which forms the basis of the approach of Ref. 12, demonstrates that if the squared RWC $C^{\lambda\ \ \nu}_{\lambda_0\ \nu_0}$ is unity, then the composition of two optimal RWF's of the same tie structure yields another RWF of the same tie structure. This observation is very useful for developing a pattern calculus to decompose a RWF into a product of elementary RWF's (up to a scalar). We draw particular attention here to maximal type RWF's for the orthogonal and unitary groups for which these pattern calculus laws hold.

We note that if v^λ_+ is a maximal weight vector of L then it is also a maximal weight vector for L_0 and is thus an eigenstate of the Gel'fand invariants of L and L_0. We denote the irreducible U_0-module, to which the maximal weight vector belongs, by $V(\lambda\,|\lambda)$, where $(\lambda\,|\lambda)$ denotes the lexical weight whose L_0-component $\lambda_0=\lambda$ takes maximal allowed values. We call the weight $(\lambda\,|\lambda)$ *maximally connected* and we call the associated RWF

$$\rho\begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda \end{pmatrix}$$

a RWF of *maximal type*. If v^λ_+, v^ν_+ $(\lambda,\nu\in\Lambda^+)$ denote two maximal weight states of L then we clearly have (modulo a phase)

$$v^\lambda_+\otimes v^\nu_+ = v^{\lambda+\nu}_+ ,$$

from which we deduce, in view of Eq. (50), the following combination law for maximal type RWF's:

$$\rho\begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda \end{pmatrix}\circ\rho\begin{pmatrix} \nu & \nu \\ \nu & \nu \end{pmatrix} = \rho\begin{pmatrix} \lambda+\nu & \lambda+\nu \\ \lambda+\nu & \lambda+\nu \end{pmatrix}. \tag{51}$$

Application of Weyl group symmetry then shows that the composition of two maximal type RWF's of the same tie structure yields another maximal type RWF of the same tie structure.

In the case of $\mathrm{U}(n+1)\supset\mathrm{U}(n)$ it turns out[12] that there is a relatively large class of RWF's (herein referred to as *extremal*) for which the composition of two RWF's of the same extremal tie structure yields another extremal RWF of the same tie structure. The class of extremal RWF's includes the maximal RWF's as a special case.

V. THE DENOMINATOR POLYNOMIAL

It is our eventual aim to express the rational polynomial functions $\rho(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ of Eqs. (35) and (36) in the form

$$\rho\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix} = C_{\lambda,\lambda_0}\frac{\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})}{\delta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})},$$
$$\Delta\in\mathrm{Sym}(\lambda),\quad \Delta_0\in\mathrm{Sym}(\lambda_0) , \tag{52}$$

where $\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ [resp. $\delta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$] is a polynomial function on $H^*\times H_0^*$ called the *numerator* (resp. *denominator*) polynomial and C_{λ,λ_0} is a numerical constant depending only on the labels λ and λ_0. In determining the numerator and denominator polynomials of Eq. (52) it suffices to evaluate the polynomial functions

$$\tau\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix},\quad \Delta\in\mathrm{Sym}(\lambda)\ ,\quad \Delta_0\in\mathrm{Sym}(\lambda_0)\ , \tag{53}$$

as defined by Eq. (31). We proceed by deducing the zeros of the polynomial function (53) and applying Weyl group symmetry. It turns out that the number of linear factors obtained by this method equals precisely the degree of the polynomial function (53). Unfortunately, however, in evaluating the numerator polynomials of Eq. (52) it is evidently necessary to treat the different canonical imbeddings $\mathrm{U}(n+1)\supset\mathrm{U}(n)$, $\mathrm{O}(2h+1)\supset\mathrm{O}(2h)$, $\mathrm{O}(2h)\supset\mathrm{O}(2h-1)$ separately. Nevertheless it is possible to present a unified treatment of the denominator polynomials of Eq. (52), which is the main concern of this section of the paper.

Following Ref. 3 we find it convenient to introduce the following subsets of $S(\lambda)$ and $S(\lambda_0)$:

$$S_\Delta(\lambda) = \{\rho\in S(\lambda)\,|\,\rho-\Delta = k\alpha \ \text{ for some } k\in\mathbb{Z};\ \alpha\in\Phi\},$$
$$\Delta\in\mathrm{Sym}(\lambda)\ ,$$
$$\bar S_\Delta(\lambda) = S(\lambda)\sim S_\Delta(\lambda)\ ,$$
$$S_{\Delta_0}(\lambda_0) = \{\rho_0\in S(\lambda_0)\,|\,\rho_0-\Delta_0 = k\alpha_0 \ \text{ for some } k\in\mathbb{Z};\ \alpha_0\in\Phi_0\},$$
$$\Delta_0\in\mathrm{Sym}(\lambda_0)\ , \tag{54}$$
$$\bar S_{\Delta_0}(\lambda_0) = S(\lambda_0)\sim S_{\Delta_0}(\lambda_0)\ .$$

We now note [see Ref. 35, Theorem (3) (ii)] that if $(\mu+\Delta|\mu_0+\Delta_0)\in\Lambda^+\times\Lambda_0^+$ then $T[\begin{smallmatrix}\lambda\\ \Delta\end{smallmatrix}]$ [cf. Eq. (30)] vanishes on $V(\lambda)\otimes V(\mu)$ if and only if

$$\beta_{\lambda,\Delta}(\mu) = \beta_{\lambda,\nu}(\mu)\ ,\quad \text{for some } \nu\in\bar S_\Delta(\lambda)\ . \tag{55a}$$

Similarly $T\left[\begin{smallmatrix}\lambda_0\\ \Delta_0\end{smallmatrix}\right]$ vanishes on $V(\lambda_0)\otimes V(\mu_0)$ if and only if

$$\alpha_{\lambda_0,\Delta_0}(\mu_0) = \alpha_{\lambda_0,\nu_0}(\mu_0)\ ,\quad \text{for some } \nu_0\in\bar S_{\Delta_0}(\lambda_0)\ . \tag{55b}$$

Thus we deduce that if either condition (55a) or (55b) holds then $T[\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix}]$ [cf. Eq. (30)] vanishes on $V(\lambda\,|\lambda_0)\otimes V(\mu|\mu_0)$ and hence $\tau(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})(\mu|\mu_0)=0$. Thus we deduce divisibility of the trace polynomial (53) by the set of factors

$$\prod_{\nu\in\bar S_\Delta(\lambda)}(\beta_{\lambda,\Delta}-\beta_{\lambda,\nu})\prod_{\nu_0\in\bar S_{\Delta_0}(\lambda_0)}(\alpha_{\lambda_0,\Delta_0}-\alpha_{\lambda_0,\nu_0})\ .$$

We may thus write

$$\tau\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = d_{\lambda,\lambda_0} \prod_{\nu\in\bar{S}_\Delta(\lambda)} (\beta_{\lambda,\Delta} - \beta_{\lambda,\nu})$$
$$\times \prod_{\nu_0\in\bar{S}_{\Delta_0}(\lambda_0)} (\alpha_{\lambda_0,\Delta_0} - \alpha_{\lambda_0,\nu_0}) \eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}, \tag{56}$$

where d_{λ,λ_0} is a numerical constant (depending only on the labels λ and λ_0) and $\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ is a (monic) polynomial function, herein referred to as the numerator polynomial.

Now, in view of Eq. (12), b_λ (resp. a_{λ_0}) is a matrix with entries from L (resp. L_0). Accordingly we see that the trace polynomial (53) is to determine a polynomial function of degree

$$(n_\lambda - 1) + (n_{\lambda_0} - 1),$$

where n_λ (resp. n_{λ_0}) is the number of distinct weights in $V(\lambda)$ [resp. $V(\lambda_0)$]: i.e.,

$$n_\lambda = |S(\lambda)|, \quad n_{\lambda_0} = |S(\lambda_0)|.$$

By comparison with Eq. (56) we see that the numerator polynomial $\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ is to have degree

$$|S_\Delta(\lambda)| - 1 + |S_{\Delta_0}(\lambda_0)| - 1 = r_\lambda + r_{\lambda_0},$$

where [cf. Ref. 35, Lemma (1)] the integers r_λ and r_{λ_0} are uniquely determined by

$$r_\lambda = |S_\Delta(\lambda)| - 1 = \sum_{\alpha\in\Phi^+} \langle\lambda,\alpha\rangle, \quad \Delta\in \mathrm{Sym}(\lambda),$$
$$r_{\lambda_0} = |S_{\Delta_0}(\lambda_0)| - 1 = \sum_{\alpha_0\in\Phi_0^+} \langle\lambda_0,\alpha_0\rangle, \quad \Delta_0\in \mathrm{Sym}(\lambda_0). \tag{57}$$

Substituting Eq. (56) into Eq. (36) we obtain the result

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = d_{\lambda,\lambda_0} \frac{D[\mu_0]}{D[\mu_0+\Delta_0]} \frac{\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})}{\gamma(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})}, \tag{58}$$

where $\gamma(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ denotes the polynomial function

$$\gamma\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = \gamma\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}\gamma\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix},$$

where

$$\gamma\begin{pmatrix}\lambda\\ \Delta\end{pmatrix} = \prod_{\substack{\nu\in\bar{S}_\Delta(\lambda)\\ \neq\Delta}} (\beta_{\lambda,\Delta} - \beta_{\lambda,\nu})$$

and

$$\gamma\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix} = \prod_{\substack{\nu_0\in\bar{S}_{\Delta_0}(\lambda_0)\\ \neq\Delta_0}} (\alpha_{\lambda_0,\Delta_0} - \alpha_{\lambda_0,\nu_0}).$$

In view of the results of Ref. 35 [see in particular Appendix B and the remarks preceding Eq. (4.28)] we have immediately the result

$$\gamma\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}(\mu) = C_\lambda \prod_{\substack{\alpha>0\\ \langle\Delta,\alpha\rangle>0}} \prod_{k=1}^{\langle\Delta,\alpha\rangle} [\langle\mu+\delta,\alpha\rangle + k - 1]$$
$$\times \prod_{\substack{\alpha>0\\ \langle\Delta,\alpha\rangle<0}} \prod_{k=1}^{-\langle\Delta,\alpha\rangle} [k - 1 - \langle\mu+\delta,\alpha\rangle],$$
$$\Delta\in \mathrm{Sym}(\lambda),$$

where

$$C_\lambda = \left(-\frac{1}{2}\right)^{r_\lambda} \prod_{\alpha>0} \langle\lambda,\alpha\rangle!(\alpha,\alpha)^{\langle\lambda,\alpha\rangle}. \tag{59}$$

We similarly obtain

$$\gamma\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}(\mu_0)$$
$$= C_{\lambda_0} \prod_{\substack{\alpha_0>0\\ \langle\Delta_0,\alpha_0\rangle>0}} \prod_{k=1}^{\langle\Delta_0,\alpha_0\rangle} [\langle\mu_0+\delta_0,\alpha_0\rangle + k - 1]$$
$$\times \prod_{\substack{\alpha_0>0\\ \langle\Delta_0,\alpha_0\rangle<0}} \prod_{k=1}^{\langle\Delta_0,\alpha_0\rangle} [k - 1 - \langle\mu_0+\delta_0,\alpha_0\rangle],$$
$$\Delta_0\in \mathrm{Sym}(\lambda_0),$$

where

$$C_{\lambda_0} = \left(-\frac{1}{2}\right)^{r_{\lambda_0}} \prod_{\alpha_0>0} \langle\lambda_0,\alpha_0\rangle!(\alpha_0,\alpha_0)^{\langle\lambda_0,\alpha_0\rangle}. \tag{60}$$

Substituting Eqs. (59) and (60) into Eq. (58) we may write [cf. Eq. (52)]

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = C_{\lambda,\lambda_0} \frac{\eta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})}{\delta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})}, \tag{61}$$

where the polynomial function $\delta(\begin{smallmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{smallmatrix})$ (herein referred to as the denominator polynomial) is given by

$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0) = \delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}(\mu)\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}(\mu_0), \tag{62a}$$

where

$$\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}(\mu) = \prod_{\substack{\alpha>0\\ \langle\Delta,\alpha\rangle>0}} \prod_{k=1}^{\langle\Delta,\alpha\rangle} [\langle\mu+\delta,\alpha\rangle + k - 1]$$
$$\times \prod_{\substack{\alpha>0\\ \langle\Delta,\alpha\rangle<0}} \prod_{k=1}^{-\langle\Delta,\alpha\rangle} [k - 1 - \langle\mu+\delta,\alpha\rangle] \tag{62b}$$

and

$$\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}(\mu_0) = \prod_{\substack{\alpha_0>0\\ \langle\Delta_0,\alpha_0\rangle>0}} \prod_{k=1}^{\langle\Delta_0,\alpha_0\rangle} [\langle\mu_0+\delta_0,\alpha_0\rangle + k]$$
$$\times \prod_{\substack{\alpha_0>0\\ \langle\Delta_0,\alpha_0\rangle<0}} \prod_{k=1}^{-\langle\Delta_0,\alpha_0\rangle} [k - \langle\mu_0+\delta_0,\alpha_0\rangle], \tag{62c}$$

where we have used the Weyl dimension formula

$$D[\mu_0] = \prod_{\alpha_0>0} \frac{\langle\mu_0+\delta_0,\alpha_0\rangle}{\langle\delta_0,\alpha_0\rangle}.$$

We note that the numerical constant C_{λ,λ_0} of Eq. (61) (which depends only on the labels λ and λ_0) is related to the numerical constant d_{λ,λ_0} of Eq. (58) by

$$d_{\lambda,\lambda_0} = C_\lambda C_{\lambda_0} C_{\lambda,\lambda_0},$$

with C_λ and C_{λ_0} as in Eqs. (59) and (60), respectively.

In the case of the canonical imbedding

$U(n+1)\supset U(n)$ the denominator polynomial of Eq. (62) agrees with the denominator polynomial obtained previously by Biedenharn and Louck.[12]

We remark that we have defined our denominator polynomials so that they satisfy the symmetry rule

$$\begin{aligned}\tilde{\sigma}\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \delta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \Delta_0\end{pmatrix}, \quad \sigma\in W,\\ \tilde{\sigma}_0\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \sigma_0(\Delta_0)\end{pmatrix}, \quad \sigma_0\in W_0.\end{aligned} \tag{63}$$

We see, therefore, in view of Eq. (38), that the numerator polynomials of Eq. (61) are to satisfy the symmetry condition

$$\begin{aligned}\tilde{\sigma}\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \eta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \Delta_0\end{pmatrix}, \quad \sigma\in W,\\ \tilde{\sigma}_0\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} &= \eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \sigma_0(\Delta_0)\end{pmatrix}, \quad \sigma_0\in W_0.\end{aligned} \tag{64}$$

We conclude by remarking that it has been implicitly assumed that there are no common factors between the numerator and denominator polynomials of Eq. (61) so that we have the correct numerator and denominator polynomials as required. Although this fact is not immediately obvious it shall be rigorously demonstrated in the following. Unfortunately, as remarked earlier, in order to determine the explicit form for the numerator polynomials it is evidently necessary to treat the different canonical subgroup imbeddings separately. Accordingly we shall devote the remainder of this paper to the canonical imbedding $U(n+1)\supset U(n)$ and leave it to the second paper of the series to consider the canonical imbedding $O(n+1)\supset O(n)$.

VI. NOTATION AND FUNDAMENTALS FOR U(n)

In this section we assume that L is the Lie algebra of $U(n+1)$ and L_0 is the Lie algebra of $U(n)$. We recall[8,10] that the $(n+1)^2$ generators a_{ij} $(i,j=1,\ldots,n+1)$ of the Lie group $U(n+1)$ satisfy the commutation relations

$$[a_{ij},a_{kl}] = \delta_{kj}a_{il} - \delta_{il}a_{kj},$$

and are moreover required to satisfy the Hermiticity condition

$$a_{ij}^\dagger = a_{ji}$$

on finite-dimensional (i.e., unitary) representations of the group. We choose as a Cartan subalgebra H for L the space spanned by the $n+1$ commuting Hermitian operators a_{ii} $(i=1,\ldots,n+1)$. The weights $\lambda\in H^*$ may be identified with the $(n+1)$-tuples $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_{n+1})$, where the components λ_i are given by $\lambda_i=\lambda(a_{ii})$. We choose as our system of positive roots Φ^+, the weights $\epsilon_i-\epsilon_j$ $(1\leqslant i<j\leqslant n+1)$, where ϵ_r denotes the fundamental weight with 1 in the rth position and zeros elsewhere. Thus δ, the half-sum of the positive roots, is given by

$$\delta = \frac{1}{2}\sum_{i<j}(\epsilon_i-\epsilon_j) = \frac{1}{2}\sum_{i=1}^{n+1}(n+2-2i)\epsilon_i.$$

In this case the inner product on H^* induced by the Killing form is given by

$$(\lambda,\mu) = \sum_{r=1}^{n+1}\lambda_r\mu_r, \quad \lambda,\mu\in H^*.$$

Also all roots $\alpha\in\Phi^+$ satisfy $(\alpha,\alpha)=2$, whence we may write

$$\langle\lambda,\alpha\rangle = (\lambda,\alpha), \quad \text{for all } \lambda\in H^*, \quad \alpha\in\Phi.$$

The Weyl group W is, of course, the symmetric group S_{n+1} on $n+1$ objects. Thus if $\sigma\in W$ is a permutation of the numbers $1,\ldots,n+1$ then the action of σ on a weight $\lambda\in H^*$ may be written, in our picture, as

$$(\sigma\lambda)_r = \lambda_{\sigma^{-1}(r)}. \tag{65}$$

We remark that Eq. (65) arises because we regard the Weyl group elements σ as permuting the fundamental weights ϵ_r (cf. Ref. 31):

$$\sigma\epsilon_r = \epsilon_{\sigma(r)}.$$

Now the components of a weight λ are given by $\lambda_r=(\lambda,\epsilon_r)$. Thus we must have

$$(\sigma\lambda)_r = (\sigma(\lambda),\epsilon_r) = (\lambda,\sigma^{-1}(\epsilon_r)) = \lambda_{\sigma^{-1}(r)},$$

where we have used the fact that the inner product on H^* is invariant under W.

The finite-dimensional irreducible representations of the Lie group $U(n+1)$ are characterized by their highest weights $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_{n+1})$ whose components are required to satisfy the conditions

$$\lambda_i-\lambda_j\in\mathbb{Z}, \quad i,j=1,\ldots,n+1,$$

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant\lambda_{n+1}.$$

We have, in particular, the elementary dominant weights Λ_r $(r=1,\ldots,n+1)$ defined by

$$\Lambda_r = \sum_{l=1}^{r}\epsilon_l. \tag{66}$$

An alternative characterization of the finite-dimensional irreducible representations is given by the eigenvalues of the $U(n+1)$ Gel'fand invariants

$$I_1 = \sum_{i=1}^{n+1}a_{ii}, \quad I_2 = \sum_{i,j=1}^{n+1}a_{ij}a_{ji},$$

$$I_3 = \sum_{i,j,k=1}^{n+1}a_{ij}a_{jk}a_{ki}, \quad \text{etc.}$$

The operators I_r $(r=1,\ldots,n+1)$ form a full set of invariants for the Lie algebra L that generate the center Z of the enveloping algebra U. Thus we may write

$$Z = \mathbb{C}[I_1,I_2,\ldots,I_{n+1}].$$

We note, in particular, that the eigenvalues of the first- and second-order invariants on the irreducible U-module $V(\lambda)$ with highest weight λ are given by

$$\chi_\lambda(I_1) = \sum_{i=1}^{n+1}\lambda_i,$$

$$\chi_\lambda(I_2) = (\lambda,\lambda+2\delta) = \sum_{i=1}^{n+1}\lambda_i(\lambda_i+n+2-2i).$$

The eigenvalues of the higher-order invariants I_r $(r>2)$ are given in Ref. 36.

The generators a_{ij} of the Lie group $U(n+1)$ fit naturally into a $(n+1)\times(n+1)$ matrix[9–11]

$$b=[a_{ij}],$$

whose (i,j) entry is the generator a_{ij}: the matrix b is a special case of the more general matrices b_λ of Eq. (12), where π_λ, in this case, corresponds to the fundamental contragredient vector representation. Polynomials in the $\mathrm{U}(n+1)$ matrix b then may be defined recursively according to

$$[b^{m+1}]_{ij}=\sum_{k=1}^{n+1}[b^m]_{ik}a_{kj}=\sum_{k=1}^{n+1}a_{ik}[b^m]_{kj}.$$

It can be shown[11,33,37] that the matrix b satisfies a polynomial identity over the center Z of U, which may be written in factorized form as

$$\prod_{r=1}^{n+1}(b-\beta_r)=0, \tag{67}$$

where the β_r are invariants of the group whose eigenvalues on a finite-dimensional irreducible module with highest weight λ are given by

$$\beta_r=\lambda_r+n+1-r. \tag{68}$$

The characteristic roots β_r are clearly a particular case of the more general roots of Eq. (26). The roots β_r determine linear polynomial functions on H^*: $\beta_r(\lambda)=\lambda_r+n+1-r$. The action of the translated Weyl group $\tilde{W}$ on the roots β_r is thus given by [cf. Eq. (27)]

$$\tilde{\sigma}\beta_r=\beta_{\sigma(r)},\quad \sigma\in W. \tag{69}$$

Thus the translated Weyl group acts on the characteristic roots β_r by permuting them among themselves.

The characteristic roots β_r will be employed repeatedly in our work: they play a role equivalent to the partial hooks employed in the work of Biedenharn and Louck.[12] Since the translated Weyl group permutes the β_r one may identify the center Z of U with the ring of symmetric polynomials in the β_r (cf. Harish Chandra's theorem):

$$Z=\mathbb{C}\{\beta_1,\beta_2,\dots,\beta_{n+1}\},$$

where $\mathbb{C}\{x_1,x_2,\dots,x_{n+1}\}$ denotes the symmetric polynomial algebra in $n+1$ indeterminates $x_1,\dots,x_{n+1}$.

With regard to the subgroup $\mathrm{U}(n)$ and its Lie algebra L_0 we follow the same notation as above except we add a subscript $_0$ to everything. Thus δ_0, the half-sum of the positive roots, is given by

$$\delta_0=\frac{1}{2}\sum_{i=1}^{n}(n+1-2i)\epsilon_{0i}.$$

The $\mathrm{U}(n)$ matrix $a=[a_{ij}]$ $(i,j=1,\dots,n)$ satisfies an nth-order polynomial identity analogous to Eq. (67):

$$\prod_{r=1}^{n}(a-\alpha_r)=0.$$

The characteristic roots in this case take a constant value on a finite-dimensional irreducible U_0-module with highest weight $\lambda_0=(\lambda_{01},\dots,\lambda_{0n})$, given by

$$\alpha_r=\lambda_{0r}+n-r. \tag{70}$$

We may also define the fundamental Gel'fand invariants $I_{0m}=\mathrm{tr}(a^m)$ of $\mathrm{U}(n)$ that generate the center Z_0 of the universal enveloping algebra U_0 of L_0. In this case the center Z_0 may be identified with the ring of symmetric polynomials in the characteristic roots α_r:

$$Z_0=\mathbb{C}\{\alpha_1,\dots,\alpha_n\}.$$

Since the imbedding $L\supset L_0$ is canonical, the centralizer of L_0 in U given by

$$\mathscr{Z}=Z\otimes Z_0=Z_0\otimes Z\quad\text{(enveloping algebra product)}.$$

We may clearly identify $\mathscr{Z}$ with the algebra

$$\mathscr{Z}=\mathbb{C}\{\beta_1,\dots,\beta_{n+1}:\alpha_1,\dots,\alpha_n\},$$

where $\mathbb{C}\{x_1,\dots,x_{n+1}:y_1,\dots,y_n\}$ denotes the algebra of all polynomials in indeterminates $x_1,\dots,x_{n+1},y_1,\dots,y_n$, which are symmetric in the x_k and also the y_r.

For example it can be shown[10,11] that the $(n+1,n+1)$ entries of the matrix powers b^m are invariants of $\mathrm{U}(n)$:

$$[b^m]_{n+1,n+1}\in\mathscr{Z}. \tag{71}$$

Thus we may express the above centralizer elements as a polynomial in the $\mathrm{U}(n+1)$ and $\mathrm{U}(n)$ invariants I_k $(k=1,\dots,n+1)$ and I_{0r} $(r=1,\dots,n)$, respectively. Alternatively we may express the invariants of Eq. (71) as a symmetric polynomial in the characteristic roots β_k, α_r, according to[10,11]

$$[b^m]_{n+1,n+1}=\sum_{k=1}^{n+1}\beta_k^m C_k,$$

where

$$C_k=\prod_{\substack{p=1\\ \neq k}}^{n+1}(\beta_k-\beta_p)^{-1}\prod_{r=1}^{n}(\beta_k-\alpha_r-1). \tag{72}$$

It should be noted that the $\mathrm{U}(n+1)$ characteristic roots β_k are invariants of $\mathrm{U}(n+1)$ and hence also of $\mathrm{U}(n)$ as distinct from the $\mathrm{U}(n)$ characteristic roots α_r, which are invariants of $\mathrm{U}(n)$ but not of $\mathrm{U}(n+1)$.

If $V(\lambda)$ denotes a finite-dimensional irreducible U-module with highest weight λ, then it is well known[8] that $V(\lambda)$ decomposes into a direct sum of irreducible U_0-modules according to [cf. Eq. (10)]

$$V(\lambda)=\bigotimes_{\lambda_0\in[\lambda]}V(\lambda|\lambda_0), \tag{73}$$

where the $\mathrm{U}(n)$ highest weights occurring are to be dominant and satisfy the betweenness conditions

$$\lambda_1\geqslant\lambda_{01}\geqslant\lambda_2\geqslant\lambda_{02}\geqslant\cdots\geqslant\lambda_{0n}\geqslant\lambda_{n+1}. \tag{74}$$

Thus, for this case, we call the weight $(\lambda|\lambda_0)$ lexical if and only if λ and λ_0 are dominant and satisfy the betweenness conditions of Eq. (74).

By repeated application of the above result we see that the group $\mathrm{U}(n+1)$ admits the canonical chain of subgroups

$$\mathrm{U}(n+1)\supset\mathrm{U}(n)\supset\cdots\supset\mathrm{U}(2)\supset\mathrm{U}(1), \tag{75}$$

whose representation labels serve to completely label the basis states of the irreducible representations. This is the state labeling scheme proposed by Gel'fand and Tsetlin.[7] The partitions for each of the groups occurring in the chain (75) are most conveniently arranged into a Gel'fand–Tsetlin (GT) pattern, which is described in detail in the work of Baird and Biedenharn.[8] We see, therefore, that the lexical weights $(\lambda|\lambda_0)$ occurring in the decomposition (73) are to form the top two rows of the GT patterns and may be written in the more suggestive form

$$(\lambda\,|\lambda_0) \equiv \begin{pmatrix} \lambda_1 & & \lambda_2 & & \lambda_3 & \cdots & \lambda_n & & \lambda_{n+1} \\ & \lambda_{01} & & \lambda_{02} & & \cdots & & \lambda_{0n} & \end{pmatrix}.$$

VII. OPTIMAL RWC's FOR U(n)

It is our aim here to determine all optimal RWC's for the canonical imbedding $U(n+1)\supset U(n)$. We adopt the notation and conventions of the previous sections. We shall express the numerator polynomial

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}, \quad \Delta\in \mathrm{Sym}(\lambda), \quad \Delta_0\in \mathrm{Sym}(\lambda_0), \tag{76}$$

of Eq. (61) as a polynomial in the characteristic roots β_k and α_r. We note that the numerator polynomial of Eq. (76) is to determine a polynomial function of degree $r_\lambda + r_{\lambda_0}$ [see remarks preceding Eq. (57)], where the integers r_λ and r_{λ_0} in this case are given by

$$\begin{aligned} r_\lambda &= \sum_{\alpha>0}(\lambda,\alpha) = 2(\lambda,\delta) = \sum_{k=1}^{n+1}(n+2-2k)\lambda_k ,\\ r_{\lambda_0} &= \sum_{\alpha_0>0}(\lambda_0,\alpha_0) = 2(\lambda_0,\delta_0) = \sum_{r=1}^{n}(n+1-2r)\lambda_{0r} . \end{aligned} \tag{77}$$

Suppose now that $(\mu|\mu_0)\in\mathscr{L}$ is a lexical weight. It follows in view of Eq. (35) and the betweenness conditions of Eq. (74) that the numerator polynomial of Eq. (76) vanishes if the weight $(\mu+\Delta|\mu_0+\Delta_0)$ is nonlexical: since in such a case the RWC

$$\left\langle \begin{matrix}\mu+\Delta\\ \mu_0+\Delta_0\end{matrix}\middle|\middle| \begin{matrix}\lambda\\ \lambda_0\end{matrix}\ ;\ \begin{matrix}\mu\\ \mu_0\end{matrix}\right\rangle$$

vanishes. Thus we deduce a vanishing contribution whenever the following situations occur:

$$\begin{aligned} &\text{(i)}\ \ (\mu_0+\Delta_0)_r > (\mu+\Delta)_r , \quad \text{some } r=1,\ldots,n ,\\ &\text{(ii)}\ \ (\mu+\Delta)_{r+1} > (\mu_0+\Delta_0)_r , \quad \text{some } r=1,\ldots,n . \end{aligned} \tag{78}$$

In case (i) above we have

$$\mu_{0r}+\Delta_{0r} = \mu_r + \Delta_r + k, \quad \text{for some } k\in\mathbb{Z}^+ .$$

To determine the possible range of k values we note that

$$k = \mu_{0r} - \mu_r + \Delta_{0r} - \Delta_r \leqslant \Delta_{0r} - \Delta_r ,$$

where the last inequality follows because the weight $(\mu|\mu_0)$ is lexical. Thus we can only get a vanishing contribution if $\Delta_{0r}-\Delta_r>0$, in which case we deduce that the possible range of k-values is given by $1\leqslant k\leqslant \Delta_{0r}-\Delta_r$. In such a case we deduce that the numerator polynomial of Eq. (76) is to be divisible by factors

$$\begin{aligned} &[(\mu_0+\Delta_0)_r - (\mu+\Delta)_r - k]\\ &\quad = (\alpha_r - \beta_r + \Delta_{0r} - \Delta_r - k + 1),\\ &\quad k = 1,\ldots,\Delta_{0r}-\Delta_r \end{aligned}$$

or equivalently by factors

$$(\alpha_r - \beta_r + m), \quad m = 1,\ldots,\Delta_{0r}-\Delta_r .$$

Thus we deduce divisibility of the numerator polynomial (76) by the set of factors

$$\prod_{m=1}^{\Delta_{0r}-\Delta_r}(\alpha_r-\beta_r+m), \quad \Delta_{0r}>\Delta_r .$$

Applying a similar argument to the conjugate numerator polynomial

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \sigma_0(\Delta_0)\end{pmatrix}, \quad \sigma\in W, \quad \sigma_0\in W_0 ,$$

we deduce divisibility of this latter polynomial by factors

$$\prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r}(\alpha_r-\beta_r+m), \quad \sigma_0(\Delta_0)_r>\sigma(\Delta)_r$$

[noting, in view of Eq. (65), that $\sigma(\Delta)_r = \Delta_{\sigma^{-1}(r)}$]. But in view of the Weyl group symmetries of Eq. (64) we may write

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = \tilde\sigma^{-1}\tilde\sigma_0^{-1}\eta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \sigma_0(\Delta_0)\end{pmatrix}.$$

This would then imply [cf. Eqs. (8) and (69)] that the numerator polynomial (76) is divisible by the further set of factors

$$\begin{aligned} &\tilde\sigma^{-1}\tilde\sigma_0^{-1}\prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r}(\alpha_r-\beta_r+m)\\ &\quad = \prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r}(\alpha_{\sigma_0^{-1}(r)}-\beta_{\sigma^{-1}(r)}+m),\\ &\quad \sigma_0(\Delta_0)_r>\sigma(\Delta)_r . \end{aligned}$$

Taking into account arbitrary permutations $\sigma\in W$, $\sigma_0\in W_0$, we thus deduce divisibility of the numerator polynomial by the set of factors

$$\prod_{k=1}^{n+1}\prod_{\substack{r=1\\ \Delta_{0r}>\Delta_k}}^{n}\prod_{m=1}^{\Delta_{0r}-\Delta_k}(\alpha_r-\beta_k+m). \tag{79}$$

In case (ii) of Eq. (78) we deduce a vanishing contribution whenever $\Delta_{r+1}>\Delta_{0r}$ and

$$(\mu+\Delta)_{r+1} = (\mu_0+\Delta_0)_r + k, \quad k\in\mathbb{Z}^+ .$$

The range of possible k-values in this case is given by $1\leqslant k\leqslant \Delta_{r+1}-\Delta_{0r}$. Thus we deduce divisibility of the numerator polynomial (76) by the set of factors

$$\begin{aligned} &(\mu+\Delta)_{r+1} - (\mu_0+\Delta_0)_r - m\\ &\quad = \beta_{r+1} - \alpha_r + \Delta_{r+1} - \Delta_{0r} - m ,\\ &\quad m = 1,\ldots,\Delta_{r+1}-\Delta_{0r} , \end{aligned}$$

or equivalently, the factors

$$\beta_{r+1} - \alpha_r + m - 1, \quad m = 1,\ldots,\Delta_{r+1}-\Delta_{0r} .$$

Thus we deduce divisibility of the numerator polynomial (76) by the set of factors

$$\prod_{m=1}^{\Delta_{r+1}-\Delta_{0r}}(\beta_{r+1}-\alpha_r+m-1), \quad \Delta_{r+1}>\Delta_{0r} .$$

Applying Weyl group symmetry as before and considering all possible permutations of the $U(n+1)$ roots β_k and

U(n) roots α_r we deduce divisibility of the numerator polynomial by the additional set of factors

$$\prod_{k=1}^{n+1}\prod_{\substack{r=1\\ \Delta_k>\Delta_{0r}}}^{n}\prod_{m=1}^{\Delta_k-\Delta_{0r}}(\beta_k-\alpha_r+m-1)\,. \tag{80}$$

It is our aim now to demonstrate that the numerator polynomial of Eq. (76) is given by the products (79) and (80): i.e.,

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\prod_{k=1}^{n+1}\prod_{\substack{r=1\\ \Delta_{0r}>\Delta_k}}^{n}\prod_{m=1}^{\Delta_{0r}-\Delta_k}(\alpha_r-\beta_k+m)$$

$$\times\prod_{k=1}^{n+1}\prod_{\substack{r=1\\ \Delta_k>\Delta_{0r}}}^{n}\prod_{m=1}^{\Delta_k-\Delta_{0r}}(\beta_k-\alpha_r+m-1)\,,$$

$$\Delta\in\mathrm{Sym}(\lambda)\,,\quad \Delta_0\in\mathrm{Sym}(\lambda_0)\,. \tag{81}$$

We note that Eq. (81) determines a monic polynomial function that satisfies the Weyl group symmetries of Eq. (64) as required. It remains to demonstrate that the total number of factors on the rhs of Eq. (81) is equal to the degree of the polynomial function (76): it is given by the integer $r_\lambda+r_{\lambda_0}$ with r_λ and r_{λ_0} as in Eq. (77).

In view of Weyl group symmetry it suffices to consider the semimaximal case: i.e., $\Delta=\lambda$, $\Delta_0=\lambda_0$ in Eq. (81). In such a case we see that the total number of factors on the rhs of (81) is given by

$$\begin{aligned}N&=\sum_{\substack{k=1\\ \lambda_{0r}>\lambda_k}}^{n+1}\sum_{r=1}^{n}(\lambda_{0r}-\lambda_k)+\sum_{\substack{k=1\\ \lambda_k>\lambda_{0r}}}^{n+1}\sum_{r=1}^{n}(\lambda_k-\lambda_{0r})\\
&=\sum_{r=1}^{n}\left\{\sum_{\substack{k=1\\ \lambda_k<\lambda_{0r}}}^{n+1}-\sum_{\substack{k=1\\ \lambda_k>\lambda_{0r}}}^{n+1}\right\}\lambda_{0r}+\sum_{k=1}^{n+1}\left\{-\sum_{\substack{r=1\\ \lambda_{0r}>\lambda_k}}^{n}+\sum_{\substack{r=1\\ \lambda_{0r}<\lambda_k}}^{n}\right\}\lambda_k\\
&=\sum_{r=1}^{n}(n+1-2r)\lambda_{0r}+\sum_{k=1}^{n+1}(n+2-2k)\lambda_k\\
&=r_{\lambda_0}+r_\lambda\,,\end{aligned}$$

as required, where we have applied the betweenness conditions (74). Thus we have proved our assertion that the numerator polynomials of Eq. (76) are given by Eq. (81). It is easily checked that our results agree with those obtained previously by Biedenharn and Louck.[12]

We remark that the numerator polynomial of Eq. (81) is a product of linear factors that involve the representation labels of both groups U$(n+1)$ and U(n) as distinct from the denominator polynomial of Eq. (62), which is a product of linear factors each of which involve only the representation labels of U$(n+1)$ or those of the group U(n). We may express the denominator polynomial of Eq. (62) in terms of the characteristic roots β_k and α_r by noting that if $\alpha\in\Phi$, $\alpha_0\in\Phi_0$, then

$$(\mu+\delta,\alpha)=\beta_i-\beta_j\,,\quad \alpha=\epsilon_i-\epsilon_j\,,$$

$$(\mu_0+\delta_0,\alpha_0)=\alpha_i-\alpha_j\,,\quad \alpha_0=\epsilon_{0i}-\epsilon_{0j}\,.$$

Substituting this into Eq. (62) we obtain the following expressions for our denominator polynomials:

$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix},$$

$$\Delta\in\mathrm{Sym}(\lambda)\,,\quad \Delta_0\in\mathrm{Sym}(\lambda_0)\,, \tag{82a}$$

where

$$\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}=\prod_{\substack{r,k=1\\ \Delta_r>\Delta_k}}^{n+1}\prod_{m=1}^{\Delta_r-\Delta_k}(\beta_r-\beta_k+m-1)\,, \tag{82b}$$

$$\delta_2\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}=\prod_{\substack{r,k=1\\ \Delta_{0r}>\Delta_{0k}}}^{n}\prod_{m=1}^{\Delta_{0r}-\Delta_{0k}}(\alpha_r-\alpha_k+m)\,. \tag{82c}$$

Thus our optimal RWF's for the imbedding U$(n+1)$ $\supset$U(n) are given by the rational polynomial functions

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=C_{\lambda,\lambda_0}\frac{\eta\binom{\lambda\ \lambda_0}{\Delta\ \Delta_0}(\mu|\mu_0)}{\delta\binom{\lambda\ \lambda_0}{\Delta\ \Delta_0}(\mu|\mu_0)}, \tag{83}$$

where the numerator and denominator polynomials are given by Eqs. (81) and (82), respectively, where it is understood that the characteristic roots β_k and α_r are given by [cf. Eqs. (68) and (70)]

$$\beta_k=\mu_k+n+1-k\,,\quad \alpha_r=\mu_{0r}+n-r\,. \tag{84}$$

We have in particular for the semimaximal case [i.e., $\Delta=\lambda$, $\Delta_0=\lambda_0$ in Eq. (83)] the results

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}=\prod_{r<k}^{n+1}\prod_{m=1}^{\lambda_{0r}-\lambda_k}(\alpha_r-\beta_k+m)$$

$$\times\prod_{k<r}^{n}\prod_{m=1}^{\lambda_k-\lambda_{0r}}(\beta_k-\alpha_r+m-1)\,, \tag{85a}$$

$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}=\delta_1\begin{pmatrix}\lambda\\ \lambda\end{pmatrix}\delta_2\begin{pmatrix}\lambda_0\\ \lambda_0\end{pmatrix},$$

where

$$\delta_1\begin{pmatrix}\lambda\\ \lambda\end{pmatrix}=\prod_{r<k}^{n+1}\prod_{m=1}^{\lambda_r-\lambda_k}(\beta_r-\beta_k+m-1)\,, \tag{85b}$$

$$\delta_2\begin{pmatrix}\lambda_0\\ \lambda_0\end{pmatrix}=\prod_{k<r}^{n}\prod_{m=1}^{\lambda_{0k}-\lambda_{0r}}(\alpha_k-\alpha_r+m)\,. \tag{85c}$$

It remains now to determine the numerical constant C_{λ,λ_0} of Eq. (83), which depends only on the labels λ,λ_0 (and is independent of the shifts Δ,Δ_0). To this end we note that Eq. (47) implies the result

$$1=\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}(0|0)=C_{\lambda,\lambda_0}\left.\frac{\eta\binom{\lambda\ \lambda_0}{\lambda\ \lambda_0}}{\delta\binom{\lambda\ \lambda_0}{\lambda\ \lambda_0}}\right|_{(\mu|\mu_0)=(0|0)},$$

whence we obtain, for the numerical constant C_{λ,λ_0},

$$C_{\lambda,\lambda_0}=\left.\frac{\delta\binom{\lambda\ \lambda_0}{\lambda\ \lambda_0}}{\eta\binom{\lambda\ \lambda_0}{\lambda\ \lambda_0}}\right|_{(\mu|\mu_0)=(0|0)}. \tag{86a}$$

Substituting $\mu_k=\mu_{0r}=0$ into Eqs. (84) and (85), we obtain the result

$$C_{\lambda,\lambda_0}=\prod_{r<k}^{n+1}\frac{(\lambda_r-\lambda_k+k-r-1)!}{(\lambda_{0r}-\lambda_k+k-r-1)!}$$

$$\times\prod_{r<k}^{n}\frac{(\lambda_{0r}-\lambda_{0k}+k-r)!}{(\lambda_r-\lambda_{0k}+k-r)!}, \tag{86b}$$

which agrees with Ref. 12 as required.

VIII. SOME EXAMPLES

We consider here some examples of our previous formulas. We shall also point out the connection between the problem of evaluating RWC's and the equivalent (algebraic) problem of evaluating the eigenvalues of certain $U(n)$-Casimir invariants.

(i) Suppose $\lambda = \epsilon_1$ is the highest weight for the fundamental vector representation of $U(n+1)$ and $\lambda_0 = 0$ [corresponding to the trivial representation of $U(n)$]. The optimal RWF's $\rho(\begin{smallmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{smallmatrix})$ $(k=1,...,n+1)$, denoted $\overline{C}_k$ in Ref. 10, in this case are given by the formula

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{pmatrix} = \overline{C}_k = \prod_{p\neq k}^{n+1} (\beta_k - \beta_p)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r),$$

which agrees with Ref. 10, Eq. (13).

(ii) Suppose $\lambda = -\epsilon_{n+1}$ is the highest weight for the contragredient vector representation and $\lambda_0 = 0$. The optimal RWF's $\rho(\begin{smallmatrix}-\epsilon_{n+1} & 0\\ -\epsilon_k & 0\end{smallmatrix})$, denoted C_k in Ref. 10, in this case reduce to [cf. Eq. (72)]

$$\rho\begin{pmatrix}-\epsilon_{n+1} & 0\\ -\epsilon_k & 0\end{pmatrix} = C_k = \prod_{p\neq k}^{n+1} (\beta_k - \beta_p)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r - 1).$$

It is interesting to see the connection between the above RWF's and the eigenvalues of the centralizer elements of Eq. (71). This connection, in fact, extends to the general case: the eigenvalues of the centralizer elements $I_{m,n}(\lambda|\lambda_0)$ of Eq. (16) are determined by the $L{:}\,L_0$ RWF's (and vice versa).

(iii) Suppose $\lambda = \epsilon_1$ and $\lambda_0 = \epsilon_{01}$ [corresponding to the highest weights of the fundamental vector representation of $U(n+1)$ and $U(n)$, respectively]. We then have the RWF's $\rho(\begin{smallmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{smallmatrix})$, herein denoted $\bar{\rho}(\begin{smallmatrix}n+1 & n\\ k & r\end{smallmatrix})$ for simplicity. In this case we obtain immediately the result

$$\bar{\rho}\begin{pmatrix}n+1 & n\\ k & r\end{pmatrix} = (-1)^n \prod_{p\neq k}^{n+1} \left(\frac{\beta_p - \alpha_r - 1}{\beta_k - \beta_p}\right) \times \prod_{l\neq r}^{n} \left(\frac{\beta_k - \alpha_l}{\alpha_r - \alpha_l + 1}\right).$$

In the notation of Ref. 10 this formula may be alternatively written [cf. Ref. 10, Eq. (20)]

$$\bar{\rho}\begin{pmatrix}n+1 & n\\ k & r\end{pmatrix} = \overline{C}_k M_r (\beta_k - \alpha_r - 1)^{-1} (\beta_k - \alpha_r)^{-1},$$

where

$$M_r = (-1)^n \prod_{l\neq r}^{n} (\alpha_r - \alpha_l + 1)^{-1} \prod_{p=1}^{n+1} (\beta_p - \alpha_r - 1).$$

This last equation determines the squared $U(n)$-reduced matrix elements of the $U(n)$ vector operator

$$\psi_i = a_{i\,n+1} \quad (i = 1,...,n).$$

(iv) Suppose $\lambda = -\epsilon_{n+1}$, $\lambda_0 = -\epsilon_{0n}$. Then we have the RWF's $\rho(\begin{smallmatrix}-\epsilon_{n+1} & -\epsilon_{0n}\\ -\epsilon_k & -\epsilon_{0r}\end{smallmatrix})$, herein denoted $\rho(\begin{smallmatrix}n+1 & n\\ k & r\end{smallmatrix})$, which are given by [cf. Ref. 10, Eq. (19)]

$$\rho\begin{pmatrix}n+1 & n\\ k & r\end{pmatrix} = C_k \overline{M}_r (\beta_k - \alpha_r - 1)^{-1} (\beta_k - \alpha_r)^{-1},$$

where

$$\overline{M}_r = (-1)^n \prod_{l\neq r}^{n} (\alpha_r - \alpha_l - 1)^{-1} \prod_{p=1}^{n+1} (\beta_p - \alpha_r).$$

(v) Consider the following $U(n)$-invariants:

$$I_{m,l} = \tau_0(b^m a^l) = \sum_{i,j=1}^{n} (b^m)_{ij} (a^l)_{ji} \in \mathscr{Z},$$

which are a special case of the more general invariants of Eq. (16). These invariants take a constant value on the irreducible $U(n)$-module $V(\mu|\mu_0) \subseteq V(\mu)$, this eigenvalue being given by the formula

$$\chi_{(\mu|\mu_0)}[I_{m,l}] = \sum_{k=1}^{n+1} \sum_{r=1}^{n} \beta_k^m \alpha_r^l \times \prod_{p\neq r}^{n} \left(\frac{\alpha_r - \alpha_p - 1}{\alpha_r - \alpha_p}\right) \rho\begin{pmatrix}n+1 & n\\ k & r\end{pmatrix} (\mu|\mu_0), \tag{87a}$$

with $\rho(\begin{smallmatrix}n+1 & n\\ k & r\end{smallmatrix})$ as in example (iv). Using the formula of example (iv) we may alternatively write

$$\chi_{(\mu|\mu_0)}[I_{m,l}] = (-1)^n \sum_{k=1}^{n+1} \sum_{r=1}^{n} \beta_k^m \alpha_r^l \prod_{p\neq k}^{n+1} \left(\frac{\beta_p - \alpha_r}{\beta_k - \beta_p}\right) \times \prod_{q\neq r}^{n} \left(\frac{\beta_k - \alpha_q - 1}{\alpha_r - \alpha_q}\right). \tag{87b}$$

The above result follows by considering the $U(n+1)$ and $U(n)$ projection operators (cf. Refs. 11 and 36)

$$P[k] = \prod_{p\neq k}^{n+1} \left(\frac{b - \beta_p}{\beta_k - \beta_p}\right), \quad P_0[r] = \prod_{l\neq r}^{n} \left(\frac{a - \alpha_l}{\alpha_r - \alpha_l}\right).$$

It follows from the $U(n+1)$ and $U(n)$ characteristic identities that we have the following resolutions[11]:

$$b^m = \sum_{k=1}^{n+1} \beta_k^m P[k], \quad a^l = \sum_{r=1}^{n} \alpha_r^l P_0[r].$$

Thus we have

$$b^m a^l = \sum_{k=1}^{n+1} \sum_{r=1}^{n} \beta_k^m \alpha_r^l P[k] P_0[r],$$

whence, taking the $U(n)$-trace, we obtain

$$I_{m,l} = \sum_{k=1}^{n+1} \sum_{r=1}^{n} \beta_k^m \alpha_r^l \tau_0(P[k] P_0[r]) = \sum_{k=1}^{n+1} \sum_{r=1}^{n} \beta_k^m \alpha_r^l \frac{\tau(\begin{smallmatrix}-\epsilon_{n+1} & -\epsilon_{0n}\\ -\epsilon_k & -\epsilon_{0r}\end{smallmatrix})}{\xi(\begin{smallmatrix}-\epsilon_{n+1} & -\epsilon_{0n}\\ -\epsilon_k & -\epsilon_{0r}\end{smallmatrix})},$$

where the polynomial functions τ and ξ are given by Eqs. (31) and (33), respectively. In view of Eq. (36), Eq. (87) is easily seen to follow.

(vi) Consider now the case $\lambda = \Lambda_k$ [see Eq. (66)], which is the highest weight for the kth-rank antisymmetric tensor representation of $U(n+1)$. The allowed $U(n)$ highest weights occurring are $\lambda_0 = \Lambda_{0k}$ $(1 \leqslant k < n+1)$ and $\lambda_0 = \Lambda_{0k-1}$. We consider first the case $\lambda_0 = \Lambda_{0k}$ $(k < n+1)$. We obtain immediately from Eqs. (83)–(85), the following result for the semimaximal RWF

$$\rho\begin{pmatrix}\Lambda_k & \Lambda_{0k}\\ \Lambda_k & \Lambda_{0k}\end{pmatrix} = \prod_{i=1}^{k}\left\{\prod_{j=k+1}^{n}\left(\frac{\beta_i-\alpha_j}{\alpha_i-\alpha_j+1}\right)\prod_{j=k+1}^{n+1}\left(\frac{\alpha_i-\beta_j+1}{\beta_i-\beta_j}\right)\right\}.$$

Applying Weyl group symmetry we obtain the general result

$$\rho\begin{pmatrix}\Lambda_k & \Lambda_{0k}\\ \Delta & \Delta_0\end{pmatrix} = \prod_{i=1}^{k}\left\{\prod_{j=k+1}^{n}\left(\frac{\beta_{r_i}-\alpha_{l_j}}{\alpha_{l_i}-\alpha_{l_j}+1}\right) \times\prod_{j=k+1}^{n}\left(\frac{\alpha_{l_i}-\beta_{r_j}+1}{\beta_{r_i}-\beta_{r_j}}\right)\right\}, \tag{88}$$

where $\sigma=(r_1,\ldots,r_{n+1})$ [resp. $\sigma_0=(l_1,\ldots,l_n)$] are the permutations such that $\Delta=\sigma(\Lambda_k)$ [resp. $\Delta_0=\sigma_0(\Lambda_{0k})$].

We consider now the case $\lambda_0=\Lambda_{0k-1}$ for which we obtain, for the semimaximal RWF

$$\rho\begin{pmatrix}\Lambda_k & \Lambda_{0k-1}\\ \Lambda_k & \Lambda_{0k-1}\end{pmatrix} = \prod_{i=1}^{k}\prod_{j=k}^{n}\left(\frac{\beta_i-\alpha_j}{\beta_i-\beta_{j+1}}\right) \times\prod_{i=1}^{k-1}\prod_{j=k+1}^{n+1}\left(\frac{\alpha_i-\beta_j+1}{\alpha_i-\alpha_{j-1}+1}\right).$$

More generally we obtain

$$\rho\begin{pmatrix}\Lambda_k & \Lambda_{0k-1}\\ \Delta & \Delta_0\end{pmatrix} = \prod_{i=1}^{k}\prod_{j=k}^{n}\left(\frac{\beta_{r_i}-\alpha_{l_j}}{\beta_{r_i}-\beta_{r_{j+1}}}\right) \times\prod_{i=1}^{k-1}\prod_{j=k+1}^{n+1}\left(\frac{\alpha_{l_i}-\beta_{r_j}+1}{\alpha_{l_i}-\alpha_{l_{j-1}}+1}\right), \tag{89}$$

where $\sigma=(r_1,\ldots,r_{n+1})$ [resp. $\sigma_0=(l_1,\ldots,l_n)$] are the permutations such that $\Delta=\sigma(\Lambda_k)$ [resp. $\Delta_0=\sigma_0(\Lambda_{0k-1})$].

The RWF's of Eqs. (88) and (89) are called *elementary* by Biedenharn and Louck.[12] Since the weights for the elementary representations $V(\Lambda_k)$ are all Weyl group conjugate to the highest weight Λ_k, Eqs. (88) and (89) enable all RWC's, and hence all Wigner coefficients, to be calculated for the antisymmetric tensor representations. Such Wigner coefficients are useful since all $U(n)$ Wigner coefficients may be obtained, at least in principle, from the elementary Wigner coefficients. From the point of view of applications, it may be demonstrated[27] that formulas (88) and (89) enable the direct evaluation of all $U(2n)\downarrow U(n)\times U(2)$ subduction coefficients for doing spin-dependent calculations in many-electron problems (cf. Harter and Patterson[28]).

IX. EXTREMAL RWC's FOR U(*n*)

The operation of composition of two RWF's was defined, quite generally, in Sec. IV. However, as noted in Sec. IV, the composition of two RWF's is not necessarily a RWF even if the composed RWF's are commuting (i.e., of the same tie structure). However, it was pointed out by Biedenharn and Louck[12] that there is a class of RWF's for the unitary groups, herein called *extremal*, such that the composition of two extremal RWF's of the same extremal tie structure yields another extremal RWF of the same tie structure. This fact forms the basis for the pattern calculus of Ref. 12 enabling a recursive evaluation of all optimal RWC's for $U(n)$. It is our aim here to present a detailed derivation of these interesting results keeping in mind extensions to more general imbeddings.

We begin by defining a lexical weight $(\lambda\,|\lambda_0)$ to be *extremal* if the $U(n+1)$ weight λ' defined by

$$\lambda'_i=\lambda_{0i},\quad i=1,\ldots,n, \qquad \lambda'_{n+1}=\sum_{i=1}^{n+1}\lambda_i-\sum_{i=1}^{n}\lambda_{0i} \tag{90a}$$

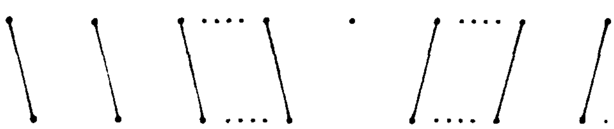
is Weyl group conjugate to the maximal weight λ. Note that the weight λ' is simply the weight of the semimaximal state

$$v_+^{(\lambda|\lambda_0)}=\left|\begin{matrix}\lambda\\ \lambda_0\\ [\mathrm{Max}]\end{matrix}\right\rangle. \tag{90b}$$

In order that the weight λ' of Eq. (90a) be W-conjugate to the highest weight λ, it is easily seen that we must have

$$\lambda_{0r}=\lambda_r \quad\text{or}\quad \lambda_{r+1}\,.$$

Another way of phrasing this is to say that the weight $(\lambda\,|\lambda_0)$ is extremal if the $U(n)$ weight λ_0 is *extremally connected* to λ. This may be visualized by writing out two rows of dots, the top row having $(n+1)$ dots and the bottom row having n dots as shown below:



(91)

We then draw a line from the rth dot of the bottom row to dot r (resp. $r+1$) of the top row according to whether $\lambda_{0r}=\lambda_r$ (resp. λ_{r+1}). It is easily seen[12] that the diagram may be drawn so that no two lines intersect, leading to a diagram of the type shown in Eq. (91): these diagrams are referred to as *extremal tie patterns* in Ref. 12. Thus associated with every extremal weight $(\lambda\,|\lambda_0)$ is an extremal tie pattern (91).

We say that two extremal weights $(\lambda\,|\lambda_0)$ and $(\nu|\nu_0)$ are *linked* if they give rise to the same extremal tie pattern. We say that an optimal RWF

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} \tag{92}$$

is extremal if the lexical weight $(\lambda\,|\lambda_0)$ is extremal. We say that an extremal RWF $\rho\binom{\lambda'\ \lambda_0'}{\lambda'\ \lambda_0'}$ is *extremally tied* to (or has the same extremal tie structure as) the RWF (92) if they are tied (in the sense of Sec. IV) and the weights $(\lambda|\lambda_0)$ and $(\lambda'|\lambda_0')$ are extremal and linked.

The significance of extremal RWF's is that the composition of two extremal RWF's of the same extremal tie structure yields another RWF of the same extremal tie structure. To clarify the situation let v^λ denote the unique (normalized) maximal weight state of the $U(n+1)$ module $V(\lambda)$. For $\sigma\in W$ we let v_σ^λ denote the unique (normalized) weight state of $V(\lambda)$ of weight $\sigma(\lambda)\in S(\lambda)$. If v^λ, v^μ are two $U(n+1)$ maximal weight states we may write (modulo a phase)

$$v^\lambda\otimes v^\mu = v^{\lambda+\mu}.$$

More generally, for $\sigma\in W$ arbitrary, we may write (up to a phase)

$$v_\sigma^\lambda\otimes v_\sigma^\mu = v_\sigma^{\lambda+\mu}. \tag{93}$$

This follows by noting that the weight $\sigma(\lambda+\mu)$ is W-conjugate to the highest weight $\lambda+\mu$ and hence occurs with unit multiplicity in $V(\lambda)\otimes V(\mu)$.

Now we call a permutation $\sigma\in W$ an *extremal* Weyl group element if the numbers $\sigma^{-1}(1), \sigma^{-1}(2),\dots,\sigma^{-1}(n)$ are in ascending order (i.e., σ^{-1} preserves lexicality of the first n components). The significance of extremal Weyl group elements σ lies in the fact that if $\lambda\in\Lambda^+$, then the $U(n)$ weight λ_0, defined by

$$\lambda_{0r} = (\sigma\lambda)_r = \lambda_{\sigma^{-1}(r)}, \quad r=1,\dots,n, \tag{94}$$

is extremally connected to λ: i.e., the weight $(\lambda|\lambda_0)$ is extremal. Thus if σ is an extremal Weyl group element it follows, from uniqueness of the weight $\sigma(\lambda)$ in $V(\lambda)$, that the state v_σ^λ corresponds to the semimaximal state $v_+^{(\lambda|\lambda_0)}$ of Eq. (90b) [with λ_0 as in Eq. (94)].

Thus in the special case that σ is an extremal Weyl group element, Eq. (93) yields the result

$$v_+^{(\lambda|\lambda_0)}\otimes v_+^{(\nu|\nu_0)} = v_+^{(\lambda+\nu|\lambda_0+\nu_0)},$$

provided $(\lambda|\lambda_0)$ and $(\nu|\nu_0)$ are extremal weights that are linked [i.e., give rise to the same tie pattern (91)]. In such a case it follows immediately, in view of Eq. (48), that the following composition law holds:

$$\begin{aligned}\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix}(\mu|\mu_0) &= \left|\left\langle v_+^{(\lambda+\nu+\mu|\lambda_0+\nu_0+\mu_0)}\middle| v_+^{(\lambda+\nu|\lambda_0+\nu_0)}\otimes v_+^{(\mu|\mu_0)}\right\rangle\right|^2\\ &= \rho\begin{pmatrix}\lambda+\nu & \lambda_0+\nu_0\\ \lambda+\nu & \lambda_0+\nu_0\end{pmatrix}(\mu|\mu_0).\end{aligned}$$

Since this is to hold for all $(\mu|\mu_0)\in\Lambda^+\times\Lambda_0^+$ we thus have the result

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \nu & \nu_0\end{pmatrix} = \rho\begin{pmatrix}\lambda+\nu & \lambda_0+\nu_0\\ \lambda+\nu & \lambda_0+\nu_0\end{pmatrix}. \tag{95a}$$

Applying Weyl group symmetry we obtain the result

$$\begin{aligned}\rho\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\lambda) & \sigma_0(\lambda_0)\end{pmatrix}\circ\rho\begin{pmatrix}\nu & \nu_0\\ \sigma(\nu) & \sigma_0(\nu_0)\end{pmatrix}\\ = \rho\begin{pmatrix}\lambda+\nu & \lambda_0+\nu_0\\ \sigma(\lambda+\nu) & \sigma_0(\lambda_0+\nu_0)\end{pmatrix},\end{aligned} \tag{95b}$$

thus proving our assertion that two extremal RWF's of the same tie structure yields, under composition, another RWF of the same extremal tie structure.

Clearly all maximal RWF's are extremal and of the same tie structure. We note also that all elementary RWF's [see example (vi) of Sec. VIII] are extremal. Moreover the elementary RWF's $\rho\binom{\Lambda_k\ \Lambda_{0k-1}}{\Lambda_k\ \Lambda_{0k-1}}$ are extremal and of the same tie structure $(k=1,\dots,n+1)$. The remaining elementary RWF's $\rho\binom{\Lambda_k\ \Lambda_{0k}}{\Lambda_k\ \Lambda_{0k}}$ $(k=1,\dots,n)$ are, of course, maximal and hence extremal of the same tie structure.

Now let $(\lambda|\lambda)$ be a maximally connected lexical weight: i.e., the $U(n)$ representation labels λ_0 take their maximum allowed values $\lambda_{0r}=\lambda_r$ $(r=1,\dots,n)$. We clearly have the following expansion for the maximal lexical weight $(\lambda|\lambda)$:

$$(\lambda|\lambda) = \lambda_{n+1}(\Lambda_{n+1}|\Lambda_{0n}) + \sum_{r=1}^n(\lambda_r-\lambda_{r+1})(\Lambda_r|\Lambda_{0r}).$$

We note that all the lexical weights occurring in the above decomposition are maximal and hence extremal and tied. Thus by repeated application of Eq. (95) [see also Eq. (51)] we obtain the following expansion for the maximal RWF's:

$$\rho\begin{pmatrix}\lambda & \lambda\\ \lambda & \lambda\end{pmatrix} = \prod_{r=1}^n \rho\begin{pmatrix}\Lambda_r & \Lambda_{0r}\\ \Lambda_r & \Lambda_{0r}\end{pmatrix}^{\lambda_r-\lambda_{r+1}}, \tag{96}$$

where we have used the result that the RWF's $\rho\binom{\Lambda_{n+1}\ \Lambda_{0n}}{\Lambda_{n+1}\ \Lambda_{0n}}$ are trivially unity. We note also that the ordering of the factors in the product (96) (shorthand notation for composition) is irrelevant since all RWF's involved are commuting. Note that Eq. (96) implies that for the maximal case $\lambda_0=\lambda$ the constant C_{λ,λ_0} of Eq. (83) is unity, as one may verify directly from Eq. (86).

We draw special attention to the maximal RWF's because Eq. (96) has a suitable extension to the orthogonal groups.

We call a lexical weight $(\lambda|\lambda_0)$ *minimally connected* if the $U(n)$ representation labels λ_0 take their minimum possible values: $\lambda_{0r}=\lambda_{r+1}$ $(r=1,\dots,n)$. We denote a general minimally connected lexical weight by the symbol $(\lambda|\bar\lambda)$. Clearly all minimally connected weights are extremal and of the same tie structure. We note that all elementary lexical weights $(\Lambda_r|\Lambda_{0r-1})$ are minimally connected. We have the following expansion for the minimally connected weights:

$$\begin{aligned}(\lambda|\bar\lambda) &= \lambda_{n+1}(\Lambda_{n+1}|\Lambda_{0n})\\ &\quad+\sum_{r=1}^n(\lambda_r-\lambda_{r+1})(\Lambda_r|\Lambda_{0r-1}).\end{aligned}$$

Thus, by repeated application of Eq. (95), we obtain the result

$$\rho\begin{pmatrix}\lambda & \bar\lambda\\ \lambda & \bar\lambda\end{pmatrix} = \prod_{r=1}^n\rho\begin{pmatrix}\Lambda_r & \Lambda_{0r-1}\\ \Lambda_r & \Lambda_{0r-1}\end{pmatrix}^{\lambda_r-\lambda_{r+1}}. \tag{97}$$

We remark that since all RWF's in the product of Eq. (97) are semimaximal they are all commuting and hence independent of ordering.

We now note that the elementary lexical weights $(\Lambda_k|\Lambda_{0k})$, $(\Lambda_k|\Lambda_{0k-1})$ $(k=1,\dots,n)$, $(\Lambda_{n+1}|\Lambda_{0n})$ form a basis for $H^*\times H_0^*$ with corresponding dual basis $(-\epsilon_{k+1}|\epsilon_{0k})$, $(\epsilon_k|-\epsilon_{0k})$ $(k=1,\dots,n)$, $(\epsilon_{n+1}|\circ)$, respectively. Thus an arbitrary lexical weight $(\lambda|\lambda_0)$ may be expanded

$$
\begin{aligned}
(\lambda|\lambda_0) = {} & \lambda_{n+1}(\Lambda_{n+1}|\Lambda_{0n}) \\
& + \sum_{r=1}^{n} [(\lambda|\lambda_0)\cdot(-\epsilon_{r+1}|\epsilon_{0r})](\Lambda_r|\Lambda_{0r}) \\
& + \sum_{r=1}^{n} [(\lambda|\lambda_0)\cdot(\epsilon_r| - \epsilon_{0r})](\Lambda_r|\Lambda_{0r-1}) \\
= {} & \lambda_{n+1}(\Lambda_{n+1}|\Lambda_{0n}) + \sum_{r=1}^{n} (\lambda_{0r} - \lambda_{r+1})(\Lambda_r|\Lambda_{0r}) \\
& + \sum_{r=1}^{n} (\lambda_r - \lambda_{0r})(\Lambda_r|\Lambda_{0r-1}), \qquad (98a)
\end{aligned}
$$

which expresses $(\lambda|\lambda_0)$ as a sum of a maximally and minimally connected weight:

$$(\lambda|\lambda_0) = (\lambda'|\lambda') + (\lambda''|\bar{\lambda}''),$$

where

$$
\begin{aligned}
\lambda' &= \sum_{r=1}^{n} (\lambda_{0r} - \lambda_{r+1})\Lambda_r, \\
\lambda'' &= \lambda_{n+1}\Lambda_{n+1} + \sum_{r=1}^{n} (\lambda_r - \lambda_{0r})\Lambda_r.
\end{aligned} \qquad (98b)
$$

It then follows from Eq. (50), that we may write

$$C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''}\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix} = \rho\begin{pmatrix}\lambda' & \lambda'\\ \lambda' & \lambda'\end{pmatrix}\circ\rho\begin{pmatrix}\lambda'' & \bar{\lambda}''\\ \lambda'' & \bar{\lambda}''\end{pmatrix}, \qquad (99a)$$

where the constant $C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''}$ is given by

$$C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''} = \rho\begin{pmatrix}\lambda' & \lambda'\\ \lambda' & \lambda'\end{pmatrix}(\lambda''|\bar{\lambda}'').$$

It is straightforwardly demonstrated that this constant is related to the constant C_{λ,λ_0} of Eq. (86) by

$$C_{\lambda,\lambda_0}C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''} = 1,$$

whence Eq. (99a) yields the result (cf. Ref. 12)

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix} = C_{\lambda,\lambda_0}\rho\begin{pmatrix}\lambda' & \lambda'\\ \lambda' & \lambda'\end{pmatrix}\circ\rho\begin{pmatrix}\lambda'' & \bar{\lambda}''\\ \lambda'' & \bar{\lambda}''\end{pmatrix}, \qquad (99b)$$

where $\rho\begin{pmatrix}\lambda' & \lambda'\\ \lambda' & \lambda'\end{pmatrix}$ and $\rho\begin{pmatrix}\lambda'' & \bar{\lambda}''\\ \lambda'' & \bar{\lambda}''\end{pmatrix}$ are given by Eqs. (96) and (97), respectively.

We note that the two RWF's on the rhs of Eq. (99) commute because they are semimaximal and hence of the same tie structure (but not the same extremal tie structure). Application of Weyl group symmetry to Eq. (99) demonstrates that all optimal RWF's are expressible as a product of elementary RWF's (times a scalar C_{λ,λ_0}).

In conclusion we remark that Eqs. (96)–(99) have a suitable extension to the orthogonal groups.

ACKNOWLEDGMENTS

The author gratefully acknowledges the financial support of an A.R.G.S. research grant.

The author thanks the referee whose useful suggestions have improved the manuscript in several places.

[1]H. Casimir, Proc. K. Akad. (Amsterdam) **38**, 844 (1931).
[2]H. Casimir and B. L. Van der Waerden, Math. Ann. **111**, 1 (1935).
[3]G. Racah, Rend. Atti Accad. Naz. Lincei. **8**, 108 (1950).
[4]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen, London, 1931); *The Classical Groups* (Princeton U. P., Princeton, NJ, 1939).
[5]J. Schwinger, "On angular momentum," preprinted in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965); see also L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum physics," in *Encyclopedia of Mathematics and Applications*, Vol. 8 (Addison–Wesley, Reading, MA, 1981).
[6]V. Bargmann, Commun. Pure Appl. Math. **4**, 198 (1961); K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience, New York, 1953); M. Moshinsky, Nucl. Phys. **31**, 384 (1962); Rev. Mod. Phys. **34**, 813 (1962).
[7]I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk. SSSR. **71**, 825 (1950). The results of this paper are an announcement only. The proof appears later in I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR **29**, 1329 (1965) [see also AMS Transl. **65**, 116 (1967), Ser. 2].
[8]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[9]M. D. Gould, J. Math. Phys. **22**, 2376 (1981).
[10]M. D. Gould, J. Math. Phys. **22**, 15 (1981).
[11]M. D. Gould, J. Math. Phys. **21**, 444 (1980); J. Aust. Math. Soc. B **20**, 401 (1978).
[12]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968); L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).
[13]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **12**, 1723 (1964); **5**, 1730 (1964); **6**, 1845 (1965).
[14]E. Chacon, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13**, 577 (1972).
[15]M. K. F. Wong, J. Math. Phys. **10**, 1065 (1969); **12**, 1530 (1971); **19**, 703 (1978); **17**, 1558 (1976); W. J. Holman, III, Nuovo Cimento A **4**, 904 (1971); T. Maekawa, J. Math. Phys. **16**, 334 (1975).
[16]M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967); S. C. Pang and K. T. Hecht, *ibid.* **8**, 1233 (1967).
[17]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970); **13**, 1985 (1972); **14**, 1336 (1973); **16**, 2408 (1975); **18**, 1833 (1977); L. C. Biedenharn, J. D. Louck, E. Chácon, and M. Ciftan, J. Math. Phys. **13**, 1957 (1972).
[18]S. A. Edwards and M. D. Gould, "Cyclic vectors, projection methods and the multiplicity problem," University of Adelaide preprint, 1980; J. Math. Phys. **25**, 2848 (1984).
[19]K. T. Hecht and J. P. Draayer, Nucl. Phys. A **223**, 285 (1974); K. T. Hecht, J. Math. Phys. **15**, 2148 (1974); S. A. Edwards, Ph.D. dissertation, Adelaide 1981 (unpublished).
[20]M. D. Gould, J. Math. Phys. **27**, 1964 (1986).
[21]S. J. Alisauskas, A. A. Jucys, and A. P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[22]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968); "The harmonic oscillator in modern physics: From atoms to quarks," in *Documents on Modern Physics* (Gordon and Breach, New York, 1969); P. Kramer and M. Moshinsky, in Ref. 23, Vol. 1.
[23]*Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968), Vols. I–III.
[24]R. Paunz, *Alternate Molecular Orbital Method* (Saunders, Philadelphia, 1967); *Spin Eigenfunctions: Construction and Use* (Plenum, New York, 1979).
[25]J. Paldus, in *Theoretical Chemistry: Advances and Perspectives*, edited by H. Eyring and D. J. Henderson (Academic, New York, 1972), Vol. 2, p. 131; Phys. Scripta **21**, 295 (1980); Int. J. Quant. Chem. Symp. **9**, 165 (1975).
[26]"The unitary group," in *Lecture Notes in Chemistry*, Vol. 22, edited by J. Hinze (Springer, New York, 1981).
[27]M. D. Gould and G. S. Chandler, Int. J. Quant. Chem. **25**, 553 (1984).
[28]W. G. Harter, Phys. Rev. A **8**, 2819 (1973); W. G. Harter and C. W. Patterson, "A unitary calculus for atomic orbitals," in *Lecture Notes in Physics*, Vol. 49 (Springer, New York, 1976).
[29]B. R. Brooks and H. F. Schaefer, III, J. Chem. Phys. **70**, 5092 (1979); B. R. Brooks, W. D. Laidig, P. Saxe, N. C. Handy, and H. F. Schaefer, III, Phys. Scripta **21**, 312 (1980); J. Chem. Phys. **72**, 2837, 4652 (1980).
[30]J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mex. Fis. **14**, 29 (1965).
[31]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[32]J. Dixmier, *Enveloping Algebras* (North-Holland, Amsterdam, 1977).
[33]M. D. Gould, J. Aust. Math. Soc. B **26**, 257 (1985).
[34]M. D. Gould, Ann. Inst. H. Poincaré A **32**, 203 (1980); J. Phys. A: Math. Gen. **17**, 1 (1984).
[35]M. D. Gould, "Casimir invariants and infinitesimal characters for semisimple Lie algebras," University of Adelaide preprint.
[36]S. A. Edwards, J. Math. Phys. **1**, 164 (1978); V. S. Popov and A. M. Perelomov, Sov. J. Nucl. Phys. **5**, 489 (1967); **3**, 676 (1966).
[37]H. S. Green, J. Math. Phys. **12**, 2106 (1971); A. J. Bracken and H. S. Green, *ibid.* **12**, 2099 (1971).

Multiplicity-free Wigner coefficients for semisimple Lie groups. II. A pattern calculus for O(n)

M. D. Gould
School of Chemistry, University of Western Australia, Nedlands, Western Australia 6009

(Received 14 November 1984; accepted for publication 16 April 1986)

In this paper a direct, algebraic derivation of a large class of (multiplicity-free) reduced Wigner coefficients is presented for the orthogonal groups. In particular all elementary reduced Wigner coefficients, including those for the fundamental spinor and tensor representations, are obtained. The results are presented in a form directly analogous to the corresponding results obtained for the unitary groups.

I. INTRODUCTION

This is the second paper in a series of two in which we consider the direct evaluation of multiplicity-free *reduced Wigner coefficients* (RWC's) for the orthogonal and unitary groups. In the first paper of the series[1] (herein referred to as I) it was demonstrated how the (multiplicity-free) RWC's for the canonical imbeddings $U(n+1) \supset U(n)$ [resp. $O(n+1) \supset O(n)$] may be obtained algebraically from the eigenvalues of certain $U(n)$ [resp. $O(n)$] Casimir invariants. In this approach the Weyl group symmetries of the RWC's are evident from the outset, in view of the simple transformation properties of polynomial functions, and the method yields a direct (i.e., nonrecursive) derivation of the RWC's.

The $U(n+1)$: $U(n)$ [resp. $O(n+1)$: $O(n)$] squared RWC's are obtained, in our approach, as a rational polynomial function (numerator polynomial divided by denominator polynomial) in the representation labels of the groups $U(n+1)$ and $U(n)$ [resp. $O(n+1)$ and $O(n)$]. A unified approach to the denominator polynomials was presented in I, where the denominator polynomials were obtained for both the unitary and orthogonal groups. The corresponding numerator polynomials for the unitary groups were also derived (algebraically), and our results yield the pattern calculus rules of Biedenharn, and co-workers.[2] It is our aim here to extend these results to obtain the numerator polynomials for the orthogonal groups. The results of this paper then enable a direct evaluation of a large class of multiplicity-free RWC's for the orthogonal groups. (The class of RWC's considered in this paper is analogous to the class of RWC's considered in the pattern calculus of Ref. 2 for the unitary groups.)

Some interesting special cases of our general formulas are also considered. In particular all vector RWC's for the orthogonal groups are given and our results are shown to agree with those of Ref. 3. All RWC's corresponding to the elementary tensor and spinor representations are also given. In particular our results enable a complete evaluation of all Wigner coefficients for the fundamental spinor (and vector) representations. The class of RWC's we consider in examples corresponds to all *elementary* RWC's of $O(n)$, which are the orthogonal group analog of the elementary Wigner operators of Ref. 2 [which form the fundamental building blocks in the $U(n)$ pattern calculus]. During the course of our investigation it shall be demonstrated that the $O(n)$ RWC formulas, developed in this paper, are extendable to a wider class of RWC's.

The algebraic methods of this paper and their relationship to the pattern calculus of Ref. 2 has been discussed in I (see also Ref. 3). Although our approach affords a direct evaluation of the (multiplicity-free) RWC's, the pattern calculus techniques of Ref. 2 are useful, particularly for computational applications. To this end we consider in this paper the beginnings of a pattern calculus for $O(n)$. It is demonstrated that the RWC's of $O(n)$ may be expressed as a commuting *product* (or composition) of elementary RWC's (up to a constant) in direct analogy with the unitary groups. It is hoped that these results, on manipulating and multiplying RWC's, may lead to the evaluation of a larger class of RWC's for the orthogonal groups in direct analogy with the $U(n)$ case. We remark, in this connection, that the pattern calculus rules developed in Ref. 2 for $U(n)$ have been applied successfully[4,5] to evaluate all RWC's for the symmetric tensor representations (only some of these latter RWC's being included in the original pattern calculus of Ref. 2).

From the point of view of future research we note that a phase calculus, for obtaining suitable phases for the $O(n)$ RWC's, needs to be developed. The beginning of such a calculus of phases is given in Ref. 3, where suitable phases were obtained for the fundamental (i.e., vector) Wigner coefficients of $O(n)$. Also, it would be desirable to extend the techniques of this paper to evaluate a larger class of RWC's for the orthogonal and unitary groups. In particular our methods are applicable, in principle, to the problem of evaluating all multiplicity-averaged RWC's (see I) for the orthogonal and unitary groups (which include, as a special case, all multiplicity-free RWC's). Also, as mentioned in I, our methods are applicable to other (noncanonical) subgroup imbeddings $G \supset G_0$ enabling an evaluation of multiplicity-averaged G: G_0 RWC's. In particular our methods enable an extension to the symplectic groups to yield multiplicity-averaged $Sp(2n)$: $Sp(2n-2) \times Sp(2)$ RWC's. Further work along these lines is now in progress.

The paper is set up as follows. Our notation and basic conventions are established in Sec. II. In Secs. III and IV all *optimal* $O(n+1)$:$O(n)$ RWC's are evaluated. In Sec. V we consider some interesting special cases of our general results and in particular give all elementary $O(n+1)$: $O(n)$

 0022-2488/86/081964-16$02.50

RWC's. We conclude in Sec. VI with the beginnings of a pattern calculus for $O(n)$.

Throughout this paper, unless otherwise stated, we adopt the notation and conventions of I. For other work on the orthogonal groups we refer to the literature cited in Refs. 1 and 3.

II. PRELIMINARIES

Following the notation of I, throughout this paper we let L (resp. L_0) denote the Lie algebra of $O(n+1)$ [resp. $O(n)$]. The $\frac{1}{2}n(n+1)$ generators a_{ij} $(i,j=1,...,n+1)$ of the Lie group $O(n+1)$ satisfy the relations[6]

$$a_{ij}=-a_{ji},\quad [a_{ij},a_{kl}]$$
$$=\delta_{kj}a_{il}-\delta_{il}a_{kj}-\delta_{ki}a_{jl}+\delta_{jl}a_{ki}$$

and are moreover required to satisfy the Hermiticity condition

$$a_{ij}^{\dagger}=a_{ji}=-a_{ij} \tag{1}$$

on finite-dimensional (i.e., unitary) representations of the group. We choose as a Cartan subalgebra H for $O(n+1)$ the space spanned by the operators

$$h_r=-ia_{2r-1,2r},\quad r=1,...,h=[(n+1)/2]. \tag{2}$$

We have included the imaginary phase $-i$ in Eq. (2) because Eq. (1) implies the Cartan subalgebra elements $a_{2r-1,2r}$ are to be represented by anti-Hermitian matrices, so we consider the Hermitian operators of Eq. (2), which have real eigenvalues leading to real weight components in analogy with the unitary groups. The weights $\lambda\in H^*$ may thus be identified with the h-tuples $\lambda=(\lambda_1,...,\lambda_h)$, where $\lambda_r=\lambda(h_r)$.

We remark that the $O(n+1)$ generators a_{ij} [which correspond to the choice of the $O(n+1)$ metric $g_{ij}=\delta_{ij}$] are not in Cartan form. Nevertheless it is easily demonstrated[6,7] that we may take, as a set of positive roots for $O(n+1)$, the weights

$$\epsilon_i\pm\epsilon_j\quad(1\leqslant i<j\leqslant h),\quad n+1=2h,$$
$$\epsilon_i\pm\epsilon_j\quad(1\leqslant i<j\leqslant h),\quad \epsilon_i\quad(1\leqslant i\leqslant h),\quad n+1=2h+1,$$

where ϵ_i denotes the fundamental weight with 1 in the ith position and zeros elsewhere. Thus δ, the half-sum of the positive roots, is given by

$$\delta=\frac{1}{2}\sum_{r=1}^{h}(n+1-2r)\epsilon_r \tag{3}$$

(which holds for both cases $n+1$ odd or even). In this case the inner product on H^* induced by the Killing form is given by

$$(\lambda,\mu)=\sum_{r=1}^{h}\lambda_r\mu_r,\quad \lambda,\mu\in H^*.$$

Thus we may write, for all roots $\alpha\in\Phi^+$,

$$\langle\lambda,\alpha\rangle=(\lambda,\alpha),\quad \alpha=\epsilon_i\pm\epsilon_j\quad(i<j),$$
$$\langle\lambda,\epsilon_i\rangle=2(\lambda,\epsilon_i)=2\lambda_i,\quad \alpha=\epsilon_i,\quad n+1=2h+1. \tag{4}$$

In the case of $O(n+1=2h+1)$ the Weyl group acts as the group of all permutations and sign changes of the set $\{\epsilon_1,...,\epsilon_h\}$ so W is isomorphic to the semidirect product of S_h (the symmetric group on h objects) and $\mathbb{Z}_2^h$:

$$W\cong S_h\circledS\mathbb{Z}_2^h,\quad n+1=2h+1. \tag{5a}$$

The Weyl group in the case of $O(n+1=2h)$ is the group of permutations and sign changes involving only even numbers of signs of the set $\{\epsilon_1,...,\epsilon_h\}$. So in this case we have

$$W\cong S_h\circledS\mathbb{Z}_2^{h-1},\quad n+1=2h. \tag{5b}$$

To clarify the action of the Weyl group on our weights we note that Eq. (5) indicates that each Weyl group element $\sigma\in W$ may be uniquely expressed in the form

$$\sigma=e\pi, \tag{6}$$

where $e\in\mathbb{Z}_2^h$ (or $\mathbb{Z}_2^{h-1}$) and $\pi\in S_h$ is a permutation (note that the elements e of $\mathbb{Z}_2^k$ are all reflections: $e^2=1$). The action of an element $\pi\in S_h$ on a weight $\lambda\in H^*$ is given by [cf. I, Eq. (65)]

$$(\pi\lambda)_r=\lambda_{\pi^{-1}(r)}. \tag{7a}$$

The action of an element $e\in\mathbb{Z}_2^h$ (or $\mathbb{Z}_2^{h-1}$) is given by

$$(e\lambda)_r=\mathrm{sgn}(e_r)\lambda_r, \tag{7b}$$

where $\mathrm{sgn}(e_r)=\pm1$ is given by

$$\mathrm{sgn}(e_r)=\begin{cases}1, & \text{if } e\epsilon_r=\epsilon_r,\\ -1, & \text{if } e\epsilon_r=-\epsilon_r.\end{cases}$$

Thus the action of the Weyl group element (6) on a weight λ is determined by

$$(\sigma\lambda)_r=\mathrm{sgn}(e_r)\lambda_{\pi^{-1}(r)}. \tag{7c}$$

The finite-dimensional irreducible representations of the Lie group $O(n+1)$ are uniquely characterized by their highest weights $\lambda=(\lambda_1,...,\lambda_h)$ whose components are required to satisfy the conditions[7,8]

$$\begin{aligned}&\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_h\geqslant0,\quad n=2h+1,\\ &\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{h-1}\geqslant|\lambda_h|,\quad n=2h,\end{aligned} \tag{8a}$$

$$\lambda_i-\lambda_j\in\mathbb{Z},\quad \lambda_i+\lambda_j\in\mathbb{Z}. \tag{8b}$$

We note that Eq. (8b) implies the components λ_r are all either integers (corresponding to tensor representations) or all half-odd integers (corresponding to spinor representations). We have in particular the fundamental dominant weights

$$\Lambda_k=\sum_{r=1}^{k}\epsilon_r,\quad 1\leqslant k\leqslant h-2,\quad h>2, \tag{9a}$$

together with the additional dominant weights

$$\begin{aligned}&\Lambda_{h-1},\quad \Lambda_s=\tfrac{1}{2}\Lambda_h,\quad n+1=2h+1,\\ &\Lambda_s^+=\tfrac{1}{2}\Lambda_h,\quad \Lambda_s^-=\tfrac{1}{2}\Lambda_{h-1}-\tfrac{1}{2}\epsilon_h,\quad n+1=2h,\end{aligned} \tag{9b}$$

which constitute a basis for H^*. We remark that the elementary weights $\Lambda_s,\Lambda_s^\pm$ correspond to the highest weights of the fundamental spinor representations.

The corresponding dual basis to (9) is given by

$$\alpha_i=\epsilon_i-\epsilon_{i+1},\quad i=1,...,h-2, \tag{10a}$$

together with

$$\begin{aligned}&\alpha_{h-1}=\epsilon_{h-1}-\epsilon_h,\quad \alpha_h=\epsilon_h,\quad n+1=2h+1,\\ &\alpha_{h-1}=\epsilon_{h-1}+\epsilon_h,\quad \alpha_h=\epsilon_{h-1}-\epsilon_h,\quad n+1=2h,\end{aligned} \tag{10b}$$

which consitute a system of simple roots for L. An arbitrary dominant weight λ may thus be expressed as a positive Z-linear combination of the dominant weights (9) according to[7] [cf. Eq. (4)]

$$\lambda = \sum_{r=1}^{h-2} \langle\lambda,\alpha_r\rangle\Lambda_r + \langle\lambda,\alpha_{h-1}\rangle\Lambda_{h-1} + \langle\lambda,\alpha_h\rangle\Lambda_s = \sum_{r=1}^{h-1} (\lambda_r - \lambda_{r+1})\Lambda_r + 2\lambda_h\Lambda_s, \quad n+1 = 2h+1, \tag{11a}$$

$$\lambda = \sum_{r=1}^{h-2} \langle\lambda,\alpha_r\rangle + \langle\lambda,\alpha_{h-1}\rangle\Lambda_s^+ + \langle\lambda,\alpha_h\rangle\Lambda_s^- = \sum_{r=1}^{h-2} (\lambda_r - \lambda_{r+1})\Lambda_r + (\lambda_{h-1} + \lambda_h)\Lambda_s^+ + (\lambda_{h-1} - \lambda_h)\Lambda_s^-, \quad n+1 = 2h. \tag{11b}$$

An alternative characterization of the finite-dimensional irreducible representations is given by the eigenvalues of the $O(n+1)$ Gel'fand invariants

$$I_2 = \sum_{i,j=1}^{n+1} \alpha_{ij}\alpha_{ji}, \quad I_3 = \sum_{i,j,k=1}^{n+1} \alpha_{ij}\alpha_{jk}\alpha_{ki}, \quad \text{etc.}$$

The operators I_r $(r = 2,...,n+1)$ form a full set of invariants for L, but it may be shown[9] that only the even invariants I_{2r} $(r = 1,...,h)$ are algebraically independent. Thus the center Z of the universal enveloping algebra U of L may be written, in polynomial algebra notation, as

$$Z = \mathbb{C}[I_2, I_4, ..., I_{2h}].$$

We note, in particular, that the eigenvalue of the second-order invariant I_2 on the irreducible U-module with highest weight $\lambda \in \Lambda^+$ is given by[9]

$$\chi_\lambda(I_2) = 2(\lambda,\lambda + 2\delta) = 2\sum_{r=1}^{h} \lambda_r(\lambda_r + n + 1 - 2r).$$

Thus we must have $I_2 = 2C_L$, where C_L is the universal Casimir element of L. The eigenvalues of the higher-order invariants I_r $(r > 2)$ are given in Refs. 9 and 10.

Following our treatment of the unitary groups we note that the generators α_{ij} of the Lie group $O(n+1)$ fit naturally into a $(n+1)\times(n+1)$ matrix

$$b = [\alpha_{ij}].$$

The matrix b is a special case of the more general matrices b_λ of I [Eq. (12)], where π_λ, in this case, corresponds to the fundamental contragredient vector representation. Polynomials in the matrix b may then be defined recursively according to

$$[b^{m+1}]_{ij} = \sum_{k=1}^{n+1} \alpha_{ik}[b^m]_{kj} = \sum_{k=1}^{n+1} [b^m]_{ik}\alpha_{kj}.$$

It can be shown[6,9] that the matrix b satisfies a $(n+1)$-order polynomial identity over the center Z of U, which may be written in its factorized form as

$$\prod_{k=1}^{n+1} (b - \beta_k) = 0, \tag{12}$$

where the characteristic roots β_k are invariants of the group whose eigenvalues on a finite-dimensional irreducible module with highest weight λ are given by

$$\begin{aligned} \beta_r &= \bar{\beta}_{n+2-r} = \lambda_r + n - r, \\ \beta_{n+2-r} &= \bar{\beta}_r = r - 1 - \lambda_r, \quad r = 1,...,h, \end{aligned} \tag{13a}$$

together with

$$\beta_{h+1} = \bar{\beta}_{h+1} = h, \quad \text{for } n+1 = 2h+1.$$

Note that we have the following relation between the characteristic roots:

$$\beta_{n+2-r} + \beta_r = n - 1, \quad r = 1,...,h. \tag{13b}$$

As in our treatment of the unitary groups the characteristic roots of Eq. (13) play a fundamental role in our work and may be regarded as the orthogonal group analog of the partial hooks employed in the work of Biedenharn and Louck.[2] The action of a translated Weyl group element $\tilde{\sigma} \in \tilde{W}$, with σ as in Eq. (6), on the characteristic roots β_k is given by

$$\tilde{\sigma}\beta_k = \begin{cases} \beta_{\pi(k)}, & \text{sgn}(e_k) = 1, \\ \bar{\beta}_{\pi(k)}, & \text{sgn}(e_k) = -1, \end{cases} \quad k = 1,...,h, \tag{14}$$

$$\tilde{\sigma}\bar{\beta}_k = (\overline{\tilde{\sigma}\beta_k}),$$

with

$$\tilde{\sigma}\beta_{h+1} = \beta_{h+1}, \quad \text{for odd } n+1 = 2h+1.$$

Thus the translated Weyl group $\tilde{W}$ permutes the characteristic roots β_k $(k = 1,...,n+1)$. Keeping in mind the special relations between the characteristic roots, implied by Eq. (13), we may identify the center Z of U with the algebra of symmetric polynomials in the β_k (cf. Harish Chandra's theorem):

$$Z = \mathbb{C}\{\beta_1,...,\beta_{n+1}\},$$

where $\mathbb{C}\{x_1,...,x_{n+1}\}$ denotes the algebra of symmetric polynomials in $n+1$ indeterminates $x_1,...,x_{n+1}$.

With regard to the subgroup $O(n)$ and its Lie algebra L_0 we follow Ref. 1 and adopt the same notation as above except we add a subscript $_0$ to everything. Thus δ_0, the half-sum of the positive roots, is given by [cf. Eq. (3)]

$$\delta_0 = \frac{1}{2}\sum_{i=1}^{h_0} (n - 2i)\epsilon_{0i}, \quad h_0 = \begin{cases} h, & n = 2h, \\ h-1, & n = 2h-1. \end{cases} \tag{15}$$

The $O(n)$ matrix $a = [\alpha_{ij}]$ $(i,j = 1,...,n)$ satisfies an nth-order polynomial identity analogous to Eq. (12):

$$\prod_{r=1}^{n} (a - \alpha_r) = 0.$$

The characteristic roots in this case take a constant value on a finite-dimensional irreducible U_0-module with highest weight $\lambda_0 = (\lambda_{01},...,\lambda_{0h_0})$, given by

$$\begin{aligned} \alpha_r &= \bar{\alpha}_{n+1-r} = \lambda_{0r} + n - 1 - r, \\ \alpha_{n+1-r} &= \bar{\alpha}_r = r - 1 - \lambda_{0r}, \quad r = 1,...,h_0, \end{aligned} \tag{16}$$

with

$$\alpha_h = \bar{\alpha}_h = h - 1, \quad \text{for odd } n = 2h-1.$$

The translated Weyl group $\tilde{W}_0$ acts on the characteristic roots α_r by permuting them among themselves in analogy with Eq. (14).

We may also define the fundamental Gel'fand invariants $I_{02r} = \mathrm{tr}[a^{2r}]$ $(r = 1,...,h_0)$ of $O(n)$, which generate the center Z_0 of the universal enveloping algebra U_0 of L_0. In this case, keeping in mind the relations implied by Eq. (16), the center Z_0 may be identified with the ring of symmetric polynomials in the characteristic roots α_r:

$$Z_0 = \mathbb{C}\{\alpha_1,...,\alpha_r\}.$$

Since the imbedding $L \supset L_0$ is canonical, the centralizer of L_0 in U is given by (cf. I, Sec. II)

$$\mathscr{Z} = Z \otimes Z_0 = Z_0 \otimes Z \text{ (enveloping algebra product).}$$

We may clearly identify $\mathscr{Z}$ with the algebra [keeping in mind the conditions implied by Eqs. (13) and (16)]

$$\mathscr{Z} = \mathbb{C}\{\beta_1,...,\beta_{n+1}:\ \alpha_1,...,\alpha_n\}$$

where $\mathbb{C}\{x_1,...,x_{n+1}:\ y_1,...,y_n\}$ denotes the algebra of all polynomials in indeterminates $x_1,...,x_{n+1}, y_1,...,y_n$, which are symmetric in the x_k and also in the y_r.

In analogy with I, Eqs. (71) and (72), it can be shown[3,6] that the $(n+1, n+1)$ entries of the matrix powers b^m are invariants of $O(n)$:

$$[b^m]_{n+1,n+1} \in \mathscr{Z}. \tag{17}$$

Thus we may express these centralizer elements as a polynomial in the $O(n+1)$ and $O(n)$ Gel'fand invariants I_k and I_{0r}. Alternatively we may express the invariants of Eq. (17) as a symmetric polynomial in the characteristic roots β_k, α_r according to[3,6]

$$[b^m]_{n+1,n+1} = \sum_{k=1}^{n+1} \beta_k^m C_k,$$

where

$$C_k = \prod_{p \neq k}^{n+1} (\beta_k - \beta_p)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r - \eta_r) \tag{18}$$

and where

$$\eta_r = \begin{cases} 1, & n = 2h, \\ 1 - \delta_{r,h}, & n = 2h - 1. \end{cases}$$

If $V(\lambda)$ denotes a finite-dimensional irreducible U-module with highest weight $\lambda \in \Lambda^+$, then it is well known[8] that $V(\lambda)$ decomposes into a direct sum of irreducible U_0-submodules according to [cf. I, Eqs. (10) and (73)]

$$V(\lambda) = \bigoplus_{\lambda_0 \in [\lambda]} V(\lambda\,|\lambda_0), \tag{19}$$

where the $O(n)$ highest weights are to be dominant [cf. Eq. (8)] and satisfy the betweenness conditions[8]

$$\lambda_1 \geqslant \lambda_{01} \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_h \geqslant \lambda_{0h} \geqslant -\lambda_h, \quad n = 2h,$$
$$\lambda_1 \geqslant \lambda_{01} \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{h-1} \geqslant h_{0h-1} \geqslant |\lambda_h|, \quad n = 2h - 1. \tag{20}$$

Thus, for this case, we call a weight $(\lambda\,|\lambda_0)$ *lexical* if and only if λ and λ_0 are dominant and satisfy the betweenness conditions of Eq. (20). We note that the components of the weight $(\lambda\,|\lambda_0)$ are all either simultaneously integers (corresponding to tensor representations) or half-odd integers (corresponding to spinor representations).

By repeated application of the above results we see that the group $O(n+1)$ admits the canonical chain of subgroups

$$O(n+1) \supset O(n) \supset \cdots O(3) \supset O(2), \tag{21}$$

whose representation labels serve to completely label the basis states of the irreducible representations. This is the state labeling scheme proposed by Gel'fand and Tsetlin.[11] The partitions for each of the groups in the chain (21) is most conveniently arranged into a Gel'fand–Tsetlin (GT) pattern, which is described in detail in Ref. 8. We see therefore that the lexical weights $(\lambda\,|\lambda_0)$ occurring in the decomposition (19) are to constitute the top two rows of the $O(n+1)$ GT patterns.

It is our aim in this paper to determine all *optimal* reduced Wigner coefficients (RWC's) (see I, Sec. III for definitions) for the canonical imbedding $O(n+1) \supset O(n)$. We recall, from I, Sec. V, that the squared optimal $O(n+1)$: $O(n)$ RWC's are determined by

$$\rho\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}(\mu|\mu_0) = \left| \left\langle \begin{matrix} \mu+\Delta \\ \mu_0+\Delta_0 \end{matrix} \right| \left| \begin{matrix} \lambda \\ \lambda_0 \end{matrix} \quad ; \quad \begin{matrix} \mu \\ \mu_0 \end{matrix} \right\rangle \right|^2,$$
$$\Delta \in \mathrm{Sym}(\lambda), \quad \Delta_0 \in \mathrm{Sym}(\lambda_0), \tag{22}$$

where $\rho\left(\begin{smallmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{smallmatrix}\right)$ is the rational polynomial function given by

$$\rho\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix} = C_{\lambda,\lambda_0} \frac{\eta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}}{\delta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}},$$
$$\Delta \in \mathrm{Sym}(\lambda), \quad \Delta_0 \in \mathrm{Sym}(\lambda_0), \tag{23}$$

where C_{λ,λ_0} is a numerical constant (depending only on labels λ and λ_0) and $\eta\left(\begin{smallmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{smallmatrix}\right)$ [resp. $\delta\left(\begin{smallmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{smallmatrix}\right)$] is a polynomial function in the representation labels of $O(n+1)$ and $O(n)$, herein referred to as the numerator (resp. denominator) polynomial. We recall from I that in Eqs. (22) and (23) we have adopted the convention that $\mathrm{Sym}(\lambda)$ [resp. $\mathrm{Sym}(\lambda_0)$] denotes the set of weights Weyl group conjugate to the weight λ (resp. λ_0).

The denominator polynomials of Eq. (23) are given explicitly by I, Eq. (62), and it remains to determine the numerator polynomials

$$\eta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix}, \quad \Delta \in \mathrm{Sym}(\lambda), \quad \Delta_0 \in \mathrm{Sym}(\lambda_0), \tag{24}$$

which, according to I, Eq. (64), are to satisfy the symmetry rules

$$\tilde{\sigma}\,\eta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix} = \eta\begin{pmatrix} \lambda & \lambda_0 \\ \sigma(\Delta) & \Delta_0 \end{pmatrix}, \quad \sigma \in W,$$
$$\tilde{\sigma}_0\eta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \Delta_0 \end{pmatrix} = \eta\begin{pmatrix} \lambda & \lambda_0 \\ \Delta & \sigma_0(\Delta_0) \end{pmatrix}, \quad \sigma_0 \in W_0. \tag{25}$$

We recall from I, Eqs. (34) and (38), that the denominator polynomials [and hence the squared RWC's of Eq. (22)] are to satisfy the same symmetry rules.

It is our aim in this paper to determine the numerator polynomials of Eq. (24) as a polynomial in the characteristic roots β_k and α_r. This then enables a complete determination of the optimal RWC's for the imbedding $O(n+1) \supset O(n)$ through Eq. (23). Unfortunately, however, it is evidently necessary to treat the cases n even and n odd separately. Accordingly, in the following section we restrict ourselves to the imbedding $O(2h+1) \supset O(2h)$ (i.e., n even) and leave it to Sec. IV to consider the case $n = 2h - 1$ (i.e., n odd).

III. OPTIMAL RWC'S FOR $O(2h+1)\supset O(2h)$

It is our aim here to determine all optimal RWC's for the canonical imbedding $O(2h+1)\supset O(2h)$: i.e., we are considering the case where $n=2h$ is even. Throughout we adopt the notation and conventions of I (see in particular I, Secs. III and V).

We recall from I, Sec. V that the numerator polynomial of Eq. (24) is to determine a polynomial in the characteristic roots β_k,α_r of degree $r_\lambda + r_{\lambda_0}$ [see remarks preceding I, Eq. (57)], where the integers r_λ and r_{λ_0} in this case are given by [cf. Eq. (4)]

$$r_\lambda = \sum_{\alpha>0} \langle\lambda,\alpha\rangle = 2(\lambda,\delta) + \sum_{i=1}^{h} \lambda_i = \sum_{k=1}^{h} (n+2-2k)\lambda_k, \tag{26}$$

$$r_{\lambda_0} = \sum_{\alpha_0>0} \langle\lambda_0,\alpha_0\rangle = 2(\lambda_0,\delta_0) = \sum_{r=1}^{h} (n-2r)\lambda_{0r},$$

where we have used Eqs. (3) and (15).

Suppose now that $(\mu|\mu_0)\in\mathscr{L}$ is a lexical weight. It follows in view of Eq. (22) and the betweenness conditions of Eq. (20) that the numerator polynomial of Eq. (24) vanishes if the weight $(\mu+\Delta|\mu_0+\Delta_0)$ is nonlexical: since in such a case the RWC

$$\left\langle \begin{matrix}\mu+\Delta\\ \mu_0+\Delta_0\end{matrix} \middle\| \begin{matrix}\lambda\\ \lambda_0\end{matrix} \;;\; \begin{matrix}\mu\\ \mu_0\end{matrix}\right\rangle$$

vanishes. Thus we deduce a vanishing contribution whenever the following situations occur:

$$\begin{aligned}&\text{(i) } (\mu_0+\Delta_0)_r > (\mu+\Delta)_r, \\ & \qquad\qquad\qquad\qquad\qquad\qquad r=1,\dots,h,\\ &\text{(ii) } (\mu+\Delta)_{r+1} > (\mu_0+\Delta_0)_r, \\ &\text{(iii) } -(\mu+\Delta)_h > (\mu_0+\Delta_0)_h.\end{aligned} \tag{27}$$

In case (i) above we have

$$(\mu_0+\Delta_0)_r = (\mu+\Delta)_r + m, \quad \text{for some } m\in\mathbb{Z}^+.$$

To determine the possible range of m-values we note that lexicality of the weight $(\mu|\mu_0)$ implies that m must lie in the range

$$1\leqslant m\leqslant \Delta_{0r}-\Delta_r.$$

Thus we can only get a vanishing contribution if $\Delta_{0r}>\Delta_r$, in which case we deduce divisibility of the numerator polynomial (24) by factors

$$\begin{aligned}\mu_{0r}+\Delta_{0r}-\mu_r-\Delta_r-m &\\ =\alpha_r-\beta_r+\Delta_{0r}-\Delta_r-m+1&,\\ m=1,\dots,\Delta_{0r}-\Delta_r&,\end{aligned}$$

or equivalently by factors

$$\alpha_r-\beta_r+m, \quad m=1,\dots,\Delta_{0r}-\Delta_r.$$

Thus we deduce divisibility of the numerator polynomial (24) by the set of factors

$$\prod_{m=1}^{\Delta_{0r}-\Delta_r} (\alpha_r-\beta_r+m), \quad \Delta_{0r}>\Delta_r.$$

Applying a similar argument to the conjugate numerator polynomial

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \sigma_0(\Delta_0)\end{pmatrix}, \quad \sigma\in W, \quad \sigma_0\in W_0,$$

we deduce divisibility of this latter polynomial by factors

$$\prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r} (\alpha_r-\beta_r+m), \quad \sigma_0(\Delta_0)_r>\sigma(\Delta)_r.$$

But in view of the Weyl group symmetries of Eq. (25) we may write

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = \tilde{\sigma}^{-1}\tilde{\sigma}_0^{-1}\eta\begin{pmatrix}\lambda & \lambda_0\\ \sigma(\Delta) & \sigma_0(\Delta_0)\end{pmatrix},$$

which would imply [cf. Eqs. (14)] that the numerator polynomial (24) is divisible by the further set of factors

$$\begin{aligned}\tilde{\sigma}^{-1}\tilde{\sigma}_0^{-1} &\prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r} (\alpha_r-\beta_r+m)\\ &= \prod_{m=1}^{\sigma_0(\Delta_0)_r-\sigma(\Delta)_r} (\alpha_{\sigma_0^{-1}(r)}-\beta_{\sigma^{-1}(r)}+m),\\ &\sigma_0(\Delta_0)_r>\sigma(\Delta)_r.\end{aligned}$$

Taking into account arbitrary *permutations* $\sigma=\pi$, $\sigma_0=\pi_0\in S_h$ [cf. Eq. (5)] we deduce divisibility of the numerator polynomial by the set of factors

$$\prod_{\substack{r,k=1\\ \Delta_{0r}>\Delta_k}}^{h} \prod_{m=1}^{\Delta_{0r}-\Delta_k} (\alpha_r-\beta_k+m). \tag{28a}$$

So far we have applied only permutational symmetries and we need to take into account the fact that the Weyl groups also contain sign changes of the weight components. In view of Eq. (14) we see that such reflections have the effect of interchanging the roots $\beta_k\leftrightarrow\bar{\beta}_k$, $\alpha_r\leftrightarrow\bar{\alpha}_r$: a reflection $e\in\mathbb{Z}^h$, which changes the sign of the k th component of a weight $(k\leqslant h)$, effects the transformation

$$\tilde{e}\beta_k=\bar{\beta}_k, \quad \tilde{e}\bar{\beta}_k=\beta_k$$

[and a similar statement for sign changes of the $O(n)$ weight components]. Thus, taking into account sign changes of the weight components, we deduce, from Eq. (28a), divisibility of the numerator polynomial (24) by the additional set of factors

$$\prod_{\substack{r,k=1\\ \Delta_{0r}>-\Delta_k}}^{h} \prod_{m+1}^{\Delta_{0r}+\Delta_k} (\alpha_r-\bar{\beta}_k+m), \tag{28b}$$

$$\prod_{\substack{r,k=1\\ \Delta_{0r}<-\Delta_k}}^{h} \prod_{m=1}^{-(\Delta_{0r}+\Delta_k)} (\bar{\alpha}_r-\beta_k+m), \tag{28c}$$

$$\prod_{\substack{r,k=1\\ \Delta_{0r}<\Delta_k}}^{h} \prod_{m=1}^{\Delta_k-\Delta_{0r}} (\bar{\alpha}_r-\bar{\beta}_k+m), \tag{28d}$$

where the roots β_k, $\bar{\beta}_k$, α_r, $\bar{\alpha}_r$ are given by Eqs. (13) and (16), respectively.

So far we have not applied conditions (ii) and (iii) of Eq. (27). However, it turns out that for the orthogonal groups these latter conditions are redundant. In fact (see the Appendix) condition (ii) of Eq. (27) (together with all permutational symmetries) yields precisely the set of factors

(28d) while condition (iii) (together with permutational symmetries) yields precisely the set of factors (28c). Thus there are no new factors to be obtained by further examination of Eq. (27).

It is our aim now to demonstrate that the numerator polynomial of Eq. (24) is given by the products of Eq. (28): i.e.,

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = \prod_{\substack{r,k=1\\ \Delta_{0r}>\Delta_k}}^{h} \prod_{m=1}^{\Delta_{0r}-\Delta_k} (\alpha_r - \beta_k + m)$$
$$\times \prod_{\substack{r,k=1\\ \Delta_{0r}>-\Delta_k}}^{h} \prod_{m=1}^{\Delta_{0r}+\Delta_k} (\alpha_r - \bar{\beta}_k + m)$$
$$\times \prod_{\substack{r,k=1\\ \Delta_{0r}<-\Delta_k}}^{h} \prod_{m=1}^{-(\Delta_{0r}+\Delta_k)} (\bar{\alpha}_r - \beta_k + m)$$
$$\times \prod_{\substack{r,k=1\\ \Delta_{0r}<\Delta_k}}^{h} \prod_{m=1}^{\Delta_k-\Delta_{0r}} (\bar{\alpha}_r - \bar{\beta}_k + m). \qquad (29)$$

We note that Eq. (29) determines a monic polynomial function, which satisfies the Weyl group symmetries of Eq. (25) as required. It remains to demonstrate that the total number of factors on the right-hand side (rhs) of Eq. (29) is equal to the degree of the polynomial function (24): it is given by the integer $r_\lambda + r_{\lambda_0}$ with r_λ and r_{λ_0} as in Eq. (26).

In view of Weyl group symmetry it suffices to consider the semimaximal case: i.e., $\Delta = \lambda$, $\Delta_0 = \lambda_0$ in Eq. (29). In such a case we see that the total number of factors on the rhs of (29) is given by

$$N = \sum_{\substack{r,k=1\\ \lambda_{0r}>\lambda_k}}^{h} (\lambda_{0r} - \lambda_k) + \sum_{r,k=1}^{h} (\lambda_{0r} + \lambda_k)$$
$$+ \sum_{\substack{r,k=1\\ \lambda_{0r}>\lambda_k}}^{h} (\lambda_k - \lambda_{0r})$$
$$= \sum_{r=1}^{h} \left\{ \sum_{\substack{k=1\\ \lambda_k<\lambda_{0r}}}^{h} + h - \sum_{\substack{k=1\\ \lambda_k>\lambda_{0r}}}^{h} \right\} \lambda_{0r}$$
$$+ \sum_{k=1}^{h} \left\{ - \sum_{\substack{r=1\\ \lambda_{0r}>\lambda_k}}^{h} + h + \sum_{\substack{r=1\\ \lambda_{0r}<\lambda_k}}^{h} \right\} \lambda_k$$
$$= \sum_{r=1}^{h} (n-2r)\lambda_{0r} + \sum_{k=1}^{h} (n+2-2k)\lambda_k$$
$$= r_{\lambda_0} + r_\lambda,$$

as required, where we have applied the betweenness conditions of Eq. (20). Thus we have proved our assertion that the numerator polynomials are given by Eq. (29).

Following our $U(n)$ derivation we may express the denominator polynomials of I, Eq. (62), in terms of the characteristic roots β_k and α_r by noting that if $\alpha \in \Phi^+$, $\alpha_0 \in \Phi_0^+$, then

$$\langle \mu + \delta, \alpha \rangle = 2(\mu + \delta, \alpha) = \beta_r - \bar{\beta}_r, \quad \alpha = \epsilon_r,$$
$$\langle \mu + \delta, \alpha \rangle = (\mu + \delta, \alpha)$$
$$= \begin{cases} \beta_r - \beta_k, & \alpha = \epsilon_r - \epsilon_k, \\ \beta_r - \bar{\beta}_k, & \alpha = \epsilon_r + \epsilon_k, \end{cases} \quad r \neq k,$$
$$\langle \mu_0 + \delta_0, \alpha_0 \rangle = (\mu_0 + \delta_0, \alpha_0)$$
$$= \begin{cases} \alpha_r - \alpha_k, & \alpha_0 = \epsilon_{0r} - \epsilon_{0k}, \\ \alpha_r - \bar{\alpha}_k, & \alpha_0 = \epsilon_{0r} + \epsilon_{0k}, \end{cases} \quad r \neq k.$$

Substituting this into I, Eq. (62), we obtain the following expressions for our denominator polynomials:

$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix} = \delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix},$$

where

$$\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix} = \prod_{\substack{r,k=1\\ \Delta_r>\Delta_k}}^{h} \prod_{m=1}^{\Delta_r-\Delta_k} (\beta_r - \beta_k + m - 1)$$
$$\times \prod_{\substack{r<k\\ \Delta_r>-\Delta_k}}^{h} \prod_{m=1}^{\Delta_r+\Delta_k} (\beta_r - \bar{\beta}_k + m - 1)$$
$$\times \prod_{\substack{r<k\\ \Delta_r<-\Delta_k}}^{h} \prod_{m=1}^{-(\Delta_r+\Delta_k)} (\bar{\beta}_r - \beta_k + m - 1) \qquad (30a)$$

and

$$\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix} = \prod_{\substack{r,k=1\\ \Delta_{0r}>\Delta_{0k}}}^{h} \prod_{m=1}^{\Delta_{0r}-\Delta_{0k}} (\alpha_r - \alpha_k + m)$$
$$\times \prod_{\substack{r<k\\ \Delta_{0r}>-\Delta_{0k}}}^{h} \prod_{m=1}^{\Delta_{0r}+\Delta_{0k}} (\alpha_r - \bar{\alpha}_k + m)$$
$$\times \prod_{\substack{r<k\\ \Delta_{0r}<-\Delta_{0k}}}^{h} \prod_{m=1}^{-(\Delta_{0r}+\Delta_{0k})} (\bar{\alpha}_r - \alpha_k + m). \qquad (30b)$$

Thus our optimal RWF's for the imbedding $O(2h+1) \supset O(2h)$ are given by the rational polynomial functions

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0) = C_{\lambda,\lambda_0} \frac{\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)}{\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)}, \qquad (31)$$

where the numerator and denominator polynomials are given by Eqs. (29) and (30), respectively, where it is understood that the characteristic roots β_k and α_r are given by [cf. Eqs. (13) and (16)]

$$\beta_k = \mu_k + n - k, \quad \alpha_r = \mu_{0r} + n - 1 - r,$$
$$\bar{\beta}_k = n - 1 - \beta_k = k - 1 - \mu_k,$$
$$\bar{\alpha}_r = n - 2 - \alpha_r = r - 1 - \mu_{0r},$$
$$r,k = 1,\ldots,h. \qquad (32)$$

We have in particular for the semimaximal case [i.e., $\Delta = \lambda$, $\Delta_0 = \lambda_0$ in Eq. (31)] the results

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix} = \prod_{r<k}^{h} \prod_{m=1}^{\lambda_{0r}-\lambda_k} (\alpha_r - \beta_k + m)$$
$$\times \prod_{k<r}^{h} \prod_{m=1}^{\lambda_k-\lambda_{0r}} (\bar{\alpha}_r - \bar{\beta}_k + m)$$

$$\times \prod_{r,k=1}^{h} \prod_{m=1}^{\lambda_{0r}+\lambda_k} (\alpha_r - \bar{\beta}_k + m),$$

$$\delta\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix} = \delta_1\begin{pmatrix} \lambda \\ \lambda \end{pmatrix} \delta_2\begin{pmatrix} \lambda_0 \\ \lambda_0 \end{pmatrix}, \tag{33a}$$

where

$$\delta_1\begin{pmatrix} \lambda \\ \lambda \end{pmatrix} = \prod_{r<k}^{h} \prod_{m=1}^{\lambda_r - \lambda_k} (\beta_r - \beta_k + m - 1)$$

$$\times \prod_{r<k}^{h} \prod_{m=1}^{\lambda_r+\lambda_k} (\beta_r - \bar{\beta}_k + m - 1), \tag{33b}$$

$$\delta_2\begin{pmatrix} \lambda_0 \\ \lambda_0 \end{pmatrix} = \prod_{r<k}^{h} \left\{ \prod_{m=1}^{\lambda_{0r}-\lambda_{0k}} (\alpha_r - \alpha_k + m) \right.$$

$$\left. \times \prod_{m=1}^{\lambda_{0r}+\lambda_{0k}} (\alpha_r - \bar{\alpha}_k + m) \right\}. \tag{33c}$$

It remains now to determine the numerical constant C_{λ,λ_0} of Eq. (31). To this end we note that I, Eq. (47), implies the result

$$1 = \rho\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}(0|0) = C_{\lambda,\lambda_0} \left. \frac{\eta\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}}{\delta\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}} \right|_{(\mu|\mu_0)=(0|0)},$$

from which we obtain [cf. I, Eq. (86a)]

$$C_{\lambda,\lambda_0} = \left. \frac{\delta\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}}{\eta\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix}} \right|_{(\mu|\mu_0)=(0|0)}. \tag{34a}$$

Substituting $\mu_k = \mu_{0r} = 0$ into Eqs. (32) and (33), we thereby obtain [cf. I, Eq. (86)]

$$C_{\lambda,\lambda_0} = \prod_{r<k}^{h} \frac{(\lambda_r - \lambda_k + k - r - 1)!(\lambda_{0r} + \lambda_{0k} + n - r - k)!}{(\lambda_{0r} - \lambda_k + k - r - 1)!(\lambda_r + \lambda_{0k} + n - r - k)!} \prod_{r<k}^{h} \frac{(\lambda_{0r} - \lambda_{0k} + k - r)!(\lambda_r + \lambda_k + n - r - k)!}{(\lambda_r - \lambda_{0k} + k - r)!(\lambda_{0r} + \lambda_k + n - r - k)!}. \tag{34b}$$

IV. OPTIMAL RWC's FOR O(2*h*)⊃O(2*h* −1)

It is our aim in this section to determine all optimal $O(n+1)$: $O(n)$ RWC's for the case where $n = 2h - 1$ is odd. In this case we recall from I, Sec. V, that the numerator polynomial of Eq. (24) is to determine a polynomial in the characteristic roots β_k, α_r of degree $r_\lambda + r_{\lambda_0}$, where the integers r_λ and r_{λ_0} are now given by [see Eq. (4) and cf. Eq. (26)]

$$r_\lambda = \sum_{\alpha>0} \langle \lambda, \alpha \rangle = 2(\lambda, \delta) = \sum_{k=1}^{h} (n + 1 - 2k)\lambda_k,$$

$$r_{\lambda_0} = \sum_{\alpha_0>0} \langle \lambda_0, \alpha_0 \rangle = 2(\lambda_0, \delta_0) + \sum_{r=1}^{h-1} \lambda_{0r}$$

$$= \sum_{r=1}^{h-1} (n + 1 - 2r)\lambda_{0r} \tag{35}$$

where we have used Eqs. (3) and (15).

Following our previous derivation suppose that $(\mu|\mu_0) \in \mathscr{L}$ is a lexical weight. We then deduce, as before, that the numerator polynomial of Eq. (24) vanishes if the weight $(\mu + \Delta|\mu_0 + \Delta_0)$ is nonlexical. Hence we deduce, from Eq. (20), a vanishing contribution whenever the following situations occur:

$$\begin{aligned} &\text{(i) } (\mu_0 + \Delta_0)_r > (\mu + \Delta)_r, \\ & \qquad\qquad\qquad\qquad\qquad r = 1,\dots,h-1, \\ &\text{(ii) } (\mu + \Delta)_{r+1} > (\mu_0 + \Delta_0)_r, \\ &\text{(iii) } -(\mu + \Delta)_h > (\mu_0 + \Delta_0)_{h-1}. \end{aligned} \tag{36}$$

In case (i) we necessarily have

$$(\mu_0 + \Delta_0)_r = (\mu + \Delta)_r + m, \quad \text{for some } m \text{ in } \mathbf{Z}^+,$$

which, in view of the lexicality of the weight $(\mu|\mu_0)$, can only occur if $\Delta_{0r} > \Delta_r$. In such a case we deduce divisibility of the numerator polynomial (24) by the set of factors

$$\begin{aligned} &(\mu_0 + \Delta_0)_r - (\mu + \Delta)_r - m \\ &\quad = \alpha_r - \beta_r + \Delta_{0r} - \Delta_r - m + 1, \\ &\quad m = 1,\dots,\Delta_{0r} - \Delta_r, \end{aligned}$$

or equivalently by factors

$$\alpha_r - \beta_r + m, \quad m = 1,\dots,\Delta_{0r} - \Delta_r.$$

Thus we deduce divisibility of the numerator polynomial (24) by the set of factors

$$\prod_{m=1}^{\Delta_{0r}-\Delta_r} (\alpha_r - \beta_r + m), \quad \Delta_{0r} > \Delta_r.$$

Applying the Weyl group symmetries of Eq. (25) as before, we then deduce (by considering permutational symmetries only) divisibility of the numerator polynomial (24) by the set of factors

$$\prod_{\substack{r=1 \\ \Delta_{0r}>\Delta_k}}^{h-1} \prod_{k=1}^{h} \prod_{m=1}^{\Delta_{0r}-\Delta_k} (\alpha_r - \beta_k + m). \tag{37a}$$

If now we take into account that the Weyl groups of $O(n+1)$ and $O(n)$ also contain sign changes of the weight components we then deduce, from the factors (37a), divisibility of the numerator polynomial (24) by the additional factors

$$\prod_{\substack{r=1 \\ \Delta_{0r}>-\Delta_k}}^{h-1} \prod_{k=1}^{h} \prod_{m=1}^{\Delta_{0r}+\Delta_k} (\alpha_r - \bar{\beta}_k + m), \tag{37b}$$

$$\prod_{r=1}^{h-1}\prod_{\substack{k=1\\ \Delta_{0r}<-\Delta_k}}^{h}\prod_{m=1}^{-(\Delta_{0r}+\Delta_k)}(\bar{\alpha}_r-\beta_k+m), \tag{37c}$$

$$\prod_{r=1}^{h-1}\prod_{\substack{k=1\\ \Delta_{0r}<\Delta_k}}^{h}\prod_{m=1}^{\Delta_k-\Delta_{0r}}(\bar{\alpha}_r-\bar{\beta}_k+m). \tag{37d}$$

As before, it turns out that conditions (ii) and (iii) of Eq. (36) are redundant. In fact (cf. the Appendix) condition (ii) of Eq. (36) (together with all permutational symmetries) yields precisely the set of factors (37d) while condition (iii) (together with all permutational symmetries) yields precisely the set of factors (37c). So there are no new factors to be obtained by further examination of Eq. (36).

It is our aim now to demonstrate that the numerator polynomial of Eq. (24) is given by the products of Eq. (37): i.e.,

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\prod_{r=1}^{h-1}\prod_{\substack{k=1\\ \Delta_{0r}>\Delta_k}}^{h}\prod_{m=1}^{\Delta_{0r}-\Delta_k}(\alpha_r-\beta_k+m)$$

$$\times\prod_{\substack{r=1\\ \Delta_{0r}>-\Delta_k}}^{h-1}\prod_{k=1}^{h}\prod_{m=1}^{\Delta_{0r}+\Delta_k}(\alpha_r-\bar{\beta}_k+m)$$

$$\times\prod_{\substack{r=1\\ \Delta_{0r}<-\Delta_k}}^{h-1}\prod_{k=1}^{h}\prod_{m=1}^{-(\Delta_{0r}+\Delta_k)}(\bar{\alpha}_r-\beta_k+m)$$

$$\times\prod_{r=1}^{h-1}\prod_{\substack{k=1\\ \Delta_{0r}<\Delta_k}}^{h}\prod_{m=1}^{\Delta_k-\Delta_{0r}}(\bar{\alpha}_r-\bar{\beta}_k+m). \tag{38}$$

We note that Eq. (38) determines a monic polynomial function, which satisfies the Weyl group symmetries of Eq. (25) as required. It remains to demonstrate that the total number of factors on the rhs of Eq. (38) is equal to the degree of the polynomial function (24): in this case it is given by the integer $r_\lambda+r_{\lambda_0}$ with r_λ and r_{λ_0} as in Eq. (35).

In view of Weyl group symmetry it suffices to consider the semimaximal case: i.e., $\Delta=\lambda$, $\Delta_0=\lambda_0$ in Eq. (38). In such a case we see that the total number of factors on the rhs of Eq. (38) is given by

$$N=\sum_{r=1}^{h-1}\sum_{\substack{k=1\\ \lambda_{0r}>\lambda_k}}^{h}(\lambda_{0r}-\lambda_k)+\sum_{r=1}^{h-1}\sum_{\substack{k=1\\ \lambda_{0r}+\lambda_k>0}}^{h}(\lambda_{0r}+\lambda_k)$$

$$+\sum_{r=1}^{h-1}\sum_{\substack{k=1\\ \lambda_{0r}<\lambda_k}}^{h}(\lambda_k-\lambda_{0r})$$

$$=\sum_{r=1}^{h-1}\left\{\sum_{\substack{k=1\\ \lambda_k<\lambda_{0r}}}^{h}+h-\sum_{\substack{k=1\\ \lambda_k>\lambda_{0r}}}^{h}\right\}\lambda_{0r}$$

$$+\sum_{k=1}^{h}\left\{-\sum_{\substack{r=1\\ \lambda_{0r}>\lambda_k}}^{h-1}+h-1+\sum_{\substack{r=1\\ \lambda_{0r}<\lambda_k}}^{h-1}\right\}\lambda_k$$

$$=\sum_{r=1}^{h-1}(n+1-2r)\lambda_{0r}+\sum_{k=1}^{h}(n+1-2k)\lambda_k$$

$$=r_{\lambda_0}+r_\lambda,$$

as required, where we have applied the betweenness conditions of Eq. (20). This is enough to establish the result that the required numerator polynomials are given by Eq. (38).

Following our previous derivation we may express the denominator polynomial of I, Eq. (62), in terms of the characteristic roots β_k and α_r by noting if $\alpha\in\Phi^+$, $\alpha_0\in\Phi_0^+$, then

$$\langle\mu+\delta,\alpha\rangle=(\mu+\delta,\alpha)=\begin{cases}\beta_r-\beta_k, & \alpha=\epsilon_r-\epsilon_k,\\ \beta_r-\bar{\beta}_k, & \alpha=\epsilon_r+\epsilon_k,\end{cases}$$

$$\langle\mu_0+\delta_0,\alpha_0\rangle=2(\mu_0+\delta_0,\alpha_0)=\alpha_r-\bar{\alpha}_r,\quad \alpha_0=\epsilon_{0r},$$

$$\langle\mu_0+\delta_0,\alpha_0\rangle=(\mu_0+\delta_0,\alpha_0)$$

$$=\begin{cases}\alpha_r-\alpha_k, & \alpha_0=\epsilon_{0r}-\epsilon_{0k},\\ \alpha_r-\bar{\alpha}_k, & \alpha_0=\epsilon_{0r}+\epsilon_{0k}.\end{cases}$$

Substituting this into I, Eq. (62), we obtain the following expressions for our denominator polynomials [cf. Eq. (30)]:

$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}=\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix},$$

where

$$\delta_1\begin{pmatrix}\lambda\\ \Delta\end{pmatrix}=\prod_{\substack{r,k=1\\ \Delta_r>\Delta_k}}^{h}\prod_{m=1}^{\Delta_r-\Delta_k}(\beta_r-\beta_k+m-1)$$

$$\times\prod_{\substack{r<k\\ \Delta_r>-\Delta_k}}^{h}\prod_{m=1}^{\Delta_r+\Delta_k}(\beta_r-\bar{\beta}_k+m-1)$$

$$\times\prod_{\substack{r<k\\ \Delta_r<-\Delta_k}}^{h}\prod_{m=1}^{-(\Delta_r+\Delta_k)}(\bar{\beta}_r-\beta_k+m-1), \tag{39a}$$

$$\delta_2\begin{pmatrix}\lambda_0\\ \Delta_0\end{pmatrix}=\prod_{\substack{r,k=1\\ \Delta_{0r}>\Delta_{0k}}}^{h-1}\prod_{m=1}^{\Delta_{0r}-\Delta_{0k}}(\alpha_r-\alpha_k+m)$$

$$\times\prod_{\substack{r<k\\ \Delta_{0r}>-\Delta_{0k}}}^{h-1}\prod_{m=1}^{\Delta_{0r}+\Delta_{0k}}(\alpha_r-\bar{\alpha}_k+m)$$

$$\times\prod_{\substack{r<k\\ \Delta_{0r}<-\Delta_k}}^{h-1}\prod_{m=1}^{-(\Delta_{0r}+\Delta_{0k})}(\bar{\alpha}_r-\alpha_k+m). \tag{39b}$$

Thus our optimal RWF's for the imbedding $O(2h)\supset O(2h-1)$ are given by the rational polynomial functions

$$\rho\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)=C_{\lambda,\lambda_0}\frac{\eta\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)}{\sigma\begin{pmatrix}\lambda & \lambda_0\\ \Delta & \Delta_0\end{pmatrix}(\mu|\mu_0)}, \tag{40}$$

where the numerator and denominator polynomials are given by Eqs. (39) and (40), respectively, it being understood that the characteristic roots β_k and α_r are given by [cf. Eqs. (13) and (16)]

$$\beta_k=\mu_k+n-k,\quad \alpha_r=\mu_{0r}+n-1-r,$$
$$\bar{\beta}_k=k-1-\mu_k,\quad \bar{\alpha}_r=r-1-\mu_{0r},$$
$$r=1,\ldots,h-1,\quad k=1,\ldots,h. \tag{41}$$

We have in particular for the semimaximal case [i.e., $\Delta=\lambda$, $\Delta_0=\lambda_0$ in Eq. (40)]

$$\eta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix} = \prod_{r<k}^{h}\prod_{m=1}^{\lambda_{0r}-\lambda_k}(\alpha_r - \beta_k + m)$$
$$\times\prod_{k<r}^{h-1}\prod_{m=1}^{\lambda_k-\lambda_{0r}}(\bar\alpha_r - \bar\beta_k + m) \qquad (42a)$$
$$\times\prod_{r=1}^{h-1}\prod_{k=1}^{h}\prod_{m=1}^{\lambda_{0r}+\lambda_k}(\alpha_r - \bar\beta_k + m),$$
$$\delta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix} = \delta_1\begin{pmatrix}\lambda\\ \lambda\end{pmatrix}\delta_2\begin{pmatrix}\lambda_0\\ \lambda_0\end{pmatrix},$$
where
$$\delta_1\begin{pmatrix}\lambda\\ \lambda\end{pmatrix} = \prod_{r<k}^{h}\left\{\prod_{m=1}^{\lambda_r-\lambda_k}(\beta_r - \beta_k + m - 1)\right.$$
$$\left.\times\prod_{m=1}^{\lambda_r+\lambda_k}(\beta_r - \bar\beta_k + m - 1)\right\}, \qquad (42b)$$
$$\delta_2\begin{pmatrix}\lambda_0\\ \lambda_0\end{pmatrix} = \prod_{r<k}^{h-1}\prod_{m=1}^{\lambda_{0r}-\lambda_{0k}}(\alpha_r - \alpha_k + m)$$
$$\times\prod_{r<k}^{h-1}\prod_{m=1}^{\lambda_{0r}+\lambda_{0k}}(\alpha_r - \bar\alpha_k + m). \qquad (42c)$$

It remains now to determine the numerical constant C_{λ,λ_0} of Eq. (40), which is given by [cf. Eq. (34a)]

$$C_{\lambda,\lambda_0} = \frac{\delta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}}{\eta\begin{pmatrix}\lambda & \lambda_0\\ \lambda & \lambda_0\end{pmatrix}}\Bigg|_{(\mu|\mu_0)=(0|0)}.$$

Substituting $\mu_k = \mu_{0r} = 0$ into Eqs. (41) and (42), we thereby obtain the result [cf. Eq. (34b)]

$$C_{\lambda,\lambda_0} = \prod_{r<k}^{h}\frac{(\lambda_r - \lambda_k + k - r - 1)!(\lambda_r + \lambda_k + n - r - k)!}{(\lambda_{0r} - \lambda_k + k - r - 1)!(\lambda_{0r} + \lambda_k + n - r - k)!}\prod_{r<k}^{h-1}\frac{(\lambda_{0r} - \lambda_{0k} + k - r)!(\lambda_{0r} + \lambda_{0k} + n - r - k)!}{(\lambda_r - \lambda_{0k} + k - r)!(\lambda_r + \lambda_{0k} + n - r - k)!}. \qquad (43)$$

V. EXAMPLES

We consider here some interesting examples of our previous formulas where we have found it convenient, as before, to consider the cases n even and n odd separately.

A. RWC's for O(2h+1)⊃O(2h)

(i) In the special case $h = 1$, all O(2) Wigner coefficients are trivially unity, hence the O(3): O(2) RWC's correspond to the full O(3) Wigner coefficients. In this case our optimal RWC's are given by

$$\rho\begin{pmatrix}l & m\\ \pm l & m\end{pmatrix}(l'|m') = \left|\left\langle\begin{matrix}l'\pm l\\ m+m'\end{matrix}\right|\begin{matrix}l\\ m\end{matrix};\begin{matrix}l'\\ m'\end{matrix}\right\rangle\right|^2 \qquad (44a)$$

$$= C_{l,m}\frac{\eta\begin{pmatrix}l & m\\ \pm l & m\end{pmatrix}(l'|m')}{\delta\begin{pmatrix}l & m\\ \pm l & m\end{pmatrix}(l'|m')}, \qquad (44b)$$

where, according to formulas (29)–(34), we have

$$\eta\begin{pmatrix}l & m\\ l & m\end{pmatrix} = \prod_{k=1}^{l-m}(\bar\alpha_1 - \bar\beta_1 + k)\prod_{k=1}^{l+m}(\alpha_1 - \bar\beta_1 + k),\quad \eta\begin{pmatrix}l & m\\ -l & m\end{pmatrix} = \prod_{k=1}^{l-m}(\bar\alpha_1 - \beta_1 + k)\prod_{k=1}^{l+m}(\alpha_1 - \beta_1 + k),$$
$$\delta_1\begin{pmatrix}l\\ l\end{pmatrix} = \prod_{k=1}^{2l}(\beta_1 - \bar\beta_1 + k - 1),\quad \delta_2\begin{pmatrix}m\\ m\end{pmatrix} = 1,\quad \delta_1\begin{pmatrix}l\\ -l\end{pmatrix} = \prod_{k=1}^{2l}(\bar\beta_1 - \beta_1 + k - 1). \qquad (45)$$

In this case our O(3) and O(2) roots are given by[3] $\beta_1 = l' + 1, \bar\beta_1 = -l', \alpha_1 = m', \bar\alpha_1 = -m'$. Substituting this into Eqs. (44) and (45) above, noting that the constant $C_{l,m}$ of Eq. (34) in this case is given by

$$C_{l,m} = (2l)!/(l-m)!(l+m)!,$$

we obtain immediately the results (cf. Ref. 12)

$$\rho\begin{pmatrix}l & m\\ l & m\end{pmatrix}(l'|m') = \frac{(2l)!(2l')!(l+l'-m-m')!(l+l'+m+m')!}{(l-m)!(l'-m')!(l+m)!(l'+m')!(2l+2l')!},$$
$$\rho\begin{pmatrix}l & m\\ -l & m\end{pmatrix}(l'|m') = \frac{(l'+m')!(l'-m')!(2l'+1-2l)!(2l)!}{(l+m)!(l-m)!(l'+m'-l+m)!(l'-l-m'-m)!(2l'+1)!}, \qquad (46)$$

where it is understood that if any of the terms in factorials are negative then the quantity vanishes. Formula (46) enables an evaluation of the optimal O(3) Wigner coefficients of Eq. (44a).

(ii) Suppose $\lambda = \Lambda_1 = \epsilon_1$ is the highest weight for the fundamental vector representation of $O(n+1)$ and $\lambda_0 = 0$ [corresponding to the trivial representation of $O(n)$]. We then obtain the fundamental RWF's $\rho\begin{pmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{pmatrix}$ [resp. $\rho\begin{pmatrix}\epsilon_1 & 0\\ -\epsilon_k & 0\end{pmatrix}$], denoted $\overline{C}_k$ (resp. C_k) in Ref. 3. We find it useful in this case to define fundamental weights ϵ_k, for $k > h$, according to

$$\epsilon_{n+2-k} = -\epsilon_k, \tag{47}$$

with $\epsilon_{h+1} = 0$ for $n+1 = 2h+1$ odd [which agrees with Eq. (47)]. Application of our previous formulas immediately implies the result

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{pmatrix} = \overline{C}_k = 2(\bar{\beta}_k - \beta_k)^{-1}(\bar{\beta}_k - \beta_k - 1)^{-1} \prod_{\substack{p=1\\ \neq k}}^{h} (\bar{\beta}_k - \beta_p)^{-1}(\bar{\beta}_k - \bar{\beta}_p)^{-1} \prod_{r=1}^{h} (\bar{\beta}_k - \alpha_r - 1)(\bar{\beta}_k - \bar{\alpha}_r - 1),$$

$$k \neq h+1,$$

where we have adopted the notation [cf. Eq. (13)]

$$\beta_{n+2-k} = \bar{\beta}_k, \quad k = 1,\ldots,n+1.$$

Using the result

$$\bar{\beta}_k - \beta_k - 1 = 2(\bar{\beta}_k - \beta_{h+1}), \quad k \neq h+1, \tag{48}$$

we may more usefully write

$$\overline{C}_k = C_{n+2-k} = \prod_{p\neq k}^{n+1} (\bar{\beta}_k - \bar{\beta}_p)^{-1} \prod_{r=1}^{n} (\bar{\beta}_k - \bar{\alpha}_r - 1), \quad k \neq h+1. \tag{49a}$$

To obtain the RWF's $C_k = \rho\begin{pmatrix}\epsilon_1 & 0\\ -\epsilon_k & 0\end{pmatrix}$, we simply interchange β_k and $\bar{\beta}_k$ in formula (49a) to give

$$C_k = \overline{C}_{n+2-k} = \prod_{p\neq k}^{n+1} (\beta_k - \beta_p)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r - 1), \quad k \neq h+1. \tag{49b}$$

Formulas (49) above, which in fact hold for all $k = 1,\ldots,n+1$, agree with Eq. (12) of Ref. 3 as required. We remark that Eq. (49b) determines the eigenvalues of the $O(n)$-invariants of Eq. (17) in direct analogy with the $U(n)$ case.

It is interesting to note that it can be shown[3,6] that Eq. (49) also holds for the case $k = h+1$ and gives the RWF

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ 0 & 0\end{pmatrix} = \prod_{p\neq h+1}^{n+1} (\beta_{h+1} - \beta_p)^{-1} \prod_{r=1}^{n} (\beta_{h+1} - \alpha_r - 1). \tag{50}$$

This latter RWF is not included in our class of RWF's since it is not optimal (the zero weight is clearly not Weyl group conjugate to the highest weight ϵ_1). This result therefore suggests that our results on $O(n)$ might be extended to a larger class of RWF's.

(iii) Consider now the case $\lambda = \epsilon_1 = \Lambda_1, \lambda_0 = \epsilon_{01} = \Lambda_{01}$ [corresponding to the highest weights of the fundamental vector representations of $O(n+1)$ and $O(n)$]. We then have the RWF's $\rho\begin{pmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{pmatrix}$, where we define fundamental weights ϵ_k ($k = 1,\ldots,n+1$) in accordance with Eq. (47) and similarly define fundamental weights ϵ_{0r} ($r = 1,\ldots,n$): i.e.,

$$\epsilon_{0r} = -\epsilon_{0n+1-r}. \tag{51}$$

Application of our previous formulas immediately yields the result

$$\rho\begin{pmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{pmatrix} = \frac{(\alpha_r - \bar{\beta}_k + 1)(\alpha_r - \bar{\beta}_k + 2)}{(\beta_k - \bar{\beta}_k)(\beta_k - \bar{\beta}_k + 1)} \prod_{\substack{p=1\\ \neq k}}^{h} \frac{(\alpha_r - \beta_p + 1)(\alpha_r - \bar{\beta}_p + 1)}{(\bar{\beta}_k - \beta_p)(\bar{\beta}_k - \bar{\beta}_p)} \prod_{\substack{l=1\\ \neq r}}^{h} \frac{(\bar{\alpha}_l - \bar{\beta}_k + 1)(\alpha_l - \bar{\beta}_k + 1)}{(\alpha_r - \alpha_l + 1)(\alpha_r - \bar{\alpha}_l + 1)},$$

$$h+1 \neq k = 1,\ldots,n+1, \quad r = 1,\ldots,n.$$

Using Eq. (48) and the relations

$$\beta_{h+1} - \bar{\alpha}_r = \alpha_r - \beta_{h+1} + 2, \quad \alpha_r - \bar{\alpha}_r = 2(\alpha_r - \beta_{h+1} + 1),$$

we may write

$$\rho\begin{pmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{pmatrix} = \overline{C}_k M_r (\bar{\beta}_k - \bar{\alpha}_r)^{-1}(\bar{\beta}_k - \bar{\alpha}_r - 1)^{-1}(\bar{\beta}_k - \alpha_r - 1)^{-1}(\bar{\beta}_k - \alpha_r - 2), \tag{52}$$

with $\overline{C}_k$ as in formula (49) and where M_r is given by

$$M_r = (-1)^n \prod_{\substack{l=1\\ \neq r}}^{n} (\bar{\alpha}_r - \bar{\alpha}_l - 1 - \delta_{l,n+1-r})^{-1} \prod_{p=1}^{n+1} (\bar{\beta}_p - \bar{\alpha}_r)$$

$$= (-1)^n \prod_{\substack{l=1\\ \neq r}}^{n} (\alpha_r - \alpha_l + 1 + \delta_{l,n+1-r})^{-1} \prod_{p=1}^{n+1} (\beta_p - \alpha_r + 1 - \delta_{p,h+1}).$$

This last equation determines the squared reduced matrix elements[3,6] of the $O(n)$ vector operator $\psi_i = \{a_{i,n+1}\}$ [where we note that Eq. (15) of Ref. 3 erroneously has the formulas for M_r and $\overline{M}_r$ interchanged].

Equation (52) above agrees with formula (19) of Ref. 3 as required. It is interesting to note[3] that Eq. (52) also extends to the case $k = h + 1$ to yield the (nonoptimal) RWF's [cf. Eq. (50)]

$$\rho\begin{pmatrix} \epsilon_1 & \epsilon_{01} \\ 0 & \epsilon_{0r} \end{pmatrix} = \overline{C}_{h+1} M_r (\beta_{h+1} - \overline{\alpha}_r)^{-1} (\beta_{h+1} - \overline{\alpha}_r - 1)^{-1} (\beta_{h+1} - \alpha_r - 1)^{-1} (\beta_{h+1} - \alpha_r - 2).$$

This last example again illustrates that our pattern calculus for $O(n)$ may be extended to include more general RWF's.

(iv) With regard to our fundamental lexical weights [cf. Eq. (11)] $(\Lambda_s|\Lambda_s^\pm)$; $(\Lambda_r|\Lambda_{0r})$, $(\Lambda_r|\Lambda_{0r-1})$ $(r = 1,...,h)$, we have, in accordance with our prescription, for the maximal cases

$$\rho\begin{pmatrix} \Lambda_s & \Lambda_s^+ \\ \Lambda_s & \Lambda_s^+ \end{pmatrix} = \prod_{r<k}^{h} \left(\frac{\alpha_r - \overline{\beta}_k + 1}{\beta_r - \overline{\beta}_k} \right) \prod_{r<k}^{h} \left(\frac{\alpha_k - \overline{\beta}_r + 1}{\alpha_r - \overline{\alpha}_k + 1} \right),$$

$$\rho\begin{pmatrix} \Lambda_l & \Lambda_{0l} \\ \Lambda_l & \Lambda_{0l} \end{pmatrix} = \prod_{r<k}^{l} \frac{(\alpha_k - \overline{\beta}_r + 1)(\alpha_k - \overline{\beta}_r + 2)}{(\beta_r - \overline{\beta}_k)(\beta_r - \overline{\beta}_k + 1)} \prod_{r<k}^{l} \frac{(\alpha_r - \overline{\beta}_k + 1)(\alpha_r - \overline{\beta}_k + 2)}{(\alpha_r - \overline{\alpha}_k + 2)(\alpha_r - \overline{\alpha}_k + 1)}$$
$$\times \prod_{\substack{r<l \\ k>l}}^{h} \frac{(\alpha_r - \beta_k + 1)(\alpha_r - \overline{\beta}_k + 1)(\overline{\alpha}_k - \overline{\beta}_r + 1)(\alpha_k - \overline{\beta}_r + 1)}{(\beta_r - \beta_k)(\beta_r - \overline{\beta}_k)(\alpha_r - \alpha_k + 1)(\alpha_r - \overline{\alpha}_k + 1)},$$

while for the *minimally tied* cases [cf. I, Eq. (97)] we have

$$\eta\begin{pmatrix} \Lambda_s & \Lambda_s^- \\ \Lambda_s & \Lambda_s^- \end{pmatrix} = \prod_{k=1}^{h} (\overline{\alpha}_h - \overline{\beta}_k + 1) \prod_{r=1}^{h-1} \prod_{k=1}^{h} (\alpha_r - \overline{\beta}_k + 1),$$

$$\delta_1\begin{pmatrix} \Lambda_s \\ \Lambda_s \end{pmatrix} = \prod_{r<k}^{h} (\beta_r - \overline{\beta}_k), \quad \delta_2\begin{pmatrix} \Lambda_s^- \\ \Lambda_s^- \end{pmatrix} = \prod_{r<h} (\alpha_r - \alpha_h + 1) \prod_{r<k}^{h-1} (\alpha_r - \overline{\alpha}_k + 1),$$

$$\eta\begin{pmatrix} \Lambda_l & \Lambda_{0l-1} \\ \Lambda_l & \Lambda_{0l-1} \end{pmatrix} = \prod_{r<l} \left\{ \prod_{k<l} (\alpha_r - \overline{\beta}_k + 1)(\alpha_r - \overline{\beta}_k + 2) \prod_{k>l}^{h} (\alpha_r - \beta_k + 1)(\alpha_r - \overline{\beta}_k + 1) \right\}$$
$$\times \prod_{\substack{r<l \\ k>l}}^{h} (\alpha_k - \overline{\beta}_r + 1)(\overline{\alpha}_k - \overline{\beta}_r + 1),$$

$$\delta_1\begin{pmatrix} \Lambda_l \\ \Lambda_l \end{pmatrix} = \prod_{r<l} \left\{ \prod_{k>l}^{h} (\beta_r - \beta_k)(\beta_r - \overline{\beta}_k) \prod_{k<r} (\beta_k - \overline{\beta}_r)(\beta_k - \overline{\beta}_r + 1) \right\},$$

$$\delta_2\begin{pmatrix} \Lambda_{0l-1} \\ \Lambda_{0l-1} \end{pmatrix} = \prod_{r<l} \left\{ \prod_{k>l}^{h} (\alpha_r - \alpha_k + 1)(\alpha_r - \overline{\alpha}_k + 1) \prod_{k<r} (\alpha_k - \overline{\alpha}_r + 2)(\alpha_k - \overline{\alpha}_r + 1) \right\}.$$

The corresponding constants C_{λ,λ_0} of Eq. (34) are given by

$$C_{\Lambda_s,\Lambda_s^-} = 1, \quad C_{\Lambda_l,\Lambda_{0l-1}} = 2.$$

We remark that for the maximal cases $(\lambda_{0r} = \lambda_r, r = 1,...,h)$ it is easily deduced from Eq. (34) that the constants C_{λ,λ_0} are unity.

In analogy with the unitary groups we call the above RWF's elementary. It can be shown that every $O(n+1):O(n)$ lexical weight can be uniquely expressed as a positive $\mathbb{Z}$-linear combination of elementary lexical weights. Hence the elementary RWF's above will play a role in the $O(n)$ pattern calculus analogous to that played by the elementary RWF's in the $U(n)$ pattern calculus.

B. RWC's for $O(2h)\supset O(2h-1)$

(v) In the special case $h = 2$ we note that a general (multiplicity-free) $O(4)$ Wigner coefficient may be expressed

$$\left\langle \begin{matrix} \mu+\Delta \\ j \\ m'' \end{matrix} \middle| \begin{matrix} \lambda \\ l \\ m \end{matrix} \; ; \; \begin{matrix} \mu \\ l' \\ m' \end{matrix} \right\rangle = \left\langle \begin{matrix} \mu+\Delta \\ j \end{matrix} \middle\| \begin{matrix} \lambda \\ l \end{matrix} \; ; \; \begin{matrix} \mu \\ l' \end{matrix} \right\rangle \left\langle \begin{matrix} j \\ m'' \end{matrix} \middle| \begin{matrix} l \\ m \end{matrix} \; ; \; \begin{matrix} l' \\ m'' \end{matrix} \right\rangle.$$

Since all $O(3)$ Wigner coefficients may be obtained (cf. Ref. 12) it is thus clear that all (multiplicity-free) $O(4)$ Wigner coefficients can be obtained from the $O(4):O(3)$ RWC's. Our formula enables us to evaluate all $O(4):O(3)$ optimal RWC's:

$$\rho\begin{pmatrix} \lambda & l \\ \Delta & \pm l \end{pmatrix}(\mu|l') = \left| \left\langle \begin{matrix} \mu+\Delta \\ l'\pm l \end{matrix} \middle\| \begin{matrix} \lambda \\ l \end{matrix} \; ; \; \begin{matrix} \mu \\ l' \end{matrix} \right\rangle \right|^2, \quad \Delta = \pm(\lambda_1,\lambda_2), \pm(\lambda_2,\lambda_1). \tag{53}$$

For the semimaximal case we have

$$\left|\left\langle\begin{matrix}\mu+\lambda\\ l'+l\end{matrix}\right|\left|\begin{matrix}\lambda\\ l\end{matrix}\ ;\ \begin{matrix}\mu\\ l'\end{matrix}\right\rangle\right|^2=\rho\begin{pmatrix}\lambda & l\\ \lambda & l\end{pmatrix}(\mu|l')=C_{\lambda,l}\,\frac{\eta\begin{pmatrix}\lambda & l\\ \lambda & l\end{pmatrix}(\mu|l')}{\delta\begin{pmatrix}\lambda & l\\ \lambda & l\end{pmatrix}(\mu|l')},$$

where, according to our prescription,

$$C_{\lambda,l}=\frac{(2l+1)!(\lambda_1-\lambda_2)!(\lambda_1+\lambda_2)!}{(l-\lambda_2)!(l+\lambda_2)!(\lambda_1-l)!(\lambda_1+l+1)!},$$

$$\eta\begin{pmatrix}\lambda & l\\ \lambda & l\end{pmatrix}=\prod_{m=1}^{l-\lambda_2}(\alpha_1-\beta_2+m)\prod_{m=1}^{\lambda_1-l}(\bar{\alpha}_1-\bar{\beta}_1+m)\prod_{m=1}^{\lambda_1+l}(\alpha_1-\bar{\beta}_1+m)\prod_{m=1}^{\lambda_2+l}(\alpha_1-\bar{\beta}_2+m),$$

$$\delta_1\begin{pmatrix}\lambda\\ \lambda\end{pmatrix}=\prod_{m=1}^{\lambda_1-\lambda_2}(\beta_1-\beta_2+m-1)\prod_{m=1}^{\lambda_1+\lambda_2}(\beta_1-\bar{\beta}_2+m-1),\quad \delta_2\begin{pmatrix}l\\ l\end{pmatrix}=\prod_{m=1}^{2l}(\alpha_1-\bar{\alpha}_1+m).$$

In this case our O(4) and O(3) characteristic roots are given by[3]

$$\beta_1=\mu_1+2,\quad \beta_2=\mu_2+1,\quad \bar{\beta}_1=-\mu_1,\quad \bar{\beta}_2=1-\mu_2,\quad \alpha_1=l'+1,\quad \bar{\alpha}_1=-l'.$$

Substituting into the above formulas we obtain the result

$$\left|\left\langle\begin{matrix}\lambda+\mu\\ l+l'\end{matrix}\right|\left|\begin{matrix}\lambda\\ l\end{matrix}\ ;\ \begin{matrix}\mu\\ l'\end{matrix}\right\rangle\right|^2$$
$$=\frac{(2l+1)!(2l'+1)!}{(2l'+2l+1)!}\,\frac{(\lambda_1-\lambda_2)!(\lambda_1+\lambda_2)!(\mu_1-\mu_2)!(\mu_1+\mu_2)!}{(l-\lambda_2)!(l+\lambda_2)!(l'-\mu_2)!(l'+\mu_2)!}$$
$$\times\frac{(l'+l-\mu_2-\lambda_2)!(l'+l+\mu_2+\lambda_2)!(\mu_1+\lambda_1-l-l')!(\mu_1+\lambda_1+l+l'+1)!}{(\lambda_1-l)!(\lambda_1+l+1)!(\mu_1-l')!(\mu_1+l'+1)!(\mu_1-\mu_2+\lambda_1-\lambda_2)!(\mu_1+\mu_2+\lambda_1+\lambda_2)!}.$$

Applying Weyl group symmetry we may similarly evaluate all the optimal RWF's of Eq. (53).

(vi) In analogy with example (ii) we consider the special case where $\lambda=\epsilon_1$, $\lambda_0=0$. Then we obtain, in view of our prescription, the RWF's

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{pmatrix}=\bar{C}_k=\frac{1}{2}\prod_{p\neq k}^{h}(\bar{\beta}_k-\beta_p)^{-1}(\bar{\beta}_k-\bar{\beta}_p)^{-1}$$
$$\times\prod_{r=1}^{h-1}(\bar{\beta}_k-\bar{\alpha}_r-1)(\bar{\beta}_k-\alpha_r-1),$$
$$k=1,\dots,h.$$

In this case we may use the result

$$\bar{\beta}_k-\beta_k=2(\bar{\beta}_k-\alpha_h)$$

to give the more suggestive formula

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ \epsilon_k & 0\end{pmatrix}=\bar{C}_k=C_{n+2-k}$$
$$=\prod_{p\neq k}^{n+1}(\bar{\beta}_k-\bar{\beta}_p)^{-1}\prod_{r=1}^{n}(\bar{\beta}_k-\bar{\alpha}_r-1+\delta_{r,h}),\tag{54a}$$

which can be shown to hold for all $k=1,\dots,n+1$, where we adopt convention (47) for defining weights ϵ_k ($k=1,\dots,n+1$). In analogy with formula (49b) we have also the result

$$\rho\begin{pmatrix}\epsilon_1 & 0\\ -\epsilon_k & 0\end{pmatrix}=C_k=\bar{C}_{n+2-k}$$
$$=\prod_{p\neq k}^{n+1}(\beta_k-\beta_p)^{-1}$$
$$\times\prod_{r=1}^{n}(\beta_k-\alpha_r-1+\delta_{r,h}).\tag{54b}$$

Formulas (54) agree with Eq. (12) of Ref. 3 as required, and determine the eigenvalues of the O(n)-invariants of Eq. (17).

(vii) In analogy with example (iii) we consider now the case $\lambda=\epsilon_1$, $\lambda_0=\epsilon_{01}$. In this case we obtain the following formula for our optimal RWF's:

$$\rho\begin{pmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{pmatrix}=\frac{(\bar{\beta}_k-\alpha_r-2)(\bar{\beta}_k-\alpha_r-1)}{(\bar{\alpha}_r-\alpha_r-2)(\bar{\alpha}_r-\alpha_r-1)}$$
$$\times\prod_{p\neq k}^{h}\frac{(\beta_p-\alpha_r-1)(\bar{\beta}_p-\alpha_r-1)}{(\bar{\beta}_k-\bar{\beta}_p(\bar{\beta}_k-\beta_p)}$$
$$\times\prod_{\substack{l=1\\ \neq r}}^{h-1}\frac{(\bar{\beta}_k-\bar{\alpha}_l-1)(\bar{\beta}_k-\alpha_l-1)}{(\alpha_r-\bar{\alpha}_l+1)(\alpha_r-\alpha_l+1)},$$
$$r=1,\dots,h-1,\quad k=1,\dots,h.$$

In this case we may use the results

$$\bar{\beta}_k-\beta_k=2(\bar{\beta}_k-\alpha_h),\quad \bar{\alpha}_r-\alpha_r-1=2(\bar{\alpha}_r-\alpha_h),$$
$$r=1,\dots,h-1,$$

to obtain the more suggestive formula

$$\rho\begin{pmatrix}\epsilon_1 & \epsilon_{01}\\ \epsilon_k & \epsilon_{0r}\end{pmatrix}=\bar{C}_kM_r(\bar{\beta}_k-\bar{\alpha}_r-1)^{-1}(\bar{\beta}_k-\bar{\alpha}_r)^{-1}$$
$$\times(\bar{\beta}_k-\alpha_r-1)^{-1}(\bar{\beta}_k-\alpha_r-2),\quad r\neq k,\tag{55}$$

which holds for all k and $r\neq h$, where $\bar{C}_k$ is given by formula (54) and M_r is given by[3,6]

$$M_r = (-1)^n \prod_{k=1}^{n+1} (\bar\beta_k - \bar\alpha_r) \times \prod_{l \neq r}^{n} (\bar\alpha_r - \bar\alpha_l - 1 - \delta_{l,n+1-r} + \delta_{l,h})^{-1}. \qquad (56)$$

This latter equation determines the squared-reduced matrix elements of the $O(n)$ vector operator $\psi_i = a_{i,n+1}$.

Formula (55) agrees with Eq. (19) of Ref. 3 as required. It is interesting to note the following formula obtained in Ref. 3 for the case $r = h$:

$$\rho\begin{pmatrix} \epsilon_1 & \epsilon_{01} \\ \epsilon_k & 0 \end{pmatrix} = \bar C_k M_h (\bar\beta_k - \alpha_h)^{-2}, \qquad (57)$$

where M_h (the zero shift reduced matrix element squared) is still given by formula (56) (cf. Ref. 3). The RWF of Eq. (57) is nonoptimal [since the zero weight is clearly not conjugate under the Weyl group to the $O(n)$ highest weight ϵ_{01}] and hence formula (57) indicates an extension of the $O(n)$ pattern calculus rules to certain nonoptimal RWF's.

(viii) With regard to our fundamental lexical weights [cf. Eq. (11)] $(\Lambda_s^\pm|\Lambda_s)$, $(\Lambda_r|\Lambda_{0r})$, $(\Lambda_r|\Lambda_{0r-1})$ ($r = 1,\ldots,h-1$) we have, in analogy with example (iv), for the maximal cases

$$\rho\begin{pmatrix} \Lambda_s^+ & \Lambda_s \\ \Lambda_s^+ & \Lambda_s \end{pmatrix} = \prod_{r<k}^{h} \left(\frac{\alpha_r - \bar\beta_k + 1}{\beta_r - \bar\beta_k}\right) \prod_{r<k}^{h} \left(\frac{\alpha_k - \bar\beta_r + 1}{\alpha_r - \bar\alpha_k + 1}\right),$$

$$\rho\begin{pmatrix} \Lambda_s^- & \Lambda_s \\ \Lambda_s^- & \Lambda_s \end{pmatrix} = \prod_{r=1}^{h-1} \left(\frac{\alpha_r - \beta_h + 1}{\beta_r - \beta_h}\right) \prod_{r<k}^{h-1} \left(\frac{\alpha_r - \bar\beta_k + 1}{\beta_r - \bar\beta_k}\right) \times \prod_{r<k}^{h-1} \left(\frac{\alpha_k - \bar\beta_r + 1}{\alpha_r - \bar\alpha_k + 1}\right),$$

$$\rho\begin{pmatrix} \Lambda_l & \Lambda_{0l} \\ \Lambda_l & \Lambda_{0l} \end{pmatrix} = \prod_{r<k}^{l} \frac{(\alpha_r - \bar\beta_k + 2)(\alpha_r - \bar\beta_k + 1)}{(\alpha_r - \bar\alpha_k + 2)(\alpha_r - \bar\alpha_k + 1)} \times \prod_{r<k}^{l} \frac{(\alpha_k - \bar\beta_r + 2)(\alpha_k - \bar\beta_r + 1)}{(\beta_r - \bar\beta_k)(\beta_r - \bar\beta_k + 1)} \times \prod_{\substack{r<l \\ k>l}}^{h} \frac{(\alpha_r - \beta_k + 1)(\alpha_r - \bar\beta_k + 1)}{(\beta_r - \beta_k)(\beta_r - \bar\beta_k)} \times \prod_{\substack{r<l \\ k>l}}^{h-1} \frac{(\bar\alpha_k - \bar\beta_r + 1)(\alpha_k - \bar\beta_r + 1)}{(\alpha_r - \bar\alpha_k + 1)(\alpha_r - \alpha_k + 1)},$$

while for the *minimally tied* cases [cf. I, Eq. (97)] we have

$$\eta\begin{pmatrix} \Lambda_l & \Lambda_{0l-1} \\ \Lambda_l & \Lambda_{0l-1} \end{pmatrix} = \prod_{\substack{r<l \\ k>l}}^{h} (\alpha_r - \beta_k + 1)(\alpha_r - \bar\beta_k + 1) \times \prod_{\substack{k<l \\ r>l}}^{h-1} (\bar\alpha_r - \bar\beta_k + 1)(\alpha_r - \bar\beta_k + 1) \times \prod_{\substack{r<l \\ k<l}} (\alpha_r - \bar\beta_k + 2)(\alpha_r - \bar\beta_k + 1),$$

$$\delta_1\begin{pmatrix} \Lambda_l \\ \Lambda_l \end{pmatrix} = \prod_{\substack{r<l \\ k>l}}^{h} (\beta_r - \beta_k)(\beta_r - \bar\beta_k) \times \prod_{r<k}^{l} (\beta_r - \bar\beta_k)(\beta_r - \bar\beta_k + 1),$$

$$\delta_2\begin{pmatrix} \Lambda_{0l-1} \\ \Lambda_{0l-1} \end{pmatrix} = \prod_{\substack{r<l \\ k>l}}^{h-1} (\alpha_r - \alpha_k + 1)(\alpha_r - \bar\alpha_k + 1) \times \prod_{r<k}^{l-1} (\alpha_r - \bar\alpha_k + 2)(\alpha_r - \bar\alpha_k + 1).$$

For the minimally tied cases we note that the constant C_{λ,λ_0} of Eq. (43) reduces to

$$C_{\Lambda_l,\Lambda_{0l-1}} = \tfrac{1}{2},$$

while for the maximally tied cases (i.e., $\lambda_{0r} = \lambda_r$, $r = 1,\ldots,h-1$) the constant C_{λ,λ_0} of Eq. (43) is easily seen to give unity.

Examples (viii) and (iv), together with the Weyl group symmetries of Eq. (25) [cf. I, Eqs. (34) and (38)] enables all elementary RWF's for the orthogonal groups to be evaluated in analogy with $U(n)$. In particular, since all weights in the fundamental spinor representations are Weyl group conjugate (via reflections only) to the highest weight, this enables all Wigner coefficients to be evaluated for the fundamental spinor representations.

VI. COMPOSITION OF RWF'S

It was shown in I that the optimal RWF's of $U(n)$ may be expressed as a product (or composition) of elementary RWF's (times a known constant C_{λ,λ_0}). Such considerations are likely to be of importance, from the practical point of view, for obtaining the optimum way of expressing a RWC in actual calculations. Moreover such a calculus of RWF's has the advantage that it may be extended to calculate nonoptimal RWF's by manipulating and multiplying elementary RWF's. To this end we conclude in this section with a beginning of a pattern calculus for $O(n)$.

Following the notation of I [see, in particular, Secs. IV and IX], let $(\lambda|\lambda)$ denote a maximal lexical weight of $O(n+1)$: $O(n)$, i.e., the $O(n)$ representation labels take their maximal allowed values

$$\lambda_{0r} = \lambda_r, \quad r = 1,\ldots,h_0 = \begin{cases} h, & n = 2h, \\ h-1, & n = 2h-1. \end{cases}$$

We clearly have the following expansion for the maximal lexical weight $(\lambda|\lambda)$ [cf. Eq. (11)]:

$$(\lambda|\lambda) = \sum_{r=1}^{h-1} (\lambda_r - \lambda_{r+1})(\Lambda_r|\Lambda_{0r}) + 2\lambda_h(\Lambda_s|\Lambda_s^+), \quad n = 2h,$$

$$= \sum_{r=1}^{h-2} (\lambda_r - \lambda_{r+1})(\Lambda_r|\Lambda_{0r}) + (\lambda_h + \lambda_{h-1})(\Lambda_s^+|\Lambda_s) + (\lambda_{h-1} - \lambda_h)(\Lambda_s^-|\Lambda_s), \quad h = 2h-1.$$

We note that all the elementary lexical weights occurring in the above decomposition are maximal. Thus, by repeated application of I, Eq. (51), we obtain the following decomposition of a maximal RWF into elementary RWF's:

$$\rho\begin{pmatrix}\lambda & \lambda \\ \lambda & \lambda\end{pmatrix} = \rho\begin{pmatrix}\Lambda_s & \Lambda_s^+ \\ \Lambda_s & \Lambda_s^+\end{pmatrix}^{2\lambda_h} \circ \prod_{r=1}^{h-1} \rho\begin{pmatrix}\Lambda_r & \Lambda_{0r} \\ \Lambda_r & \Lambda_{0r}\end{pmatrix}^{\lambda_r - \lambda_{r+1}},$$

$$n = 2h,$$

$$= \rho\begin{pmatrix}\Lambda_s^+ & \Lambda_s \\ \Lambda_s^+ & \Lambda_s\end{pmatrix}^{\lambda_{h-1}+\lambda_h} \circ \rho\begin{pmatrix}\Lambda_s^- & \Lambda_s \\ \Lambda_s^- & \Lambda_s\end{pmatrix}^{\lambda_{h-1}-\lambda_h}$$

$$\circ \prod_{r=1}^{h-2} \rho\begin{pmatrix}\Lambda_r & \Lambda_{0r} \\ \Lambda_r & \Lambda_{0r}\end{pmatrix}^{\lambda_r - \lambda_{r+1}}, \quad n = 2h-1. \quad (58)$$

As in the case of $\mathrm{U}(n)$ we note that all RWF's on the rhs are semimaximal and hence commuting so that the ordering is irrelevant. Equation (58) is clearly analogous to I, Eq. (94), obtained for the unitary groups.

Now let $(\lambda|\bar{\lambda})$ be a minimally connected lexical weight; i.e., the $\mathrm{O}(n)$ representation labels take their minimal allowed values

$$\lambda_{0h} = -\lambda_h, \quad \lambda_{0r} = \lambda_{r+1}, \quad r = 1,\ldots,h-1, \quad n = 2h,$$

$$\lambda_{0h-1} = |\lambda_h|, \quad \lambda_{0r} = \lambda_{r+1}, \quad r = 1,\ldots,h-2,$$

$$n = 2h-1.$$

Clearly the fundamental lexical weights $(\Lambda_r|\Lambda_{0r-1})$ $(r = 1,\ldots,h-1)$ together with $(\Lambda_s|\Lambda_s^-)$ [resp. $(\Lambda_s^\pm|\Lambda_s)$] for $n = 2h$ even (resp. $n = 2h-1$ odd) are minimally connected. We note that, with the above definitions, the elementary lexical weights $(\Lambda_s^\pm|\Lambda_s)$, for $n = 2h-1$ odd, are both minimally and maximally connected.

We have the following expansion of a minimally connected lexical weight $(\lambda|\bar{\lambda})$ into elementary minimal weights:

$$(\lambda|\bar{\lambda})$$

$$= \sum_{r=1}^{h-1} (\lambda_r - \lambda_{r+1})(\Lambda_r|\Lambda_{0r-1}) + 2\lambda_h(\Lambda_s|\Lambda_s^-),$$

$$n = 2h,$$

$$= \sum_{r=1}^{h-2} (\lambda_r - \lambda_{r+1})(\Lambda_r|\Lambda_{0r-1})$$

$$+ (\lambda_h + |\lambda_h|)(\Lambda_s^+|\Lambda_s)$$

$$+ (|\lambda_h| - \lambda_h)(\Lambda_s^-|\Lambda_s), \quad n = 2h-1.$$

Thus by repeated application of I, Eq. (50), we may expand a general minimally connected RWF into elementary RWF's according to

$$\rho\begin{pmatrix}\lambda & \bar{\lambda} \\ \lambda & \bar{\lambda}\end{pmatrix}$$

$$= C^\lambda \rho\begin{pmatrix}\Lambda_s & \Lambda_s^- \\ \Lambda_s & \Lambda_s^-\end{pmatrix}^{2\lambda_h} \circ \prod_{r=1}^{h-1} \rho\begin{pmatrix}\Lambda_r & \Lambda_{0r-1} \\ \Lambda_r & \Lambda_{0r-1}\end{pmatrix}^{\lambda_r - \lambda_{r+1}}$$

$$n = 2h,$$

$$= C^\lambda \rho\begin{pmatrix}\Lambda_s^+ & \Lambda_s \\ \Lambda_s^+ & \Lambda_s\end{pmatrix}^{\lambda_h + |\lambda_h|} \circ \rho\begin{pmatrix}\Lambda_s^- & \Lambda_s \\ \Lambda_s^- & \Lambda_s\end{pmatrix}^{|\lambda_h| - \lambda_h}$$

$$\circ \prod_{r=1}^{h-2} \rho\begin{pmatrix}\Lambda_r & \Lambda_{0r-1} \\ \Lambda_r & \Lambda_{0r-1}\end{pmatrix}^{\lambda_r - \lambda_{r+1}} \quad n = 2h-1, \quad (59)$$

for some numerical constant C^λ depending only on the labels $\lambda, \bar{\lambda}$ (and hence λ). It is easily seen, by comparison with examples (iv) and (viii) of Sec. V, that the constant C^λ of Eq. (59) is given by

$$C^\lambda = 2^a C_{\lambda,\bar{\lambda}}, \quad (60a)$$

where $C_{\lambda,\bar{\lambda}}$ is given by Eqs. (34) and (43) and

$$a = \begin{cases} -\sum_{r=1}^{h-1} (\lambda_r - \lambda_{r+1}), & n = 2h, \\ \sum_{r=1}^{h-2} (\lambda_r - \lambda_{r+1}), & n = 2h-1. \end{cases} \quad (60b)$$

Equation (59) is clearly analogous to I, Eq. (97), obtained for the unitary groups. However, unlike in the $\mathrm{U}(n)$ case, the constant C^λ of Eq. (59) is not generally unity. The proof of this result for $\mathrm{U}(n)$ follows because (see I, Sec. IX) the minimally connected $\mathrm{U}(n+1)$: $\mathrm{U}(n)$ lexical weights correspond to semimaximal states whose weights are Weyl group conjugate to the highest weight. However, in the case of the orthogonal groups the maximal state is clearly a GT state but the remaining W-conjugate weight states are not GT states. Nevertheless Eq. (59) may be applied in analogy with I, Eq. (97), for the unitary groups.

We now note that the fundamental lexical weights

$$(\Lambda_r|\Lambda_{0r}) \quad (r = 1,\ldots,h-1),$$

$$(\Lambda_r|\Lambda_{0r-1}) \quad (r = 1,\ldots,h-1),$$

$$(\Lambda_s|\Lambda_s^+), \quad (\Lambda_s|\Lambda_s^-), \quad n = 2h,$$

$$(\Lambda_r|\Lambda_{0r}) \quad (r = 1,\ldots,h-2),$$

$$(\Lambda_r|\Lambda_{0r-1}) \quad (r = 1,\ldots,h-1),$$

$$(\Lambda_s^+|\Lambda_s), \quad (\Lambda_s^-|\Lambda_s), \quad n = 2h-1,$$

form a basis for $H^* \times H_0^*$ with corresponding dual basis

$$(-\epsilon_{r+1}|\epsilon_{0r}) \quad (r = 1,\ldots,h-1),$$

$$(\epsilon_r| - \epsilon_{0r}) \quad (r = 1,\ldots,h-1),$$

$$(\epsilon_h|\epsilon_{0h}), \quad (\epsilon_h| - \epsilon_{0h}), \quad n = 2h,$$

$$(-\epsilon_{r+1}|\epsilon_{0r}) \quad (r = 1,\ldots,h-2),$$

$$(\epsilon_r| - \epsilon_{0r}) \quad (r = 1,\ldots,h-1),$$

$$(\epsilon_h|\epsilon_{0h-1}), \quad (-\epsilon_h|\epsilon_{0h-1}), \quad n = 2h-1,$$

respectively. Thus we may expand an arbitrary $\mathrm{O}(n+1)$: $\mathrm{O}(n)$ lexical weight $(\lambda|\lambda_0)$ into elementary lexical weights according to [cf. I, Eq. (98)]

$$(\lambda|\lambda_0) = \sum_{r=1}^{h-1} (\lambda_r - \lambda_{0r})(\Lambda_r|\Lambda_{0r-1})$$

$$+ \sum_{r=1}^{h-1} (\lambda_{0r} - \lambda_{r+1})(\Lambda_r|\Lambda_{0r})$$

$$+ (\lambda_h + \lambda_{0h})(\Lambda_s|\Lambda_s^+)$$

$$+ (\lambda_h - \lambda_{0h})(\Lambda_s|\Lambda_s^-),$$

$$n = 2h,$$

$$= \sum_{r=1}^{h-1} (\lambda_r - \lambda_{0r})(\Lambda_r|\Lambda_{0r-1})$$

$$+ \sum_{r=1}^{h-2} (\lambda_{0r} - \lambda_{r+1})(\Lambda_r|\Lambda_{0r})$$

$$+ (\lambda_h + \lambda_{0h-1})(\Lambda_s^+|\Lambda_s)$$

$$+ (\lambda_{0h-1} - \lambda_h)(\Lambda_s^- | \Lambda_s),$$
$$n = 2h - 1.$$

We may thus express an arbitrary lexical weight $(\lambda | \lambda_0)$ as a sum of maximally and minimally connected lexical weights according to [cf. I, Eq. (98)]

$$(\lambda | \lambda_0) = (\lambda' | \lambda') + (\lambda'' | \bar{\lambda}''), \tag{61a}$$

where

$$\lambda' = \sum_{r=1}^{h-1} (\lambda_{0r} - \lambda_{r+1})\Lambda_r + (\lambda_h + \lambda_{0h})\Lambda_s, \quad n = 2h,$$
$$= \sum_{r=1}^{h-2} (\lambda_{0r} - \lambda_{r+1})\Lambda_r + (\lambda_h + \lambda_{0h-1})\Lambda_s^+$$
$$+ (\lambda_{0h-1} - \lambda_h)\Lambda_s^-, \quad n = 2h - 1, \tag{61b}$$

and

$$\lambda'' = \sum_{r=1}^{h-1} (\lambda_r - \lambda_{0r})\Lambda_r + (\lambda_h - \lambda_{0h})\Lambda_s, \quad n = 2h,$$
$$= \sum_{r=1}^{h-2} (\lambda_r - \lambda_{0r})\Lambda_r, \quad n = 2h - 1. \tag{61c}$$

Equations (61a)–(61c) above together with I, Eq. (50), immediately implies a decomposition

$$C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''} \rho\begin{pmatrix} \lambda & \lambda_0 \\ \lambda & \lambda_0 \end{pmatrix} = \rho\begin{pmatrix} \lambda' & \lambda' \\ \lambda' & \lambda' \end{pmatrix} \circ \rho\begin{pmatrix} \lambda'' & \bar{\lambda}'' \\ \lambda'' & \bar{\lambda}'' \end{pmatrix}, \tag{62}$$

where the constant on the lhs above is given by

$$C^{\lambda'\lambda''}_{\lambda'\bar{\lambda}''} = \rho\begin{pmatrix} \lambda' & \lambda' \\ \lambda' & \lambda' \end{pmatrix} (\lambda'' | \bar{\lambda}'') = \rho\begin{pmatrix} \lambda'' & \bar{\lambda}'' \\ \lambda'' & \bar{\lambda}'' \end{pmatrix} (\lambda' | \lambda'),$$

which may be evaluated using Eqs. (33) and (42).

Equation (62) is clearly analogous to I, Eq. (99), obtained for the unitary groups. The RWF's $\rho(\begin{smallmatrix} \lambda' & \lambda' \\ \lambda' & \lambda' \end{smallmatrix})$ and $\rho(\begin{smallmatrix} \lambda'' & \bar{\lambda}'' \\ \lambda'' & \bar{\lambda}'' \end{smallmatrix})$ may be decomposed according to Eqs. (58) and (59), respectively, whence Eq. (62), together with Weyl group symmetry, yields a decomposition of an arbitrary $O(n+1):O(n)$ optimal RWF into a commuting product of elementary RWF's in direct analogy with the unitary groups.

ACKNOWLEDGMENT

The author gratefully acknowledges the financial support of an A.R.G.S. research grant.

APPENDIX

It is our aim here to demonstrate that the factors dividing the numerator polynomial (24), deduced from conditions (ii) and (iii) of Eq. (27) (together with all permutational symmetries), yield precisely the set of factors (28d) and (28c), respectively. Similarly conditions (ii) and (iii) of Eq. (36), yield the set of factors (37d) and (37c), respectively.

To see this we deduce, from condition (ii) of Eq. (27), a vanishing contribution to the numerator polynomial (24) whenever the situation

$$(\mu + \Delta)_{r+1} = (\mu_0 + \Delta_0)_r + m$$

occurs for some $m \in \mathbb{Z}^+$ (since, in such a case, the corresponding RWC vanishes). Using the assumed lexicality of the weight $(\mu | \mu_0)$, we see that the possible range of m-values is given by $1 \leqslant m \leqslant \Delta_{r+1} - \Delta_{0r}$. Thus we can only get a vanishing contribution if $\Delta_{r+1} > \Delta_{0r}$, in which case we deduce divisibility of the numerator polynomial by the set of factors

$$\mu_{r+1} + \Delta_{r+1} - \mu_{0r} - \Delta_{0r} - m, \quad m = 1,...,\Delta_{r+1} - \Delta_{0r},$$

or, equivalently, factors

$$\mu_{r+1} - \mu_{0r} + m - 1 = \beta_{r+1} - \alpha_r + m - 1,$$
$$m = 1,...,\Delta_{r+1} - \Delta_{0r}.$$

Thus we deduce divisibility of the numerator polynomial (24) by the set of factors

$$\prod_{m=1}^{\Delta_{r+1} - \Delta_{0r}} (\beta_{r+1} - \alpha_r + m - 1), \quad \Delta_{r+1} > \Delta_{0r}.$$

Taking into account all permutational symmetries on the roots β_k and α_r, we deduce, from the above, divisibility of the numerator polynomial (24) by the set of factors

$$\prod_{\substack{r,k=1 \\ \Delta_k > \Delta_{0r}}}^{h} \prod_{m=1}^{\Delta_k - \Delta_{0r}} (\beta_k - \alpha_r + m - 1). \tag{A1}$$

If we apply condition (iii) of Eq. (27) we similarly deduce divisibility of the numerator polynomial (24) by factors

$$m - 1 - \mu_h - \mu_{0h} = m + n - 2 - \beta_h - \alpha_h,$$
$$m = 1,..., -(\Delta_h + \Delta_{0h}),$$

i.e., we deduce divisibility of the numerator polynomial by the set of factors

$$\prod_{m=1}^{-(\Delta_h + \Delta_{0h})} (m + n - 2 - \beta_h - \alpha_h), \quad \Delta_h + \Delta_{0h} < 0.$$

Applying all permutational symmetries to the roots β_k, α_r we thus deduce divisibility of the numerator polynomial (24) by the additional set of factors

$$\prod_{\substack{r,k=1 \\ \Delta_{0r} + \Delta_k < 0}}^{h} \prod_{m=1}^{-(\Delta_{0r} + \Delta_k)} (m + n - 2 - \beta_k - \alpha_r). \tag{A2}$$

We now note that Eqs. (13) and (16) imply the following relations between the $O(n+1)$ and $O(n)$ roots:

$$\bar{\alpha}_r = n - 2 - \alpha_r, \quad \bar{\beta}_k = n - 1 - \beta_k.$$

Thus we have, for the factors of Eq. (A1),

$$\beta_k - \alpha_r + m - 1$$
$$= (n - 1 - \bar{\beta}_k) - (n - 2 - \bar{\alpha}_r) + m - 1$$
$$= \bar{\alpha}_r - \bar{\beta}_k + m.$$

Thus Eq. (A1) may be alternatively written

$$\prod_{\substack{r,k=1 \\ \Delta_k > \Delta_{0r}}}^{h} \prod_{m=1}^{\Delta_k - \Delta_{0r}} (\bar{\alpha}_r - \bar{\beta}_k + m),$$

which are the factors of Eq. (28d) as required.

Similarly, for the factors of Eq. (A2), we have

$$m + n - 2 - \beta_k - \alpha_r$$
$$= m + (n - 2 - \alpha_r) - \beta_k$$
$$= \bar{\alpha}_r - \beta_k + m,$$

which are the factors of Eq. (28c). In an analogous way it is easily deduced that conditions (ii) and (iii) of Eq. (36) (together with all permutational symmetries) yield the factors (37d) and (37c), respectively.

[1]M. D. Gould, J. Math. Phys. **27**, 1944 (1986).
[2]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967); L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[3]M. D. Gould, J. Math. Phys. **22**, 2376 (1981).
[4]E. Chacon, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13**, 577 (1972).
[5]S. J. Alisauskas, A. A. Jucys, and A. P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[6]M. D. Gould, J. Math. Phys. **21**, 444 (1980); J. Aust. Math. Soc. B **20**, 401 (1978).
[7]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[8]M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967); S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[9]A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971); S. Okubo, J. Math. Phys. **12**, 2382 (1977).
[10]S. A. Edwards, J. Math. Phys. **19**, 164 (1978).
[11]I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).
[12]M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3j and 6j Symbols* (MIT, Cambridge, MA, 1959).

Superunitary reduction of the orthosymplectic supertableaux

Michel Gourdin
Laboratoire de Physique Théorique et Hautes Energies, Université Pierre et Marie Curie, Tour 16-1er étage, 4 place Jussieu, 75230 Paris Cedex 05, France

(Received 2 January 1986; accepted for publication 26 March 1986)

The supertableaux of the orthosymplectic groups OSP$(2\nu|2p)$ and OSP$(2\nu+1|2p)$ are decomposed in sums of supertableaux of the superunitary group SU$(\nu|p)$.

I. INTRODUCTION

The problem of the relation between orthosymplectic and superunitary supergroups, besides its mathematical interest, might be relevant for the study of supersymmetry and supergravity in physics. The oscillator method has been proposed for the construction of unitary irreducible representations of noncompact supergroups[1] and applications to the orthosymplectic supergroups of the form OSP$(m|4)$ have been investigated.[2-4]

The viewpoint chosen in this paper is different. First, we describe the relation between the superalgebra of OSP$(2\nu|2p)$ and its maximal subsuperalgebra of U$(\nu|p)$. Second, we translate the results in the language of supertableaux and we give the decomposition of the orthosymplectic supertableaux into sums of superunitary supertableaux for both cases of irreducible supertableaux and generalized atypical supertableaux. The extension of the results to the orthosymplectic supergroups OSP$(2\nu+1|2p)$ is straightforward.

The method is economical in the sense that we essentially use the properties of supersymmetry of the boxes of a supertableau avoiding tedious algebraic computations. The interpretation of the supertableaux in terms of representations of the corresponding superalgebra is known for the superunitary case[5-7] and the orthosymplectic one.[8-10] The considerations made in this paper exclude the particular case of the supertableaux of OSP$(2|2)$ already discussed in a previous paper.[11]

II. ORTHOSYMPLECTIC LIE SUPERALGEBRA OF OSP$(2\nu|2p)$

The Lie superalgebra of the orthosymplectic group OSP$(2\nu|2p)$ has the rank $\nu+k$ and it is simple. The notation introduced by Kac[12] is $D(\nu,p)$ for $\nu\geqslant 2$ and $C(p+1)$ for $\nu=1$. The sets Δ_0 and Δ_1 of even and odd roots are given by

$$\Delta_0=\{\epsilon_\alpha d_\alpha+\epsilon_\beta d_\beta;\epsilon_i e_i+\epsilon_j e_j,2\epsilon_i e_i\},$$

$$\Delta_1=\{\epsilon_\alpha d_\alpha+\epsilon_i e_i\},$$

with $\alpha<\beta=1,2,\ldots,\nu$, $i<j=1,2,\ldots,p$, and $\epsilon=\pm 1$. The dimensions of the sets Δ_0 and Δ_1 are

$$\dim\Delta_0=2\nu(\nu-1)+2p^2,$$

$$\dim\Delta_1=\Delta\nu p.$$

The even part L_0 of the orthosymplectic superalgebra is the Lie algebra of the direct product $G_0=\mathrm{SO}(2\nu)\otimes\mathrm{Sp}(2p)$. For $\nu\geqslant 2$, L_0 is semisimple, $L_0=D_\nu+C_p$. For $\nu=1$, G_0 has a U(1) factor and $L_0=A_0+C_p$.

The odd part L_1 of the orthosymplectic superalgebra is irreducible and it behaves like a $(2\nu,2p)$ representation of L_0.

The following notations are used for the infinitesimal generators.

(i) For symplectic even generators,

$$B_{ij}\begin{cases} i=j, & \text{Cartan subalgebra}, \quad p,\\ i\neq j, & \text{root } e_i-e_j, \quad p(p-1);\end{cases}$$

$$C_{ij}\begin{cases} i=j, & \text{root } 2e_i, \quad p,\\ i\neq j, & \text{root } e_i+e_j,\quad i<j,\quad \tfrac12 p(p-1);\end{cases}$$

$$D_{ij}\begin{cases} i=j, & \text{root } -2e_i, \quad p,\\ i\neq j, & \text{root } -e_i-e_j,\quad i<j,\quad \tfrac12 p(p-1).\end{cases}$$

(ii) For orthogonal even generators,

$$H_\alpha,\quad \text{Cartan subalgebra},\quad \nu,$$

$$E_{\epsilon_\alpha\alpha\,\epsilon_\beta\beta},\quad \text{root } \epsilon_\alpha d_\alpha+\epsilon_\beta d_\beta,\quad \alpha<\beta,\quad 2\nu(\nu-1).$$

(iii) For odd generators,

$$F_{\epsilon_\alpha\alpha j},\quad \text{root } \epsilon_\alpha d_\alpha+e_j,\quad 2\nu p,$$

$$G_{\epsilon_\alpha\alpha j},\quad \text{root } \epsilon_\alpha d_\alpha-e_j,\quad 2\nu p.$$

III. MAXIMAL SUBALGEBRA U$(\nu|p)$

The maximal subsuperalgebra of the orthosymplectic group OSP$(2\nu|2p)$ is the superalgebra of the superunitary group U$(\nu|p)$ (see Ref. 4). Both superalgebras have the same Cartan subalgebra and therefore the same rank $\nu+p$.

The superalgebra of U$(\nu|p)$ is not simple. It contains a U(1) factor of generator Q_N commuting with all the other generators of U$(\nu|p)$. The superalgebra of the special superunitary group SU$(\nu|p)$ does not contain Q_N and all its generators are supertraceless.

For $\nu\neq p$ the superalgebra of SU$(\nu|p)$ is simple and it is noted $A(\nu-1,p-1)$ by Kac.[12] For $\nu=p$ the center of SU$(p|p)$ is not trivial and we must eliminate this center before to get a simple superalgebra $A(p-1,p-1)$.

The even part l_0 of the superalgebra of U$(\nu|p)$ is the Lie algebra of the direct product $g_0=\mathrm{U}(\nu)\times\mathrm{U}(p)$ with the inclusions

$$\mathrm{U}(\nu)\subset\mathrm{SO}(2\nu),\quad \mathrm{U}(p)\subset\mathrm{Sp}(2p).$$

Both unitary groups U(ν) and U(p) have a U(1) factor of respective generators B_0 and B_S and the structure of l_0 is simply

$$l_0=A_{\nu-1}\oplus A_0\oplus A_{p-1}\oplus A_0.$$

Of course the U(1) generator Q_N previously introduced is a

 0022-2488/86/081980-07$02.50

linear combination of B_0 and B_S commuting with the $U(\nu|p)$ generators.

We first study the inclusion $U(\nu) \subset SO(2\nu)$ and we choose the place of $U(1)$ in the $SO(2\nu)$ space by defining

$$B_0 = \sum_1^{\nu} H_\alpha .$$

Using the commutation relations of B_0 with the orthogonal generators we obtain a Jordan structure for the Lie algebra of $SO(2\nu)$:

$$\begin{array}{lll} B_0 = +2, & \{E_{\alpha\beta}\}, & \tfrac{1}{2}\nu(\nu-1), \\ B_0 = 0, & \{H_\alpha; E_{\pm\alpha\mp\beta}\}, & \nu^2, \\ B_0 = -2, & \{E_{-\alpha-\beta}\}, & \tfrac{1}{2}\nu(\nu-1). \end{array}$$

The generator of the $U(\nu)$ subgroup is the subset $B_0 = 0$.

We now consider the inclusion $U(p) \subset Sp(2p)$ and the natural choice for $U(p)$ is the set of the p^2 generators B_{ij}. The $U(1)$ generator B_S is then defined by

$$B_S = \sum_1^{p} B_{ii} .$$

Using the commutation relations of B_S with the symplectic generators we obtain again a Jordan structure for the Lie algebra of $Sp(2p)$:

$$\begin{array}{lll} B_S = +2, & \{C_{ij}\}, & \tfrac{1}{2}p(p+1), \\ B_s = 0, & \{B_{ij}\}, & p^2 \\ B_S = -2, & \{D_{ij}\}, & \tfrac{1}{2}p(p+1). \end{array}$$

TABLE I. Quantum numbers of the generators of $OSP(2\nu|2p)$.

Generators	B_0	B_S
H_α	0	0
$E_{+\alpha+\beta}$	+2	0
$E_{\pm\alpha\mp\beta}$	0	0
$E_{-\alpha-\beta}$	−2	0
B_{ij}	0	0
C_{ij}	0	+2
D_{ij}	0	−2
$F_{+\alpha j}$	+1	+1
$F_{-\alpha j}$	−1	+1
$G_{+\alpha j}$	+1	−1
$G_{-\alpha j}$	−1	−1

We have given, in Table I, the values of B_0 and B_S for the orthosymplectic generators of $OSP(2\nu|2p)$.

The location of $U(\nu|p)$ in $OSP(2\nu|2p)$ is defined by choosing the $U(1)$ generator Q_N not in $SU(\nu|p)$:

$$Q_N = B_0 + B_S .$$

We then get a Jordan structure for the orthosymplectic superalgebra and we now exhibit the $SU(\nu) \otimes SU(p) \otimes U(1)$ behavior of the various components, the $U(1)$ generator Q being normalized as usual:

$$Q = (1/\nu)B_0 + (1/p)B_S ,$$

$$Q_N = +2 \begin{cases} \{E_{+\alpha+\beta}\}, & (\tfrac{1}{2}\nu(\nu-1),1), \quad Q = 2/\nu, \\ \{F_{+\alpha j}\}, & (\nu,p), \quad Q = 1/\nu + 1/p, \\ \{C_{ij}\}, & (1,\tfrac{1}{2}p(p+1)), \quad Q = 2/p; \end{cases}$$

$$Q_N = 0 \begin{cases} \{G_{+\alpha j}\}, & (\nu,\bar{p}), \quad Q = 1/\nu - 1/p, \\ \{H_\alpha; E_{\pm\alpha\mp\beta}\}, & (\nu^2-1,1) \oplus (1,1), \quad Q = 0, \\ \{B_{ij}\}, & (1,p^2-1) \oplus (1,1), \quad Q = 0, \\ \{F_{-\alpha j}\}, & (\bar{\nu},p), \quad Q = -1/\nu + 1/p; \end{cases}$$

$$Q_N = -2 \begin{cases} \{D_{ij}\}, & (1,\tfrac{1}{2}p(p+1)) \quad Q = -2/p, \\ \{G_{-\alpha j}\}, & (\bar{\nu},\bar{p}), \quad Q = -1/\nu - 1/p, \\ \{E_{-\alpha-\beta}\}, & (\tfrac{1}{2}\nu(\nu-1),1), \quad Q = -2/\nu. \end{cases}$$

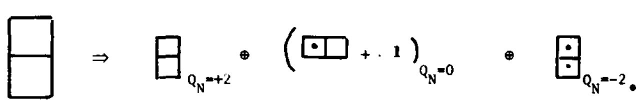
The subsuperalgebra of $U(\nu|p)$ is the set of generators $Q_N = 0$ and, as expected, its dimension is $(\nu+p)^2$. The subset $Q = 0$ corresponds to the even part l_0 and the two singlets $(1,1)$ are the generators B_0 and B_S.

The components $Q_N = \pm 2$ have the same dimension,

$$\tfrac{1}{2}\nu(\nu-1) + 2\nu p + \tfrac{1}{2}p(p+1),$$

and they are associated to two contragradient irreducible representations of $SU(\nu|p)$, namely the covariant and contravariant superantisymmetric tensors of order 2.

It is then particularly attractive to use the supertableau language for the reduction of the adjoint representation of $OSP(2\nu|2p)$ and the previous results are simply written



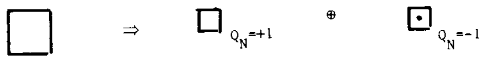
(1)

The component $Q_N = 0$ is fully reducible when $\nu \neq p$ and nonfully reducible for $\nu = p$ (see Ref. 13).

IV. REDUCTION OF THE SUPERTABLEAUX OF OSP($2\nu|2p$)

Consider first the one-box supertableau of $OSP(2\nu|2p)$ of dimension $2\nu + 2p$. The reduction of $OSP(2\nu|2p)$

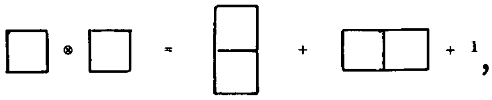
$\Rightarrow U(\nu|p)$ is very simple and we get



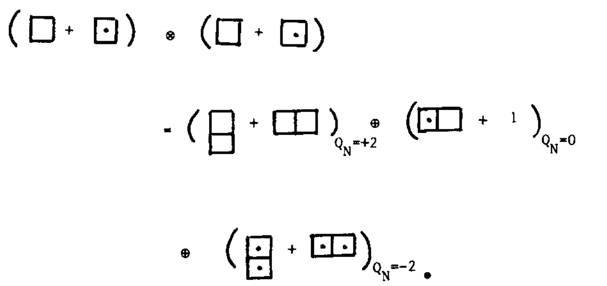
(2)

$$2\nu+2p = \quad (\nu+p) \quad + \quad (\nu+p).$$

The reduction $OSP(2\nu|2p) \Rightarrow U(\nu|p)$ of a legal orthosymplectic supertableau T is then obtained by using, step by step, the tensor product method. As an illustration consider the tensor product in $OSP(2\nu|2p)$,

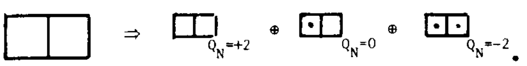

(3)

and its analog in $U(\nu|p)$,



(4)

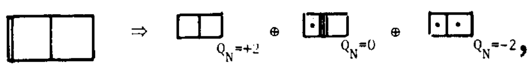
After comparison with the results obtained for the adjoint supertableau we get



(5)

When $\nu \neq p$ the tensor product (3) is fully reducible and the one-row supertableau of $OSP(2\nu|2p)$ with two boxes is irreducible. As a consequence of the relation (5) the irreducible representation of $OSP(2\nu|2p)$ associated to this supertableau is decomposed in a direct sum of three irreducible representations of $SU(\nu|p)$.

When $\nu = p$ the singlet part of the supersymmetry subspace in (3) cannot be isolated[13] and we have a two-generalized atypical supertableau and a nonfully reducible representation of $OSP(2p|2p)$ (see Ref. 13). The decomposition (5) is now replaced by



(6)

and the second term in the right-hand side is a two-generalized atypical supertableau of $SU(p|p)$.

The reduction $OSP(2\nu|2p) \Rightarrow U(\nu|p)$ of a legal orthosymplectic supertableau T is a decomposition of T in legal supertableaux of the superunitary group $SU(\nu|p)$. We write symbolically

$$T = \sum t_{Q_N(t)}. \tag{7}$$

Let us call as N_c ($N_{\bar{c}}$) the number of covariant (contravariant) boxes of the supertableau t. As an immediate consequence of the tensor product method the value $Q_N(t)$ for the supertableau t is given by

$$Q_N(t) = N_c - N_{\bar{c}}. \tag{8}$$

The supersymmetry properties of the boxes of the supertableaux t are obviously governed by those of the orthosymplectic supertableau T. In particular the non-negative integers N_c and $N_{\bar{c}}$ are restricted by

$$0 \leqslant N_c + N_{\bar{c}} \leqslant N, \tag{9}$$

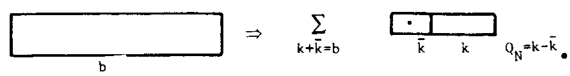
and the allowed values of $Q_N(t)$ are

$$Q_N(t) = N, N-2, N-4, \ldots, 2-N, -N. \tag{10}$$

The maximal value $Q_N = N$ corresponds to $t = T$ and the minimal value $Q_N = -N$ to the contragradient supertableau $t = \overline{T}$.

Let us emphasize that the supertableaux t entering in the right-hand side of Eq. (7) must not only respect the supersymmetry properties of the supertableaux T but also be legal supertraceless supertableaux of $SU(\nu|p)$.

Moreover the orthosymplectic supertableau T being self-contragradient the superunitary supertableaux t appearing in the decomposition (7) either are self-contragradient supertableaux—and this may occur only when zero is an allowed value of Q_N, which implies N even—or come into pairs of contragradient supertableaux with opposite values of Q_N. Let us give, as a simple illustration, the reduction of the one-row supertableaux of $OSP(2\nu|2p)$. The superunitary supertableaux t have at most one covariant row and one contravariant row:



(11)

When the orthosymplectic supertableau T is irreducible the sum (7) is a direct sum of irreducible supertableaux and of generalized atypical supertableaux of $SU(\nu|p)$. In particular when a t is not irreducible we find in the sum (7) the partners of t forming, with t, a generalized atypical supertableau. Of course, we have a full reducibility with respect to Q_N.

When the orthosymplectic supertableau T is not irreducible the previous property for a nonirreducible t is no longer true and in order to reconstruct correctly the generalized atypical supertableau of $OSP(2\nu|2p)$ we must simultaneously consider the decomposition of the various orthosymplectic supertableaux forming with T the generalized atypical orthosymplectic supertableau.

V. REDUCTION OF THE SUPERTABLEAUX OF OSP(2ν+1|2p)

The Lie superalgebra of the orthosymplectic group $OSP(2\nu+1|2p)$ has the rank $\nu+p$ and it is simple. The notation used by Kac[12] is $B(\nu,p)$. The even part L_0 of $B(\nu,p)$ is $L_0 = B_\nu + C_p$. It is semisimple for $\nu \geqslant 1$ and simple for $\nu = 0$. The odd part L_1 of $B(\nu,p)$ is irreducible and it behaves like a $(2\nu+1, 2p)$ representation of L_0. The dimensions of L_0 and L_1 are, respectively,

$$\dim L_0 = \nu(2\nu+1) + p(2p+1),$$
$$\dim L_1 = (2\nu+1)2p.$$

The generators of $OSP(2\nu|2p)$ are also generators of $OSP(2\nu+1|2p)$ and, in addition, for $B(\nu,p)$ we have (i) 2ν orthogonal even generators,

$E_{\epsilon_\alpha \alpha}$, root $\epsilon_\alpha d_\alpha$;

and (ii) $2p$ odd generators,

F_j, root e_j,

G_j, root $-e_j$.

The subsuperalgebra of $U(\nu|p)$ is defined as in the previous case (Sec. III) and the quantum numbers B_0 and B_S of the new generators are given in Table II.

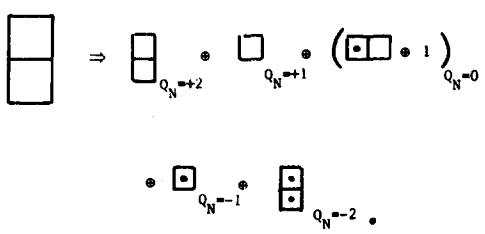
As a consequence, in the reduction $OSP(2\nu+1|2p) \Rightarrow U(\nu|p)$, the adjoint representation of $OSP(2\nu+1|2p)$ is decomposed in five components according to Q_N. The three components $Q_N = \pm 2,0$ are the same as previously and the two new components $Q_N = \pm 1$ are given, with $SU(\nu) \otimes SU(p) \oplus U(1)$ properties as follows:

$$Q_N = +1 \begin{cases} \{E_{+\alpha}\}, & (\nu,1), \quad Q = 1/\nu, \\ \{F_j\}, & (1,p), \quad Q = 1/p; \end{cases}$$

$$Q_N = -1 \begin{cases} \{G_j\}, & (1,\bar{p}), \quad Q = -1/p, \\ \{E_{-\alpha}\}, & (\bar{\nu},1), \quad Q = -1/\nu. \end{cases}$$

These components have the same dimension $(\nu + p)$ and they are associated to the two contragradient fundamental representations of $SU(\nu|p)$ described by the one covariant box and one contravariant box supertableaux.

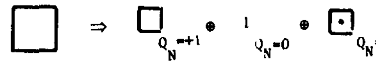
The decomposition formula (1) for the adjoint representation of $B(\nu,p)$ has the form



(12)

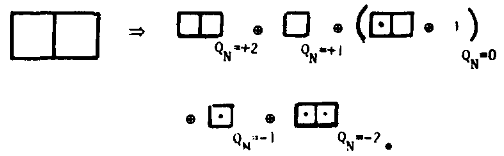
The considerations developed in Sec. IV for the superunitary reduction of the orthosymplectic supertableaux of $OSP(2\nu|2p)$ extend in a straightforward way with trivial modifications.

For the fundamental supertableau of $OSP(2\nu+1|2p)$ we have one new contribution and the formula (2) has to be replaced by



(13)

and, as given by the tensor product method, the decomposition of the supersymmetric tensor of rank 2 takes now the form



(14)

Equations (8) and (9), restricting the superunitary supertableau t entering in the decomposition formula (7), are still valid. However all integer values of $Q_N(t)$ between $+N$ and $-N$ are now allowed and we replace Eq. (10) by

$$Q_N(t) = N, N-1, N-2, \ldots, 1-N, -N. \tag{15}$$

The reduction $OSP(2\nu+1|2p) \Rightarrow U(\nu|p)$ is also conveniently obtained through the chain of subsupergroups

$OSP(2\nu+1|2p) \Rightarrow OSP(2\nu|2p) \Rightarrow U(\nu|p)$.

Combining the results of the Sec. IV and of Appendix A we get the desired result. As an illustration we give, in Fig. 1, the reduction of the fourth-order tensor with mixed supersymmetry whose supertableau is a square.

TABLE II. Generators of $OSP(2\nu+1|2p)$ not in $OSP(2\nu|2p)$.

Generators	B_0	B_S
$E_{+\alpha}$	$+1$	0
$E_{-\alpha}$	-1	0
F_j	0	$+1$
G_j	0	-1

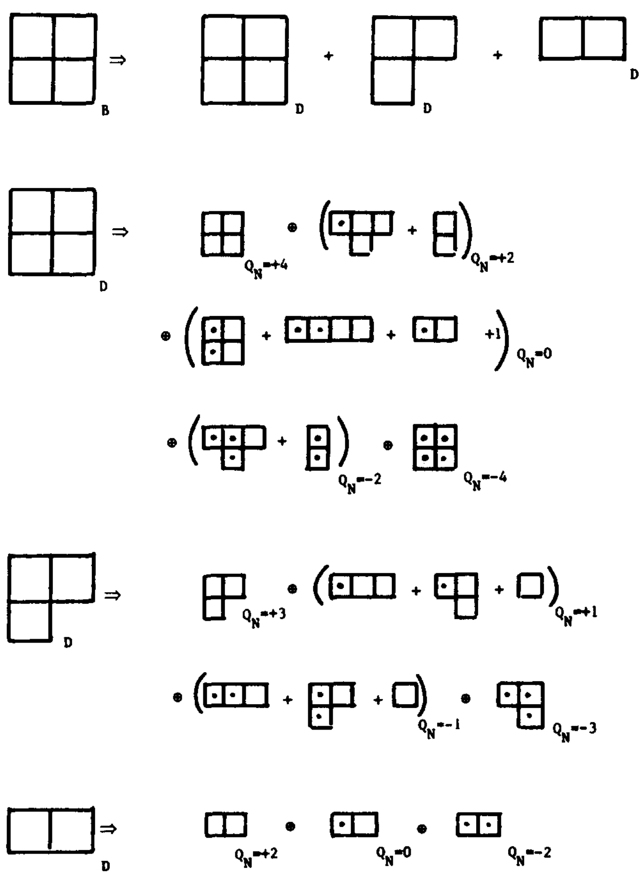


FIG. 1. Reduction of the square supertableau of $OSP(2\nu+1|2p)$.

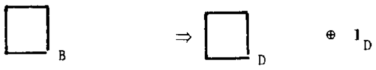
VI. CONCLUDING REMARKS

By using the tensor product method we have obtained the reduction of the supertableau of the orthosymplectic groups OSP$(2\nu|2p)$ and OSP$(2\nu+1|2p)$ in superunitary supertableaux of SU$(\nu|p)$. As a by-product of these results it is straightforward through the decomposition formula (7) for supertableaux to obtain the reduction of the irreducible and nonfully reducible representation of OSP$(2\nu|2p)$ and OSP$(2\nu+1|2p)$ described by supertableaux in terms of irreducible and nonfully reducible representations of the superunitary group SU$(\nu|p)$. This problem, being just a question of careful and patient computation, will not be treated here and we shall only give a few examples in Appendix B for the case of reduction OSP$(2|4)\Rightarrow$SU$(1|2)$.

APPENDIX A: SUPERTABLEAUX OF OSP($2\nu|2p$) AND OSP($2\nu+1|2p$)

The two sets of legal supertableaux of OSP$(2\nu|2p)$ and OSP$(2\nu+1|2p)$ are identical and the corresponding supertableaux T have, at most, ν rows and p columns of arbitrary length.[8] However, according as T is viewed as a supertableau of OSP$(2\nu|2p)$ or as a supertableau of OSP$(2\nu+1|2p)$, its dimension, the nature of its atypicity, and its interpretation in terms of representation of the relevant superalgebra are different.

Obviously OSP$(2\nu|2p)$ is a subsupergroup of OSP$(2\nu+1|2p)$ and the reduction OSP$(2\nu+1|2p)\Rightarrow$OSP$(2\nu|2p)$ of the fundamental representation of $B(\nu,p)$ is conveniently written in the supertableau language as



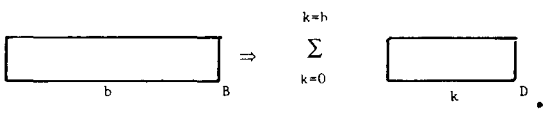
$$(2\nu+1+2p)=(2\nu+2p)+1. \qquad \text{(A1)}$$

We call x_0 the bosonic index associated to the OSP$(2\nu|2p)$ singlet of the previous decomposition. The decomposition of a supertableau T of OSP$(2\nu+1|2p)$ is then obtained, by using the tensor product method and the result is a sum of supertableaux of OSP$(2\nu|2p)$:

$$T_B \Rightarrow T_D + \sum \widetilde{T}_D.$$

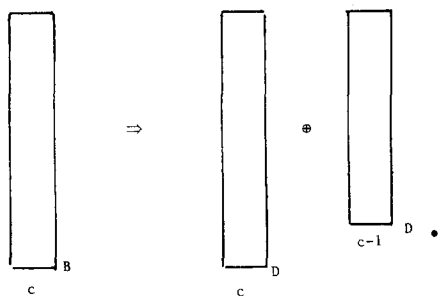
The supertableaux $\widetilde{T}$ of the previous sum are constructed from T by suppressing 1,2,... boxes, where the index x_0 can be inserted in a way compatible with the supersymmetry properties of the boxes of T.

As an illustration consider the one-row and the one-column supertableaux. In the case of a one-row supertableaux all boxes can receive the bosonic index x_0 and we get



(A2)

We only have one place for the bosonic index x_0 in one-column supertableau and the decomposition is simply



(A3)

The two relations (A2) and (A3) are easily checked by using the dimension formulas for orthosymplectic supertableau and for that purpose they just emerge as consequences of the properties of the combinatory coefficients.

APPENDIX B: REDUCTION OF THE OSP(2|4) SUPERTABLEAUX

The supertableaux of the orthosymplectic group OSP$(2|4)$ have been classified and interpreted in Ref. 9. The results are the following.

(1) The one column supertableaux ($C_2=0$) are irreducible and atypical; they describe a self-contragradient irreducible atypical representation of OSP$(2|4)$.

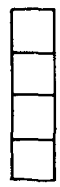
(2) The two-column supertableaux ($C_2\geqslant 1$, $C_3=0$) are irreducible and typical; they describe a self-contragradient irreducible typical representation of OSP$(2|4)$.

(3) The two-column and one-row supertableaux ($C_3=1$) are typical or atypical.

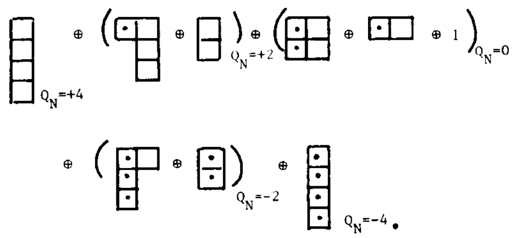
(3a) When typical they are irreducible and they describe a direct sum of contragradient irreducible typical representations of OSP$(2|4)$.

(3b) When nontypical they are nonreducible and with a second atypical supertableau they generate a two-generalized atypical supertableau, which describes either a self-contragradient or a direct sum of two contragradient nonfully reducible representations of OSP$(2|4)$.

Case 1: T is a one-column supertableau with four boxes:

$T=$  $\Rightarrow\{4|0,0\}_{85}.$

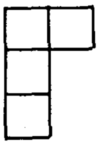
The reduction of T in supertableau of SU$(1|2)$ is the following:



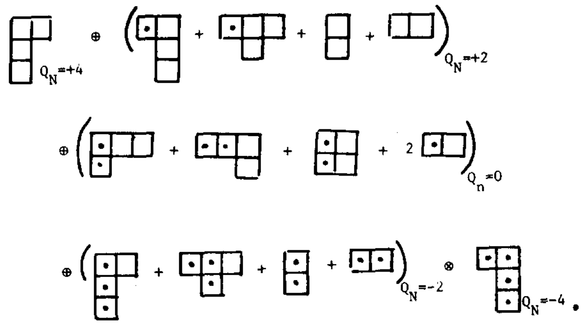
For the decomposition of the atypical representation $\{4|0,0\}_{85}$ we get

$$[\{4|3\}_9]_{Q_N=+4} \oplus [\{3|3\}_{10} \oplus \{2|1\}_5]_{Q_N=+2}$$
$$\oplus [\{2|3\}_{16} \oplus \{1|1\}_8 \oplus \{0|0\}_1]_{Q_N=0}$$
$$\oplus [\{1|3\}_{10} \oplus \{0|2\}_5]_{Q_N=-2}$$
$$\oplus [\{0|4\}_9]_{Q_N=-4}.$$

Case 2: T is a two-column supertableau

$T=$  $\Rightarrow \{4|2,0\}_{160}$.

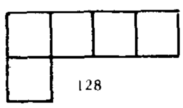
The reduction of T in supertableaux of SU(1|2) is the following:



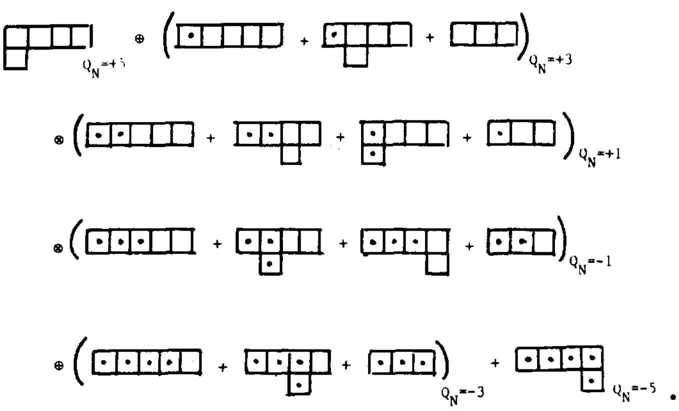
For the irreducible typical representation $\{4|2,0\}_{160}$ we get

$$[\{4|2\}_{12}]_{Q_N=+4} \oplus [\{3|3\}_{16} \oplus \{2|0\}_4 \oplus (\{3|2\}_7 + 2\{2|1\}_5 + \{1|0\}_8)_{20}]_{Q_N=+2}$$
$$\oplus [\{2|2\}_{12} \oplus \{1|2\}_{12} + \{2|3\}_6 + 2\{1|1\}_8]_{Q_N=0}$$
$$\oplus [\{1|3\}_{16} \oplus \{-1|0\}_4 \oplus (\{0|1\}_3 + 2\{0|2\}_5 + \{0|3\}_7)_{20}]_{Q_N=-2} \oplus [\{-1|2\}_{12}]_{Q_N=-4}.$$

Case 3: T is a two-column, one-row, typical supertableau

$T=$  $\Rightarrow \{5|1,0\}_{64} \oplus \{1|1,0\}_{64}$.

The reduction of T in supertableaux of SU(1|2) is the following:



For the decomposition of the two typical representations of OSP(2|4) we get

$$\{5|1,0\}_{64} \Rightarrow [\{5|1\}_8]_{Q_N=+5} \oplus [\{4|1\}_8 \oplus \{5|2\}_{12} \oplus \{3|0\}_4]_{Q_N=+3}$$
$$\oplus [\{2|0\}_4 \oplus \{1|1\}_8 \oplus (\{3|2\}_7 + 2\{2|1\}_5 + \{1|0\}_3)_{20}]_{Q_N=+1},$$
$$\{1|1,0\}_{64} \Rightarrow [\{-1|0\}_4 \oplus \{1|1\}_8 \oplus (\{0|3\}_7 + 2\{0|2\}_5 + \{0|1\}_3)_{20}]_{Q_N=-1}$$
$$\oplus [\{2|1\}_8 \oplus \{-1|2\}_{12} + \{-2|0\}_4]_{Q_N=-3} \oplus [\{-3|1\}_8]_{Q_N=-5}.$$

Let us notice the appearance, in the components $Q_N = \pm 1$, of nonfully reducible representations of SU(1|2) of dimension 20.

Case 4: T_1 is an atypical nonirreducible supertableau. The simplest cases correspond to one row with three and four boxes. The partners of T are one-row supertableaux with, respectively one and zero boxes:

$$(T_1,T_2) = \underset{32}{\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}}$$

$$\Rightarrow\{3|0,0\}_{10} + 2\{1|0,0\}_6 + \{0|0,1\}_{10},$$

$$(T_3,T_2) = \underset{32}{\begin{array}{|c|c|c|c|}\hline & & & \\ \hline \end{array}}$$

$$\Rightarrow\{4|0,0\}_{15} + 2\{0|0,0\}_1 + \{0|1,0\}_{15}.$$

The reduction of the one-row supertableaux of OSP(2|4) is given by the general equation (10) and it will not be repeated now. For the atypical representations of OSP(2|4) involved in these two supertableaux we get

$$\{3|0,0\}_{10}\Rightarrow\{3|0\}_4 + \{2|1\}_5 + \{0|0\}_1,$$
$$\{0|0,1\}_{10}\Rightarrow\{0|0\}_1 + \{0|2\}_5 + \{-2|0\}_4,$$
$$\{4|0,0\}_{15}\Rightarrow\{4|0\}_4 + \{3|1\}_8 + \{1|0\}_3,$$
$$\{0|1,0\}_{15}\Rightarrow\{0|1\}_3 + \{-1|1\}_8 + \{-3|0\}_4.$$

In the reduction of the nonfully reducible representations of OSP(2|4) associated to the two-generalized supertableaux, the components $Q_N = \pm 1$ in the first case and $Q_N = 0$ in the second one are nonfully reducible representations of SU(1|2):

$$Q = +1, \quad (\{2|1\}_5 + 2\{1|0\}_3 + \{0|0\}_1)_{12},$$
$$Q = -1, \quad (\{0|0\}_1 + 2\{0|1\}_3 + \{0|2\}_5)_{12},$$
$$Q = 0, \quad (\{1|0\}_3 + 2\{0|0\}_1 + \{0|1\}_3)_8.$$

[1]I. Bars and M. Günaydin, Commun. Math. Phys. **91**, 31 (1983).
[2]D. Z. Freedman and H. Nicolai, Nucl. Phys. B **237**, 342 (1984).
[3]A. Ceresole, P. Fre, and H. Nicolai, CERN Preprint TH 3957, 1984.
[4]M. Günayoin and N. P. Warner, Caltech Preprint CALT 88-1202, 1985.
[5]I. Bars, B. Morel, and H. Ruegg, J. Math. Phys. **24**, 2253 (1983).
[6]F. Delduc and M. Gourdin, J. Math. Phys. **25**, 1651 (1984).
[7]F. Delduc and M. Gourdin, J. Math. Phys. **26**, 1865 (1985).
[8]R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **17**, 2365 (1984).
[9]M. Gourdin, Preprint Paris LPTHE 85/9.
[10]M. Gourdin, Preprint Paris LPTHE 86/08.
[11]M. Gourdin, J. Math. Phys. **27**, 61 (1986).
[12]V. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); *Lecture Notes in Mathematics*, Vol. 676 (Springer, New York, 1978), p. 597.
[13]A. B. Balentekin and I. Bars, J. Math. Phys. **22**, 1149, 1810 (1981).

Closed, analytic, boson realizations for Sp(4)

Abraham Klein and Qing-Ying Zhang[a)]
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6396

(Received 28 January 1986; accepted for publication 8 April 1986)

The problem of determing a boson realization for an arbitrary irrep of the unitary simplectic algebra Sp($2d$) [or of the corresponding discrete unitary irreps of the unbounded algebra Sp($2d,R$)] has been solved completely in recent papers by Deenen and Quesne [J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982); **25**, 1638 (1984); **26**, 2705 (1985)] and by Moshinsky and co-workers [O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985); M. Moshinsky, "Boson realization of symplectic algebras," to be published]. This solution is not known in closed analytic form except for $d = 1$ and for special classes of irreps for $d > 1$. A different method of obtaining a boson realization that solves the full problem for Sp(4) is described. The method utilizes the chain Sp($2d$) $\supset$SU(2) $\times$SU(2) $\times\cdots\times$SU(2) (d times), which, for $d \geqslant 4$, does not provide a complete set of quantum numbers. Though a simple solution of the missing label problem can be given, this solution does not help in the construction of a mapping algorithm for general d.

I. INTRODUCTION

Depending on how much further we shall be able to develop the ideas to be presented in this paper, they may be viewed either as an addendum to recent systematic studies of boson realizations for symplectic algebras, or as an alternative approach to the entire subject.

The current interest in symplectic algebras has its genesis in two distinct developments. One is associated with the Sp($6,R$) model of collective motion,[1] where the introduction of boson realizations may be viewed *inter alia* as a convenience in the identification or isolation of specific collective degrees of freedom contained in the model. Here three sets of authors[2–4] have contributed to the solution of the problem of obtaining a closed form of boson mapping of the algebra SP($2d,R$), where d is any integer.

Quite independently, interest has been regenerated in the compact unitary symplectic algebras Sp($2d$) [in connection with the problem of understanding the microscopic basis of the interacting boson model (IBM)] starting with a discussion of Sp(6) by Ginocchio.[5] This same model has been infused with potentially new physical significance in recent work, which attempts to free the concept of dynamical symmetry from its boson realization.[6]

Stimulated by Ginocchio's work, one of the present authors and his collaborators[7] developed two methods and three different boson realizations for a subset of irreps of Sp(4) [= SO(5)], namely the irreps containing the vacuum state of some simplified shell models. One of the methods and the corresponding realization derived by its means is a special case of the method applied to Sp($2d,R$) in the previously cited literature.[2–4]

Further work along these lines was inspired by the continuing search for a microscopic basis for the IBM for deformed nuclei. In this development, Bonatsos and Klein[8] studied the shell model algebra for identical nucleons coupled to spin-0. This is an algebra Sp($2d$) for various d (depending on the shell) that is analyzed through the chain U(d) $\supset$SU(3), with boson realizations constructed in an angular momentum coupled basis. This chain is of physical interest in connection with the pseudo-SU(3) approximation[9] for heavy deformed nuclei, which may provide a basis for or an alternative to the IBM.

In other recent work, Hecht and Elliott[10] have applied the method described in Refs. 2–4 to various physically interesting examples of Sp(4) and Sp(6).

In Refs. 3 and 4 the problem of obtaining boson realizations of an arbitrary irrep of Sp($2d,R$) has been completely solved by a method that proceeds in two steps: (i) a non-Hermitian Dyson mapping[11] is obtained first; and (ii) this mapping is Hermitized. The first step is simple and yields closed, analytic expressions for the generators of the algebra. Except for the case of Sp($2,R$) and special classes of representations for Sp($2d,R$), $d > 1$, the second step involves numerical processes and thus the general result is not yet known in *explicit*, *analytic*, and *closed* form. The same statement holds for Sp($2d$).

In this note, we shall derive a boson mapping for Sp(4) in an *explicit*, *analytic*, and *closed* form by generalizing another of the mappings derived in Ref. 7, where it was introduced in connection with the monopole plus pairing model,[12] a simplified nuclear model. The remark of the opening paragraph of this paper refers to the absence, so far, of an algorithm extending the results to follow to arbitrary Sp($2d$).

In Sec. II, we review some basic properties of the algebra Sp($2d$) and outline our approach to obtaining a boson realization. In Sec. III, we illustrate our method by deriving a realization for a special irrep of Sp(4) (the so-called vacuum irrep). The same problem is solved for Sp(6) in Sec. IV. The point of this latter exercise is made clear in Sec. V, where it is shown that we have thereby also solved the problem of finding a general boson realization of Sp(4), requiring only a

[a)] On leave from Physics Department, Hunan University, Changsha, People's Republic of China.

 0022-2488/86/081987-07$02.50

renaming of some symbols. In Sec. VI, we remark how the results are trivially modified so as to apply to $Sp(4,R)$ or $Sp(6,R)$. Finally Sec. VII is devoted to a preliminary discussion of an extension of the results of this paper.

II. ALGEBRA AND METHOD OF MAPPING

We concentrate on the bounded algebra $Sp(2d)$. The corresponding results for $Sp(2d,R)$, which require only trivial modifications, will be presented in Sec. VI. Let $\psi_{mi},\psi^\dagger_{mi}$ be a set of fermion destruction and creation operators satisfying standard anticommunication relations, e.g.,

$$\{\psi_{mi},\psi^\dagger_{m'i'}\} = \delta_{mm'}\delta_{ii'} . \tag{2.1}$$

Here m are the components of a half-integral angular momentum j,

$$\sum_{m=-j}^{j} 1 = 2j+1 \equiv 2\Omega , \tag{2.2}$$

but the physical nature of the index i is left unspecified (temporarily) except that it takes on the d values $i = 1,2,3,...,d$. The pair operators ($\bar{m} = -m$)

$$A^\dagger_i = (2\sqrt{\Omega})^{-1} \sum_m (-1)^{j-m}\psi^\dagger_{mi}\psi^\dagger_{\bar{m}i} = (A_i)^\dagger , \tag{2.3a}$$

$$A^\dagger_{ik} = (2\Omega)^{-1/2} \sum_m (-1)^{j-m}\psi^\dagger_{mi}\psi^\dagger_{\bar{m}k} = A^\dagger_{ki} = (A_{ik})^\dagger , \tag{2.3b}$$

and the multipole operators

$$N_i = \sum_m \psi^\dagger_{mi}\psi_{mi} , \tag{2.4a}$$

$$B_{ik} = (2\Omega)^{-1/2} \sum_m \psi^\dagger_{mi}\psi_{mk} \equiv (2\Omega)^{-1/2} N_{ik} \tag{2.4b}$$

and [up to an additive constant for (2.4a)] the generators of $Sp(2d)$ satisfying the algebra (with the omission of obvious relations)

$$[A_i,A^\dagger_k] = \delta_{ik}[1-(N_i/\Omega)] , \tag{2.5a}$$

$$[A_{ik},A^\dagger_{ik}] = 1-(2\Omega)^{-1}(N_i+N_k) , \tag{2.5b}$$

$$[A_{ki},A^\dagger_{li}] = -(2\Omega)^{-1/2}B_{lk} , \quad l\neq k , \tag{2.5c}$$

$$[A_i,A^\dagger_{ki}] = -\Omega^{-1/2}B_{ki} , \tag{2.5d}$$

$$[B_{ik},B_{ki}] = (2\Omega)^{-1}(N_i-N_k) , \tag{2.5e}$$

$$[B_{ik},B_{kl}] = (2\Omega)^{-1/2}B_{il} , \quad i\neq l . \tag{2.5f}$$

We have chosen deliberately not to define the generators in as concise a form as possible, distinguishing between the diagonal ($i=k$) and nondiagonal ($i\neq k$) generators because of the physical picture we wish to associate with the mapping we have in mind. In this interpretation the sum over m couples to angular momentum zero and the index i enumerates different single-particle levels, so that $Sp(2d)$ is an algebra of angular momentum zero operators referring to d single-particle levels.

This shell model picture suggests a method for characterizing the irreps of $Sp(2d)$. The simplest irrep is the one that contains the vacuum state as the state of minimum weight. This irrep is identified as the collective irrep in the corresponding work on $Sp(2d,R)$.[3,4] In the present context we prefer to call it the vacuum irrep, the only one studied in our previous work.[7] A nonorthonormal but linearly independent basis for this representation is provided by the set of vectors

$$|n_1 \cdots n_d, n_{12} \cdots n_{d-1,d}\rangle = \prod_i (A^\dagger_i)^{n_i} \prod_{i<j} (A^\dagger_{ij})^{n_{ij}}|0\rangle , \tag{2.6}$$

where $|0\rangle$ satisfies the equations

$$A_i|0\rangle = A_{ik}|0\rangle = N_i|0\rangle = N_{ik}|0\rangle = 0 . \tag{2.7}$$

The importance for us of these well-known observations is that, just as in our previous work, we can, for our purposes, bypass a good deal of (elegant) formalism and note immediately that if we introduce bosons $a^\dagger_i,a_i$ and $a^\dagger_{ik},a_{ik}$ associated with each of the corresponding capital letter operators, where

$$[a_i,a^\dagger_k] = \delta_{ik}, \tag{2.8a}$$

$$[a_{ik},a^\dagger_{jl}] = \delta_{ij},\delta_{kl} , \tag{2.8b}$$

$$[a_i,a^\dagger_{jk}] = 0 , \tag{2.8c}$$

then the orthonormal basis is a Heisenberg–Weyl space of $\frac{1}{2}d(d+1)$ dimensions, namely,

$$|n_1 \cdots n_{d-1,d}) = \prod_i \frac{(a^\dagger_i)^{n_i}}{\sqrt{n_i!}} \prod_{i<j} \frac{(a^\dagger_{ij})^{n_{ij}}}{\sqrt{n_{ij}!}} |0) \tag{2.9}$$

can serve as a basis for the vacuum irrep. This is of interest provided we can realize the generators of $Sp(2d)$ as functions of the boson operators with the following properties: (i) they satisfy the commutation relations (2.5), and (ii) they satisfy the conditions (2.7).

We note that any such realization solves the problem of orthonormalizing the basis (2.6). This is done[3,4,7] by inverting the realization so as to express the boson operators as functions of the generators and therefore as operators in the shell model space.

The realization to be studied will involve the chain of subalgebras

$$Sp(2d) \supset SU(2)_1 \times SU(2)_2 \times \cdots \times SU(2)_d , \tag{2.10}$$

where $SU(2)_i$ is generated by [cf. (2.5a)] A_i, $A^\dagger_i$, and $\frac{1}{2}(\Omega_i - N_i)$ playing the roles of J_-, J_+, and J_z. Each $SU(2)_i$ furnishes a pair of quantum numbers n_i,v_i, associated with the operators n_i and v_i, where

$$\hat{n}_i = a^\dagger_i a_i , \tag{2.11}$$

$$v_i = \sum_{j\neq i} \hat{n}_{ij} \equiv \sum_{j\neq i} a^\dagger_{ij}a_{ij} , \tag{2.12}$$

$$\hat{n}_{ij} = \hat{n}_{ji} , \tag{2.13}$$

and v_i the seniority of the ith level, i.e., the number of fermions of type i present not coupled to pairs described by $A^\dagger_i$. Observe that for $d = 2,3$ these determine completely the basis (2.9). In fact, for $d=2$, we even have from (2.12), $v_1 = v_2$. However, from $d=4$ on we encounter the missing label problem.

Consider, for example, $d=4$, where the set (2.11) and (2.12) provide eight of ten needed quantum numbers. A possible choice for the two additional operators is the pair

$$\Lambda_2 = \sum_{i_1\neq i_2\neq i_3\neq i_4} a^\dagger_{i_1i_2}a^\dagger_{i_3i_4}a_{i_4i_2}a_{i_3i_1} , \tag{2.14}$$

$$\Lambda_4 = \sum_{i_1\neq i_2\neq i_3\neq i_4} (a^\dagger_{i_1i_2})^2(a^\dagger_{i_3i_4})^2 a_{i_1i_3}a_{i_2i_3}a_{i_1i_4}a_{i_2i_4} + \text{H.c.} \quad (2.15)$$

The utilization of such operators would, however, defeat our purposes, as will be clear below, since it would require us to introduce eigenvectors of these operators by carrying out a unitary transformation of the basis (2.9). Instead we shall work exclusively with the basis (2.9), the simplest of all possible choices, which are simultaneous eigenstates of the operators $\hat{n}_i$ and $\hat{n}_{ij}$, which provide a full characterization. The fact that n_i and v_i are diagonal will remain of fundamental importance.

It is best, at this point, to turn to specifics. In the next section we will develop the vacuum mapping for Sp(4) and in Sec. IV the corresponding mappings for Sp(6). Some aspects of the problem for higher d are discussed in Sec. VII.

III. VACUUM IRREP FOR Sp(4)

The realization to be developed below was first given without proof in Ref. 7. Here we shall provide some details. Our method is firmly rooted in the requirements dictated by Eq. (2.10). The expressions ($i = 1,2$)

$$N_i = 2\hat{n}_i + \hat{n}_{12}, \quad (3.1)$$

$$A^\dagger_i = (A_i)^\dagger = a^\dagger_i r_i, \quad (3.2)$$

$$r_i = [1 - \Omega^{-1}(\hat{n}_i + \hat{v}_i)]^{1/2}, \quad (3.3)$$

are the well-known Holstein–Primakoff[13] realizations of $SU(2)_i$. They provide a partial solution to our problem since they also satisfy the appropriate parts of Eq. (2.7). Thus we have only to find the realizations of $A^\dagger_{12}$ (or A_{12}) and B_{21} (or $B_{12} = B^\dagger_{21}$).

The most general form allowed for $A^\dagger_{12}$, for example, is

$$A^\dagger_{12} = a^\dagger_{12}\Phi_1(\hat{n}_1,\hat{n}_2,\hat{n}_{12}) + a^\dagger_1 a^\dagger_2 a_{12}\Phi_2(\hat{n}_1,\hat{n}_2,\hat{n}_{12}) . \quad (3.4)$$

This is because $A^\dagger_{12}$ is the component of a spinor [under both $SU(2)_1$ and $SU(2)_2$] satisfying the selection rules (i) $\Delta N_i = 1, i = 1,2$; and (ii) $\Delta v_i = \pm 1$. As expressd by (3.4), a basis for operators satisfying these restrictions is given by $a^\dagger_{12}$ and $a^\dagger_1 a^\dagger_2 a_{12}$; in turn each may be multiplied by a (so far) arbitrary function of the diagonal operators $\hat{n}_i$, $\hat{n}_{ij}$.

The great convenience of the form (3.4) associated with the basis (2.9) will be evident to anyone who checks the manipulations described below, which depend only on versions of the relation

$$a^\dagger\Phi(\hat{n}) = \phi(\hat{n} - 1)a^\dagger , \quad (3.5)$$

or its H.c., valid for any boson $a^\dagger$ and its associated number operator, and nothing more.

To determine the functions Φ_1 and Φ_2, we first utilize the Wigner–Eckart theorem in the form

$$[A^\dagger_i, A^\dagger_{12}] = 0 , \quad (3.6)$$

which leads easily to the restrictions

$$\Phi_1 = r_1 r_2 F_1(\hat{n}_{12}) , \quad (3.7)$$

$$\Phi_2 = F_2(\hat{n}_{12}) . \quad (3.8)$$

Next the commutation relation (CR)

$$[A_1, A^\dagger_{12}] = -\Omega^{-1/2}B_{21} \quad (3.9)$$

yields the form for B_{21} in terms of the same functions F_1 and F_2,

$$B_{21} = \Omega^{-1/2}a^\dagger_{21}a_1 r_2 F_1 - \Omega^{1/2}r_1 F_2 a^\dagger_2 a_{21} . \quad (3.10)$$

Because $B_{21} = B^\dagger_{12}$ and $B_{21} = B_{12}(1 \leftrightarrow 2)$, by comparing these two ways of obtaining B_{21} from B_{12}, we find

$$-\Omega F_2(\hat{n}_{12} + 1) = F_1(\hat{n}_{12}) \equiv F(\hat{n}_{12}) . \quad (3.11)$$

Thus the problem is reduced to that of the determination of a single function $F(\hat{n}_{12})$ in terms of which the generators of interest are expressed as

$$A^\dagger_{12} = a^\dagger_{12} r_1 r_2 F(\hat{n}_{12}) - \Omega^{-1}F(\hat{n}_{12})a^\dagger_2 a^\dagger_1 a_{12} , \quad (3.12)$$

$$\Omega^{1/2}B_{21} = r_1 F(\hat{n}_{12})a^\dagger_2 a_{21} + a^\dagger_{21}a_1 F(\hat{n}_{12})r_2 . \quad (3.13)$$

Finally to determine the function F, we may utilize either Eq. (2.5b) or (2.5e), which contain equivalent information. For example, the latter yields a single difference equation that can be reduced to the form (with $n_{12} \to n$)

$$1 = (n+1)(1 - \Omega^{-1}n)F^2(n) - n[1 - \Omega^{-1}(n-2)]F^2(n-1) . \quad (3.14)$$

This is to be solved subject to the boundary condition $F^2(0) = 1$, which follows from (3.14) provided $F^2(-1)$ is nonsingular, which is verified *a posteriori*.

The solution of (3.14), derived in the Appendix, is

$$F^2(n) = \Omega(\Omega + 1 - \tfrac{1}{2}n)/(\Omega - n)(\Omega - n + 1) . \quad (3.15)$$

Without loss of generality we may choose the positive square root.

IV. VACUUM IRREP FOR Sp(6)

We now show how the method of the previous section can be applied to Sp(6). As previously remarked, the reason for doing this calculation is that it turns out, as shown in the next section, to be technically equivalent to a solution for an arbitrary representation of Sp(4) and is thus worth our attention.

We start again with realizations of $SU(2)_i$, $i = 1,2,3$, as in Eqs. (3.1)–(3.3) except that the full definition (2.12) of v_i is applicable. For example,

$$\hat{v}_1 = \hat{n}_{12} + \hat{n}_{13} . \quad (4.1)$$

Then the form for A_{12} is generalized to

$$A^\dagger_{12} = a^\dagger_{12}\Phi_1(\hat{n}_1,\hat{n}_2,\hat{n}_3,\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) + a^\dagger_1 a^\dagger_2 a_{12}\Phi_2 + a^\dagger_1 a^\dagger_{23}a_{13}\Phi_3 + a^\dagger_2 a^\dagger_{13}a_{23}\Phi_4 , \quad (4.2)$$

where the $\Phi_{2,3,4}$ depend on the same set of variables as Φ_1. Below these variables will be indicated explicitly only when they take on values shifted by some integer from their reference values, e.g.,

$$\Phi_i(\hat{n}_1 + 1) \equiv \Phi_i(\hat{n}_1 + 1,\hat{n}_2,\hat{n}_3,\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) . \quad (4.3)$$

We shall, as shown below, be able to discover the detailed form of $A^\dagger_{12}$ without bringing in either of the other off-diagonal generators $A^\dagger_{13}$ or $A^\dagger_{23}$, the form of these following simply from the appropriate change of indices. It can also be verified that once the appropriate solution is uniquely specified, all CR's not explicitly utilized in the derivation are, in fact, satisfied.

The procedure to be followed therefore parallels closely that given in the previous section, with some technical complications. The application of Eq. (3.5) yields the restrictions

$$\Phi_1 = r_1 r_2 F_1(\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) , \tag{4.4}$$

$$\Phi_2 = F_2(\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) , \tag{4.5}$$

$$\Phi_3 = r_2 F_3(\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) , \tag{4.6}$$

$$\Phi_4 = r_1 F_4(\hat{n}_{12},\hat{n}_{13},\hat{n}_{23}) . \tag{4.7}$$

Furthermore the symmetry of $A^\dagger_{12}$ under the interchange $1 \leftrightarrow 2$ implies ($i = 1,2$)

$$F_i(n_{12},n_{13},n_{23}) = F_i(n_{12},n_{23},n_{13}) \tag{4.8}$$

and

$$F_3(n_{12},n_{13},n_{23}) = F_4(n_{12},n_{23},n_{13}) . \tag{4.9}$$

The CR (3.9) together with the requirement

$$B^\dagger_{12} = B_{12}(1 \leftrightarrow 2) \tag{4.10}$$

provides the restrictions

$$-\Omega F_2(n_{12}+1) = F_1 \equiv F , \tag{4.11}$$

$$F_4(n_{13}-1,n_{23}+1) = F_3 \equiv G , \tag{4.12}$$

as well as the final forms

$$A^\dagger_{12} = a^\dagger_{12} r_1 r_2 F - \Omega^{-1} F a^\dagger_2 a^\dagger_1 a_{12} + a^\dagger_1 a^\dagger_{23} a_{13} r_2 G + r_1 G a^\dagger_2 a^\dagger_{13} a_{23} , \tag{4.13}$$

$$\Omega^{1/2} B_{21} = a^\dagger_{12} a_1 r_2 F + r_1 F a^\dagger_2 a_{12} - \Omega a^\dagger_{23} a_{13} r_1 (\hat{n}_1 - 1) r_2 G + G a^\dagger_2 a_1 a^\dagger_{13} a_{23} . \tag{4.14}$$

The remaining conditions for the determination of the functions F and G can be derived from the CR

$$[B_{21},B_{12}] = (2\Omega)^{-1}(N_2 - N_1) = (2\Omega)^{-1}[2n_2 - 2n_1 + v_2 - v_1] . \tag{4.15}$$

In working out the commutator, we get some terms that cancel identically. The remaining terms either (i) involve operators not diagonal in the basis studied, whose coefficients must therefore vanish, or (ii) diagonal operators that can have a general dependence on n_{ij} but at most a polynomial structure in n_1 and n_2, which can be compared with the right-hand side of (4.15). Terms of type (i), which did not occur in Sec. III, end up yielding only two distinct relations

$$\frac{FG(n_{12}+1)}{F(n_{13}-1,n_{23}+1)G} = \frac{\Omega - v_1 + 2}{\Omega - v_1} , \tag{4.16}$$

$$\frac{F(n_{12}-1,n_{13}-1,n_{23}+1)G}{F(n_{12}-1)G(n_{12}-1)} = \frac{\Omega - v_2 + 2}{\Omega - v_2} . \tag{4.17}$$

These equations can be combined so as to yield

$$\frac{F^2(n_{13}-1,n_{23}+1)}{F^2} = \frac{(\Omega - v_1)(\Omega - v_2 + 1)}{(\Omega - v_1 + 2)(\Omega - v_2 - 1)} , \tag{4.18}$$

$$\frac{G^2(n_{12}+1)}{G^2} = \frac{(\Omega - v_1 + 2)(\Omega - v_2 + 1)}{(\Omega - v_1)(\Omega - v_2 - 1)} , \tag{4.19}$$

which further imply

$$F^2(n_{12},n_{13},n_{23}) = [(\Omega - v_1)(\Omega - v_1 + 1)(\Omega - v_2)(\Omega - v_2 + 1)]^{-1} \times \phi^2(n_{12},v_3) , \tag{4.20}$$

$$G^2(n_{12},n_{13},n_{23}) = [(\Omega - v_1 + 1)(\Omega - v_1 + 2) \times (\Omega - v_2)(\Omega - v_2 + 1)]^{-1} \times \psi^2(n_{13},n_{23}) . \tag{4.21}$$

A reduction from three to two variables is thus achieved.

Terms of type (ii) yield the difference equations that generalize (3.14). We find

$$\Omega = (n_{12}+1)(\Omega - v_1)F^2 - n_{12}(\Omega - v_1 + 2)F^2(n_{12}-1) + \Omega n_{13}(n_{23}+1)(\Omega - v_1 + 2)G^2 - n_{23}(n_{13}+1)(\Omega - v_1)G^2(n_{13}+1,n_{23}-1) . \tag{4.22}$$

We also find the same equation with $1 \leftrightarrow 2$ and a third equation that can be obtained by combining the other two.

Further progress depends on the manipulation of Eqs. (4.18)–(4.22). By combining (4.22) with the equation obtained by $(1 \leftrightarrow 2)$ and remembering (4.12), we can eliminate $G^2(n_{13}+1,n_{23}-1)$ and obtain an equation for G^2, namely,

$$n_{13}(n_{23}+1)G^2 = \frac{1}{2} + \frac{n_{12}(\Omega - v_1 + 1)(\Omega - v_2 + 2)}{\Omega(2\Omega - v_1 - v_2 + 2)} F^2(n_{12}-1) - \frac{(1+n_{12})(\Omega - v_1)(\Omega - v_2 + 1)}{\Omega(2\Omega - v_1 - v_2 + 2)} F^2 . \tag{4.23}$$

Replacing $n_{12} \to n_{12}+1$ in this equation, and utilizing Eq. (4.19), we are led, after some algebra, to a second-order difference equation for F^2. With the definitions

$$\Omega - v_i \equiv u_i , \quad i = 1,2 , \tag{4.24}$$

this equation is

$$\Omega(u_1 + u_2 + 1) = \frac{-(2+n_{12})u_1(u_1-1)u_2(u_2-1)}{(u_1+u_2)} F^2(n_{12}+1) + \frac{(1+n_{12})2u_1(u_1+1)u_2(u_2+1)(u_1+u_2+1)}{(u_1+u_2)(u_1+u_2+2)} F^2 - \frac{n_{12}(u_1+1)(u_1+2)(u_2+1)(u_2+2)}{(u_1+u_2+2)} F^2(n_{12}-1) . \tag{4.25}$$

The solution of this equation, given in the Appendix is

$$F^2(n_{12},n_{13},n_{23})=\frac{\Omega}{8}\frac{(2\Omega+2-n_{12}-v_3)(2\Omega-2n_{12}-v_3)(2\Omega+2-2n_{12}-v_3)}{(\Omega-v_1)(\Omega-v_1+1)(\Omega-v_2)(\Omega-v_2+1)}. \tag{4.26}$$

This reduces to Eq. (3.15), as it should, if we set $n_{13}=n_{23}=v_3=0$. If we insert (4.26) into (4.23), we can calculate G^2. Some details are again found in the Appendix. The result is

$$G^2=(\Omega-n_{23}+1)(\Omega-n_{13}+2)[4(\Omega-n_{23})(\Omega-n_{13}+1)-(2\Omega+2-v_3)(2\Omega-v_3)]$$
$$\times[8n_{13}(n_{23}+1)(\Omega-v_2)(\Omega-v_2+1)(\Omega-v_1+1)(\Omega-v_1+2)]^{-1}. \tag{4.27}$$

V. GENERAL SOLUTION FOR Sp(4)

We initiate this discussion by some general remarks applicable to Sp$(2d)$ for any d. A basis for a more general irrep than that displayed in (2.6) is obtained by replacing the vacuum state $|0\rangle$ by an irrep of U(d) describing some number $\mathcal{N}$ of unpaired fermions. For each $\mathcal{N}$, we have a single irrep, the antisymmetric one, of U$(2\Omega d)$, which decomposes under U$(2\Omega d)\supset$U(d) into irrep $[\mathcal{N},0,\ldots]$, $[\mathcal{N}-1,1,0,\ldots]\cdots[1,1,\ldots,1]$ as long as $\mathcal{N}<2\Omega=2j+1$, and with one further set of restrictions specified below. In general each of these irreps will occur more than once. Let us designate an arbitrary such representation as $|\lambda_1\cdots\lambda_d(\alpha)\rangle$, where $\lambda_1\cdots\lambda_d$ is a partition and (α) designates the additional quantum numbers necessary to specify a row of the representation. By letting the double product of pair creation operators in (2.6) act on the set $|\lambda_1\cdots\lambda_d(\alpha)\rangle$, we obtain a set of nonorthonormal states $|n_1\cdots n_{d-1,d};\lambda_1\cdots\lambda_d(\alpha)\rangle$, which form a basis for an irrep of Sp$(2d)$ as long as one further set of conditions is satisfied by the $|\lambda_1\cdots\lambda_d(\alpha)\rangle$. We must replace (2.7) by the less restrictive conditions

$$A_i|\lambda_1\cdots\lambda_d(\alpha)\rangle=A_{ik}|\lambda_1\cdots\lambda_d(\alpha)\rangle=0. \tag{5.1}$$

We do not wish to enter into a detailed discussion of the consequences of (5.1), since this will only deflect us from our main path, and we intend to return to matters related to these conditions in later work. For present purposes, the main effect of (5.1) is to reduce the multiplicity of *occurrence* of the irreps of U(d) described above.

The associated generalization of (2.9) is a set of states in a direct product space

$$|n_1\cdots n_{d-1,d};\lambda_1\cdots\lambda_d(\alpha))$$
$$=|n_1\cdots n_{d-1,d})\otimes|\lambda_1\cdots\lambda_d(\alpha)), \tag{5.2}$$

where $|n_1\cdots n_{d-1,d})$ are, in effect, the states (2.9) and $|\lambda_1\cdots\lambda_d(\alpha))$ are a basis for an irrep of U(d) with generators

$$\hat{N}_i\to\mathcal{N}_i, \tag{5.3}$$

$$\hat{N}_{ik}\to\mathcal{N}_{ik}. \tag{5.4}$$

It is thus assumed that $\mathcal{N}_i$ and $\mathcal{N}_{ik}$ commute with all boson operators. In this case, it is natural to ask for a realization of (5.3) and (5.4). The solution of a similar problem is well known within the framework of the orthogonal algebras that occur in the usual shell model[14] and a similar solution can be constructed here. This question is intimately tied to the conditions (5.1) and thus is subsumed by the the previous commitment to an independent discussion. It suffices to guarantee the consistency of the requirements and to proceed.

Let us now consider the case of Sp(4). Before springing our main point that the results for the vacuum irrep of Sp(6) essentially solves the general Sp(4) problem, let us try to see how the results of Sec. III *need* to be generalized. For example, a realization of SU$(2)_i$ in the basis (5.2) is obtained by writing $(i=1,2)$

$$N_i=2\hat{n}_i+\hat{n}_{12}+\mathcal{N}_i, \tag{5.5}$$

$$A_i^\dagger=a_i^\dagger r_i, \tag{5.6}$$

$$r_i=[1-\Omega^{-1}(\hat{n}_i+\hat{n}_{12}+\hat{\mathcal{N}}_i)]^{-1/2}, \tag{5.7}$$

which also satisfies (5.1).

Imposing the required selection rules stated subsequent to (3.4) the latter is generalized to

$$A_{12}^\dagger=a_{12}^\dagger\Phi_i(\hat{n}_1,\hat{n}_2,\hat{n}_{12},\mathcal{N}_1,\mathcal{N}_2)+a_1^\dagger a_2^\dagger a_{12}\Phi_2$$
$$+a_1^\dagger\mathcal{N}_{21}\Phi_3+a_2^\dagger\mathcal{N}_{12}\Phi_4. \tag{5.8}$$

Here it is clearly assumed that $\mathcal{N}_1$ and $\mathcal{N}_2$ are among the operators diagonal in $|\lambda_1\lambda_2(\alpha))$. (In fact since $\lambda_1+\lambda_2=\mathcal{N}_1+\mathcal{N}_2$, it suffices to choose $\alpha=\mathcal{N}_1-\mathcal{N}_2$.) The analogy between (4.2) and (5.8) is striking. If we introduce the *purely formal* correspondence

$$\hat{n}_{i3}\to\mathcal{N}_i\quad(i=1,2), \tag{5.9}$$

$$a_{13}^\dagger a_{32}\to\mathcal{N}_{12}=\mathcal{N}_{21}^\dagger \tag{5.10}$$

(and suppress the operator $\hat{n}_3$), it is straightforward to verify that Eqs. (4.4)–(4.12) can be copied under this correspondence, and we end up with the analogs of (4.13) and (4.14), namely,

$$A_{12}^\dagger=a_{12}^\dagger r_1r_2F-\Omega^{-1}Fa_2^\dagger a_1^\dagger a_{12}$$
$$+a_1^\dagger\mathcal{N}_{21}r_2G+r_1Ga_2^\dagger\mathcal{N}_{12}, \tag{5.11}$$

$$\Omega^{1/2}B_{21}=a_{12}^\dagger a_1r_2F+r_1Fa_2^\dagger a_{12}$$
$$-\Omega\mathcal{N}_{21}r_1(\hat{n}_1-1)r_2G+Ga_2^\dagger a_1\mathcal{N}_{12}, \tag{5.12}$$

where, e.g.,

$$F=F(\hat{n}_{12},\mathcal{N}_1,\mathcal{N}_2), \tag{5.13}$$

and similarly for G.

In fact because the CR

$$[\mathcal{N}_{12},\mathcal{N}_{21}]=\mathcal{N}_1-\mathcal{N}_2 \tag{5.14}$$

also emerges from the correspondence (5.9) and (5.10), since

$$[a_{13}^\dagger a_{32},a_{32}^\dagger a_{13}]=\hat{n}_{13}-\hat{n}_{23}, \tag{5.15}$$

the remainder of the calculation of Sec. IV goes through unchanged provided we make the replacements

$$\hat{v}_3\to\mathcal{N}_1+\mathcal{N}_2, \tag{5.16}$$

$$\hat{n}_{23}(1+\hat{n}_{13})\to\mathcal{N}_{21}\mathcal{N}_{12}, \tag{5.17}$$

$$\hat{n}_{13}(1+\hat{n}_{23})\to\mathcal{N}_{12}\mathcal{N}_{21}, \tag{5.18}$$

in Eqs. (4.16)–(4.17), where the right-hand sides of (5.16)–(5.18) are clearly diagonal in the basis (5.2). With these specifications, we have completed the task of this section.

Some remarks concerning whether the *tour de force* of this section can be extended to higher values of d will be undertaken in the concluding section (Sec. VII).

VI. REMARKS CONCERNING Sp(*2d,R*)

We take the generators of Sp($2d,R$) in the form

$$A_i^\dagger = (2\sqrt{\Omega})^{-1} \sum_{s=1}^{2\Omega} b_{si}^\dagger b_{si}^\dagger , \tag{6.1a}$$

$$A_{ik}^\dagger = (2\Omega)^{-1/2} \sum_s b_{si}^\dagger b_{sk}^\dagger , \tag{6.1b}$$

$$N_{ik} = \sum_s b_{si}^\dagger b_{sk}^\dagger , \tag{6.2}$$

where 2Ω is an integer, $i = 1,\ldots d$, and the $b_{si}^\dagger$ are boson creation operators. If we compute the analogs of (2.5) in terms of the operators (6.1) and (6.2), we find

$$[A_i, A_k^\dagger] = \delta_{ik}[1 + (N_i/\Omega)] , \tag{6.3a}$$

$$[A_{ik}, A_{ik}^\dagger] = 1 + (2\Omega)^{-1}(N_i + N_k) , \tag{6.3b}$$

$$[A_{ki}, A_{li}^\dagger] = (2\Omega)^{-1}N_{lk} , \quad l \neq k , \tag{6.3c}$$

$$[A_i, A_{ki}^\dagger] = (\sqrt{2}\Omega)^{-1}N_{ki} , \tag{6.3d}$$

$$[N_{ik}, N_{ki}] = N_i - N_k , \tag{6.3e}$$

$$[N_{ik}, N_{kl}] = N_{il}, \quad i \neq l . \tag{6.3f}$$

This differs from the corresponding algebra (2.5) only in the sign of Ω. It follows that the results of Secs. III–V hold also for Sp($4,R$) and Sp($6,R$) as soon as we replace $\Omega \to -\Omega$ in all formulas. As expected, the various radicals involved in the functions F and G become real for all positive values of the various occupation numbers.

VII. CONCLUDING REMARKS

The natural question at this point is the extent to which we can generalize the methods described to higher d. In this section we shall provide only the start to an answer to this question by discussing first, in outline, the construction of vacuum realization of Sp(8) and then confronting the problem of whether this provides the general realization for Sp(6): It does not, but it "helps."

The essential step of our method is to write down a general form for the generator $A_{12}^\dagger$. This form for Sp(8) is

$$A_{12}^\dagger = \sum_{i=1}^{10} T_i \Phi_i(\hat{n}_1 \cdots \hat{n}_4, \hat{v}_1 \cdots \hat{v}_4, \Lambda_2, \Lambda_4) , \tag{7.1}$$

where Λ_2 and Λ_4 are defined in (2.14) and (2.15) and $\{T_i\}$ is the ordered set

$$\{T_i\} = \{a_{12}^\dagger, a_1^\dagger a_2^\dagger a_{12}, a_1^\dagger a_{23}^\dagger a_{13}, a_2^\dagger a_{13}^\dagger a_{23}, a_1^\dagger a_{24}^\dagger a_{14}, a_2^\dagger a_{14}^\dagger a_{24}, a_{42}^\dagger a_{31}^\dagger a_{43}, a_{32}^\dagger a_{41}^\dagger a_{43}, a_{43}^\dagger a_1^\dagger a_2^\dagger a_{42} a_{31} , a_{43}^\dagger a_1^\dagger a_2^\dagger a_{32} a_{41}\} . \tag{7.2}$$

The only question that arises is whether there are additional members of the set (7.2) that we have failed to recognize. We argue as follows that this is not the case. Let us then try to add to the set (7.2). The simplest additional possibility is

$$a_{12}^\dagger (a_{23}^\dagger a_{14}^\dagger a_{24} a_{13} + 1 \leftrightarrow 2) . \tag{7.3}$$

But (7.3) can be written as $a_{12}^\dagger \Lambda_{12}$ + terms of the form already included in (7.1). The same argument holds *a fortiori* for monomials of higher degree in the $a^\dagger$ and a satisfying the allowed selection rules. Thus the set (7.2) is complete. The task of determining the Φ_i is being studied. It is complicated by the fact that Λ_2 and Λ_4 are not diagonal in the direct product boson basis.

Supposing that we had the result, we could subsequently ask to what extent it provides us with a solution of the general Sp(6) problem. There are several ways of seeing that something is lacking. The vacuum irrep of Sp(8) is determined by ten quantum numbers n_i and n_{ij} or their equivalent, but the general irrep of Sp(6) requires 12, six for the boson part and six for the U(3) basis $|\lambda_1, \lambda_2, \lambda_3(\alpha))$ [$\frac{1}{2}d(d+1)$ for each part]. [In the case we solved relating Sp(6) and Sp(4) the quantum number count was the same!] Nevertheless, if in the expression (7.1) we enter the correspondence

$$\hat{n}_{i4} \to \mathcal{N}_i , \tag{7.4}$$

$$a_{i4}^\dagger a_{4j} \to \mathcal{N}_{ij} \quad (i,j = 1,2,3) , \tag{7.5}$$

we should thereby obtain a set of irreps of Sp(6) in which the associated U(3) has at most two rows. In this restricted form the relationship between Sp($2d+2$) and Sp($2d$) may generalize to any d.

Finally, to see what is missing, consider a typical term, e.g., T_5. Under the correspondence (7.5) we have

$$a_1^\dagger a_{24}^\dagger a_{41} \to a_1^\dagger \mathcal{N}_{21} . \tag{7.6}$$

But the quantity $a_1^\dagger \mathcal{N}_{23} \mathcal{N}_{31}$ is another independent term satisfying the same selection rules and must be included multiplied by an independent function of the diagonal operators. On the face of it, therefore, the problem of finding the general irrep of Sp(6) is considerably more complicated than finding the vacuum irrep of Sp(8).

ACKNOWLEDGMENT

This work was supported in part by the U.S. Department of Energy under Contract No. 40132-5-20441.

APPENDIX: SOLUTION OF DIFFERENCE EQUATIONS

We first consider the difference equation (3.14). We substitute

$$(n+1)F^2 = f^2 , \quad f^2(0) = 1 . \tag{A1}$$

The equation for f^2 can be displayed in the form

$$1 + \Omega^{-1} f^2 = [1 - \Omega^{-1}(n-1)] f^2(n) - [1 - \Omega^{-1}(n-2)] f^2(n-1) . \tag{A2}$$

This suggests the substitution

$$[1 - \Omega^{-1}(n-1)] f^2 = g^2 , \quad g^2(0) = 1 + \Omega^{-1} , \tag{A3}$$

which simplifies (A2) to the form

$$1 - \Omega^{-1}(n-1) = g^2(n)(1 - \Omega^{-1} n) - g^2(n-1)[1 - \Omega^{-1}(n-1)]$$

$$\equiv h^2(n) - h^2(n-1)\,, \tag{A4}$$

$$h^2(0) = 1 + \Omega^{-1}\,. \tag{A5}$$

Equation (A4) shows clearly that $h^2(n)$ is a quadratic function of n, which turns out to have the form

$$\begin{aligned} h^2(n) &= (1+\Omega^{-1}) + [1 + (2\Omega)^{-1}]n - (2\Omega)^{-1}n^2 \\ &= (1+n)[1+\Omega^{-1} - (2\Omega)^{-1}n]\,. \end{aligned} \tag{A6}$$

Unraveling the various substitutions yields Eq. (3.15).

Next we study the solution of Eq. (4.25). First we substitute Eq. (4.20). In the resulting equation for $\phi^2(n_{12},v_3)$, introduce the notation

$$\mu_1 + \mu_2 = 2 - v_3 - 2n_{12} \equiv \mu_3 - 2n_{12} \tag{A7}$$

and put $n_{12}\to n$, $\mu_3 \to u$. Then with the further substitution (suggested by the form of the equation)

$$\phi^2(n,v_3) = \Omega f^2(n,u)(u-2n)(u-2n+2)\,, \tag{A8}$$

we obtain the final difference equation

$$\begin{aligned} (u-2n+1) = &-(2+n)f^2(n+1,u)(u-2n+2) \\ &+ (1+n)f^2(n,u)(u-2n+1) \\ &- nf^2(n-1,u)(u-2n+4)\,. \end{aligned} \tag{A9}$$

Equation (A9) is a linear combination of first and second differences. It is straightforward to verify that it has the solution

$$f^2(n,u) = \tfrac{1}{8}(u+2-n)\,. \tag{A10}$$

This leads directly back to (4.26).

Substituting now into (4.23) yields an expression for G^2. The simplest way to evaluate this expression is as follows: Also substitute Eq. (4.21) for G^2. We thus obtain the equation

$$\begin{aligned} &n_{13}(n_{23}+1)\psi^2(n_{13},n_{23}) \\ &\quad = \tfrac{1}{2}(u_1+1)(u_1+2)u_2(u_2+1) \\ &\qquad + (n_{12}/8)(u_3+3-n_{12}) \\ &\qquad \times (u_3+4-2n_{12})u_2(u_1+1) \\ &\qquad - [(1+n_{12})/8](u_3+2-n_{12}) \\ &\qquad \times (u_3-2n_{12})(u_1+2)(u_2+1)\,. \end{aligned} \tag{A11}$$

Here the left-hand side is independent of n_{12}. Therefore so is the right-hand side. This can be verified explicitly with great pain, but it is simplest to set $n_{12}=0$ everywhere here. This yields the result

$$\begin{aligned} &\psi^2(n_{12},n_{23}) \\ &\quad = [(\Omega - n_{13}+2)(\Omega - n_{23}+1)/8n_{13}(n_{23}+1)] \\ &\qquad \times [4(\Omega - n_{13}+1)(\Omega - n_{23}) \\ &\qquad - (2\Omega + 2 - v_3)(2\Omega - v_3)]\,, \end{aligned} \tag{A12}$$

and this leads immediately to (4.27).

[1]For a recent review of the very extensive literature, see D. J. Rowe, Rep. Prog. Phys. **48**, 1419 (1985).

[2]D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984); D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).

[3]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982); **25**, 1638 (1984); **26**, 2705 (1985). The last reference contains an excellent short review of the previous literature.

[4]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985); M. Moshinsky, "Boson realization of symplectic algebras," to be published.

[5]J. N. Ginocchio, Ann. Phys. (NY) **126**, 234 (1980).

[6]C. L. Wu, D. H. Feng, X. G. Chen, J. Q. Chen, and M. W. Guidry, submitted to Phys. Lett. B; submitted to Ann. Phys. (NY).

[7]A. Klein, T. D. Cohen, and C. T. Li, Ann. Phys. (NY) **141**, 382 (1982).

[8]D. Bonatsos and A. Klein, Phys. Rev. C **31**, 992 (1985); D. Bonatsos and A. Klein, Ann. Phys. (NY), to be published. In the former reference, the algebra Sp(6) is incorrectly identified as SO(7).

[9]J. P. Draayer and K. J. Weeks, Ann. Phys. (NY) **156**, 41 (1984). Previous literature may be traced from this paper.

[10]K. T. Hecht and J. P. Elliot, Nucl. Phys. **438**, 29 (1985); K. T. Hecht, *Proceedings of the VIIIth Symposium on Nuclear Physics*, Oaxtepec, Mexico, 1985, to be published.

[11]F. J. Dyson, Phys. Rev. **102**, 1217 (1956).

[12]D. Agassi, Nucl. Phys. A **116**, 49 (1968); D. Shutte and K. Bleuler, **119**, 221 (1968); J. A. Evans and N. Kraus, Phys. Lett. B **37**, 455 (1972).

[13]T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).

[14]E. R. Marshalek, Nucl. Phys. A **357**, 398 (1981); T. Suzuki and K. Matsuyanagi, Prog. Theor. Phys. **56**, 1156 (1976); H. B. Geyer and F. T. W. Hahne, Phys. Lett. B **90**, 6 (1980).

Finite-dimensional irreducible representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis

T. D. Palev[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 5 February 1986; accepted for publication 26 March 1986)

A concept of a Gel'fand–Zetlin pattern for the Lie superalgebra sl(1,3) is introduced. Within every finite-dimensional irreducible sl(1,3) module the set of the Gel'fand–Zetlin patterns constitute an orthonormed basis, called a Gel'fand–Zetlin basis. Expressions for the transformation of this basis under the action of the generators are written down for every finite-dimensional irreducible representation.

I. INTRODUCTION

Using the results obtained in Refs. 1 and 2 (hereafter referred to as I and II) we introduce in the present paper a concept of a Gel'fand–Zetlin pattern (GZ pattern) for the special linear Lie superalgebra (LS) sl(1,3). Similarly as in the representation theory of the classical Lie algebras, the GZ patterns constitute a basis in the finite-dimensional irreducible sl(1,3) modules. We write explicit expressions for the transformation of this basis, called a Gel'fand–Zetlin basis (GZ basis), under the action of the generators for every finite-dimensional irreducible representation.

From the point of view of the sl(1,3) representations, the present paper contains no new results comparing to those obtained in I and II. However, here we succeed in presenting the results in a much more compact form. This is mainly due to the fact that in this paper we introduce a unique notation for the basis vectors within the irreducible sl(1,3) module, whereas in I and II we characterized the basis through its properties with respect to the even subalgebra gl(3). There we were considering a given finite-dimensional irreducible sl(1,3) module V as a representation space of $gl(3) \subset sl(1,3)$ and represented it as a direct sum of its irreducible gl(3) submodules V_k,

$$V=\sum_k \oplus V_k, \quad k=1,2,...,8. \tag{1.1}$$

As a basis Γ_k within every V_k we choose the gl(3) Gel'fand–Zetlin basis[3] and define an orthonormal GZ basis in V to be

$$\Gamma = \bigcup_k \Gamma_k. \tag{1.2}$$

Thus, the basis vectors, used in I and II, are

$$\left|\begin{matrix} m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11}\end{matrix}\right\rangle_k, \quad k=1,...,8, \tag{1.3}$$

where k distinguishes between the different gl(3) submodules V_k of V. In terms of this notation we did not succeed in expressing the transformation properties of Γ under the action of any generator $e \in sl(1,3)$ in a compact form. In the case of the typical representations we used four relations (one relation for each group of indices: $k=1$, $k=2,3,4$, $k=5,6,7$, and $k=8$) in order to represent

$$e\left|\begin{matrix} m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11}\end{matrix}\right\rangle_k \tag{1.4}$$

as a linear combination of vectors from Γ. For each of the three classes of nontypical representations we also used four relations, so that altogether we needed 16 different equations in order to express the transformation of the GZ basis under the action of only one generator. In the present paper we succeed in writing down all these 16 relations in terms of only one. This considerably simplifies the final results, making them more transparent and putting them in a form similar to that known from the Lie algebra representation theory.

In Sec. II we introduce as a first step an intermediate basis, which is appropriate for the typical representations. However, it contains some superficial vectors in the nontypical cases. In Sec. III we define a GZ pattern for sl(1,3) and write down the transformation of the GZ basis under the action of the genrators for every (typical or nontypical) irreducible sl(1,3) module [see (3.16)–(3.21)].

II. INDUCED REPRESENTATIONS OF sl(1,3)

Throughout the paper we use terminology, notation, and assertions that were introduced in I or II. Therefore, here we briefly mention some notation and results, stressing a certain change in the notation.

To begin with we recall that the LS sl(1,3) can be defined through its four-dimensional representation as follows. Let e_{AB}, $A,B=0,1,2,3$, be a 4×4 matrix with 1 on the A th row and the B th column and zero elsewhere. Then sl(1,3) is the linear span of its even generators

$$E_{ij}=e_{ij}+\delta_{ij}e_{00}, \quad i,j=1,2,3, \tag{2.1}$$

which are the generators of the Lie algebra gl(3), and the odd generators

$$e_{0i},e_{i0}, \quad i=1,2,3. \tag{2.2}$$

By $V([m]_3)$, $[m]_3 \equiv [m_{13},m_{23},m_{33}]$, we denote the irreducible gl(3) module with $m_{i3}=\Lambda(E_{ii})$, $i=1,2,3$, being the coordinates of the highest weight Λ in the dual to E_{11},E_{22},E_{33} basis of the Cartan subalgebra. We always

a) Permanent address: Institute of Nuclear Research and Nuclear Energy, boul. Lenin 72, 1184 Sofia, Bulgaria.

choose the gl(3) GZ basis[3] in $V([m]_3)$:

$$\left|\begin{matrix} m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle \equiv \left|\begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix}\right\rangle. \tag{2.3}$$

The complex numbers m_{13},m_{23},m_{33} are fixed. They label the irreducible gl(3) module. In general, the representations corresponding to different triplets $[m]_3$ are inequivalent. The numbers m_{12}, m_{22}, and m_{11} label the basis vectors in $V([m]_3)$. They run over all possible values, consistent with the "betweenness condition" ($\mathbb{Z}_+ =$ all non-negative integers)

$$\begin{aligned} &m_{13}-m_{12},m_{12}-m_{23},m_{23}-m_{22}, \\ &\quad m_{22}-m_{33},m_{12}-m_{11},m_{11}-m_{22}\in\mathbb{Z}_+. \end{aligned} \tag{2.4}$$

In I we have studied in detail the so-called induced modules[4] over the LS sl(1,3). The relevance of these modules stems from the observation that every finite-dimensional irreducible representation of a basic LS [5] can be realized either in some induced module (typical representations) or in a factor module of it (nontypical representations). The induced modules of sl(1,3) are labeled by three complex numbers,

$$[m]_4 \equiv [m_{14},m_{24},m_{34}], \tag{2.5}$$

satisfying the conditions

$$m_{14}-m_{24},m_{24}-m_{34}\in\mathbb{Z}_+. \tag{2.6}$$

By $\overline{V}([m]_4)$ we denote the induced sl(1,3) module corresponding to $[m]_4$. We point out that in I and II instead of $\overline{V}([m]_4)$ we were writing $\overline{V}([m]_3)$ and $[m]_3 \equiv [m_{13},m_{23},m_{33}]$ instead of (2.5).

Since every $\overline{V}([m]_4)$ is also a gl(3) module and every finite-dimensional gl(3) module is completely reducible, one can decompose $\overline{V}([m]_4)$ into a direct sum of irreducible gl(3) modules [I,(5.15)]:

$$\begin{aligned} \overline{V}([m]_4) = V([m]_4) \oplus \sum_{i=1}^{3} \oplus V([m-1]_4^i) \\ \oplus \sum_{i=1}^{3} V([m-1]_4^{-i} \oplus V([m-2]), \end{aligned} \tag{2.7}$$

where here and everywhere in the paper ($c\in\mathbb{C}$) we use the notation

$$[m+c]_4 = [m_{14}+c,m_{24}+c,m_{34}+c], \tag{2.8}$$

$$[m+c]_n = [m_{1n}+c,\ldots,m_{nn}+c], \quad n=2,3, \tag{2.9}$$

$$[m]_4^{\pm i} = [m_{14}\pm\delta_{1i},m_{24}\pm\delta_{2i},m_{34}\pm\delta_{3i}], \tag{2.10}$$

$$[m]_n^{\pm i} = [m_{1n}\pm\delta_{1i},\ldots,m_{nn}\pm\delta_{ni}], \quad n=2,3. \tag{2.11}$$

In order to indicate that the gl(3) module $V([m]_3)$ is a submodule of the sl(1,3) module $\overline{V}([m]_4)$, i.e., it is a direct summand in (2.7), we modify the notation for $V([m]_3)$ and write

$$V([m]_3) = V\begin{pmatrix}[m]_4 \\ [m]_3\end{pmatrix}. \tag{2.12}$$

If $V([m]_3)$ is not a direct summand in (2.7), we assume

$$V\begin{pmatrix}[m]_4 \\ [m]_3\end{pmatrix} = 0\in\overline{V}([m]_4). \tag{2.13}$$

In terms of this notation (2.7) can be written as

$$\overline{V}([m]_4) = \sum_{[m]_3} \oplus V\begin{pmatrix}[m]_4 \\ [m]_3\end{pmatrix}, \tag{2.14}$$

where the sum is over all triplets $[m_{13},m_{23},m_{33}]$, satisfying the conditions

$$m_{i4}-m_{i3}=0,1,2, \quad i=1,2,3, \tag{2.15}$$

$$|(m_{i4}-m_{i3})-(m_{j4}-m_{j3})|\leqslant 1, \quad i,j=1,2,3, \tag{2.16}$$

$$\frac{1}{2}\sum_{i=1}^{3}(m_{i4}-m_{i3})\in\mathbb{Z}_+ \quad (\text{or } =0,1,2,3), \tag{2.17}$$

$$m_{13}-m_{23},m_{23}-m_{33}\in\mathbb{Z}_+. \tag{2.18}$$

We observe from (2.14) that the spectrum of gl(3) in $\overline{V}([m]_4)$ is simple, i.e., $\overline{V}([m]_3)$ is a direct sum of inequivalent gl(3) modules. Therefore, the decomposition (2.14) is also unique. Hence, the basis $\Gamma([m]_4)$ in $\overline{V}([m]_4)$ is uniquely defined in terms of the basis vectors of each $V(\begin{smallmatrix}[m]_4 \\ [m]_3\end{smallmatrix})$. As such a basis $\Gamma(\begin{smallmatrix}[m]_4 \\ [m]_3\end{smallmatrix})$ in $V(\begin{smallmatrix}[m]_4 \\ [m]_3\end{smallmatrix})$ we choose a GZ basis and introduce a new notation for the vector

$$\left|\begin{matrix} m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle \in \Gamma\begin{pmatrix}[m]_4 \\ [m]_3\end{pmatrix},$$

namely

$$\left|\begin{matrix} m_{14},m_{24},m_{34} \\ m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle \equiv \left|\begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix}\right\rangle. \tag{2.19}$$

Assuming that the different gl(3) submodules in the sum (2.14) are orthogonal to each other, we obtain an orthonormal basis in $\overline{V}([m]_4)$:

$$\Gamma([m]_4) = \bigcup_{[m]_3} \Gamma\begin{pmatrix}[m]_4 \\ [m]_3\end{pmatrix}. \tag{2.20}$$

The union in (2.20) is over all $[m]_3$, which are in agreement with the conditions (2.15)–(2.18). We call the vectors (2.19) from the basis $\Gamma([m]_4)$ induced patterns (I-patterns) or I-basis vectors in $\overline{V}([m]_4)$. In order to characterize all I-patterns, i.e., the basis in $\overline{V}([m]_4)$, one has to add to the conditions (2.15)–(2.18) also the inclusions (2.4) and (2.6). Thus, we have the following proposition.

Proposition 2.1: The vector (2.19) is an I-pattern, i.e., a basis vector in $\overline{V}([m]_4)$, iff the numbers m_{ij}, which characterizes it, satisfy the conditions

(1) $m_{14}-m_{24},m_{24}-m_{34}\in\mathbb{Z}_+$,

(2) $m_{i4}-m_{i3}=0,1,2, \quad i=1,2,3$,

(3) $|(m_{i4}-m_{i3})-(m_{j4}-m_{j3})|\leqslant 1, \quad i,j=1,2,3$,

(4) $\frac{1}{2}\sum_{i=1}^{3}(m_{i4}-m_{i3})\in\mathbb{Z}_+ \quad (\text{or } =0,1,2,3)$,

(5) $m_{13}-m_{12},m_{12}-m_{23},m_{23}-m_{22},m_{22}-m_{33}$,

$$m_{12}-m_{11},m_{11}-m_{22}\in\mathbb{Z}_+. \tag{2.21}$$

Observe that the sum (2.14) contains eight direct summands. In certain cases, however, some of the terms may not satisfy (2.15)–(2.18). For instance [in the notation (2.7)],

$$\overline{V}([1,1,0]) = V([1,1,0]) \oplus V([0,0,0]) \oplus V([1,0,-1]) \oplus V([0,0,-2]) \oplus V([-1,-1,-2]). \quad (2.22)$$

Therefore, in general, $\overline{V}([m]_4)$ is a direct sum of no more than eight irreducible gl(3) submodules.

Associate with every I-pattern (2.19) the numbers

$$\xi_i \begin{pmatrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{pmatrix} \equiv \xi_i = m_{i4} - m_{i3}, \quad i = 1,2,3, \quad (2.23)$$

$$\xi \begin{pmatrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{pmatrix} \equiv \xi = \frac{1}{2} \sum_{i=1}^{3} (m_{i4} - m_{i3}), \quad (2.24)$$

$$\xi_0 = \xi_3, \quad \xi_4 = \xi_1. \quad (2.25)$$

Let, moreover,

$$\theta(x) = \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{for } x < 0, \end{cases} \quad (2.26)$$

$$S(i,j) = \begin{cases} 1, & \text{for } i \leqslant j, \\ -1, & \text{for } i > j, \end{cases} \quad (2.27)$$

$$\delta(x) = \begin{cases} 1, & \text{for } x = 0, \\ 0, & \text{for } x \neq 0, \end{cases} \quad (2.28)$$

$$l_{ij} = m_{ij} - i. \quad (2.29)$$

In terms of this notation the transformation of the basis in $\overline{V}([m]_4)$ under the action of the odd generators reads

$$e_{10} \left| \begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{2} \theta(3\xi_i - 2\xi)\theta(2-\xi)(-1)^{(\xi_i - \xi_{i-1})\xi} S(i,j)S(j,1)$$
$$\times \left| \frac{\Pi_{k \neq j=1}^{2} (l_{k2} - l_{11})(l_{k2} - l_{i3} - 1)\Pi_{k \neq i=1}^{3} (l_{k3} - l_{j2})}{(l_{12} - l_{22} - j + 1)(l_{12} - l_{22} - j + 2)\Pi_{k \neq i=1}^{3} (l_{k4} - l_{i4})} \right|^{1/2} \left| \begin{matrix} [m]_4 \\ [m-1]_3^i \\ [m-1]_2^j \\ m_{11} \end{matrix} \right\rangle, \quad (2.30)$$

$$e_{20} \left| \begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{2} \theta(3\xi_i - 2\xi)\theta(2-\xi)(-1)^{(\xi_i - \xi_{i-1})\xi} S(i,j)$$
$$\times \left| \frac{(l_{j2} - l_{11} + 1)\Pi_{k \neq j=1}^{2} (l_{k2} - l_{i3} - 1)\Pi_{k \neq i=1}^{3} (l_{k3} - l_{j2})}{(l_{12} - l_{22} - j + 1)(l_{12} - l_{22} - j + 2)\Pi_{k \neq i=1}^{3} (l_{k4} - l_{i4})} \right|^{1/2} \left| \begin{matrix} [m]_4 \\ [m-1]_3^i \\ [m-1]_2^j \\ m_{11} - 1 \end{matrix} \right\rangle, \quad (2.31)$$

$$e_{30} \left| \begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \theta(3\xi_i - 2\xi)\theta(2-\xi)(-1)^{(\xi_i - \xi_{i-1})\xi} \left| \frac{\Pi_{k=1}^{2} (l_{k2} - l_{i3} - 1)}{\Pi_{k \neq i=1}^{3} (l_{k4} - l_{i4})} \right|^{1/2} \left| \begin{matrix} [m]_4 \\ [m-1]_3^i \\ [m-1]_2 \\ m_{11} - 1 \end{matrix} \right\rangle, \quad (2.32)$$

$$e_{01} \left| \begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{2} \theta(2\xi - 3\xi_i)\theta(\xi - 1)(-1)^{(\xi_i - \xi_{i+1})(\xi - 1)} S(i,j)S(j,1)$$
$$\times (l_{i4} + 1) \left| \frac{\Pi_{k \neq j=1}^{2} (l_{k2} - l_{11} + 1)(l_{k2} - l_{i3})\Pi_{k \neq i=1}^{3} (l_{k3} - l_{j2} + 1)}{(l_{12} - l_{22} + j - 1)(l_{12} - l_{22} + j - 2)\Pi_{k \neq i=1}^{3} (l_{k4} - l_{i4})} \right|^{1/2} \left| \begin{matrix} [m]_4 \\ [m+1]_3^{-i} \\ [m+1]_2^{-j} \\ m_{11} \end{matrix} \right\rangle, \quad (2.33)$$

$$e_{02}\begin{vmatrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\theta(2\xi-3\xi_i)\theta(\xi-1)(-1)^{(\xi_i-\xi_{i+1})(\xi-1)}S(i,j)$$

$$\times(l_{i4}+1)\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m+1]_3^{-i}\\ [m+1]_2^{-j}\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.34}$$

$$e_{03}\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\theta(2\xi-3\xi_i)\theta(\xi-1)(-1)^{(\xi_i-\xi_{i+1})(\xi-1)}(l_{i4}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m+1]_3^{-i}\\ [m+1]_2\\ m_{11}+1\end{matrix}\right\rangle. \tag{2.35}$$

We wish to underline that in all the above formulas

$$\xi_i = m_{i4}-m_{i3},\quad \xi=\frac{1}{2}\sum_{i=1}^{3}(m_{i4}-m_{i3}). \tag{2.36}$$

One can check by a straightforward computation that the relations (2.30)–(2.35) hold if and only if the relations (5.27), (5.29)–(5.31), (5.36)–(5.38), (5.44), (5.46), (5.47), (5.51)–(5.53), (5.60)–(5.62), and (5.66)–(5.68) derived in I hold.

III. IRREDUCIBLE REPRESENTATIONS

Every sl(1,3) module is either irreducible or nondecomposable. In I (Proposition 3) we have shown that $\overline{V}([m]_4)$ is irreducible iff

$$m_{14}\neq 0,\quad m_{24}\neq 1,\quad m_{34}\neq 2. \tag{3.1}$$

The representations of sl(1,3), realized in these irreducible modules (and also the modules themselves) are said to be typical.[4] Therefore the relations (2.30)–(2.35) describe all typical irreducible representations of the LS sl(1,3).

Each $\overline{V}([m]_4)$ that is not irreducible contains a maximal sl(1,3) invariant subspace $\overline{I}([m]_4)\neq 0$. The factor module

$$\overline{W}([m]_4)=\overline{V}([m]_4)/\overline{I}([m]_4) \tag{3.2}$$

carries an irreducible representation of sl(1,3). All such factor modules (and also the corresponding representations) are called nontypical. It is remarkable that the typical and the nontypical representations exhaust the set of all finite-dimensional irreducible representations of sl(1,3) (see Ref. 4). Therefore, it remains to construct the nontypical representations. This task was solved in II.

The reducible modules $\overline{V}([m]_4)$ resolve into three nonintersecting classes[2]: class I,

$$\{\overline{V}([0,m_{24},m_{34}])\,|\,0\geqslant m_{24}\geqslant m_{34}\}, \tag{3.3}$$

class II,

$$\{\overline{V}([m_{14},1,m_{34}])\,|\,m_{14}\geqslant 1\geqslant m_{34}\}, \tag{3.4}$$

and class III,

$$\{\overline{V}([m_{14},m_{24},2])\,|\,m_{14}\geqslant m_{24}\geqslant 2\}. \tag{3.5}$$

The corresponding maximal invariant submodules, written in the notations of this paper, are (II, Propositions 5–7)

$$\overline{I}([0,m_{24},m_{34}])=V\begin{pmatrix}[0,m_{24},m_{34}]\\ [0,m_{24}-1,m_{34}-1]\end{pmatrix}\oplus V\begin{pmatrix}[0,m_{24},m_{34}]\\ [-1,m_{24}-1,m_{34}-2]\end{pmatrix}$$
$$\oplus V\begin{pmatrix}[0,m_{24},m_{34}]\\ [-1,m_{24}-2,m_{34}-1]\end{pmatrix}\oplus V\begin{pmatrix}[0,m_{24},m_{34}]\\ [-2,m_{24}-2,m_{34}-2]\end{pmatrix}, \tag{3.6}$$

$$\overline{I}([m_{14},1,m_{34}])=V\begin{pmatrix}[m_{14},1,m_{34}]\\ [m_{14}-1,1,m_{34}-1]\end{pmatrix}\oplus V\begin{pmatrix}[m_{14},1,m_{34}]\\ [m_{14}-2,0,m_{34}-1]\end{pmatrix}$$
$$\oplus V\begin{pmatrix}[m_{14},1,m_{34}]\\ [m_{14}-1,0,m_{34}-2]\end{pmatrix}\oplus V\begin{pmatrix}[m_{14},1,m_{34}]\\ [m_{14}-2,-1,m_{34}-2]\end{pmatrix}, \tag{3.7}$$

$$\bar{I}(m_{14},m_{24},2]) = V\binom{[m_{14},m_{24},2]}{[m_{14}-1,m_{24}-1,2]} \oplus V\binom{[m_{14},m_{24},2]}{[m_{14}-1,m_{24}-2,1]}$$

$$\oplus V\binom{[m_{14},m_{24},2]}{[m_{14}-2,m_{24}-1,1]} \oplus V\binom{[m_{14},m_{24},2]}{[m_{14}-2,m_{24}-2,0]}. \tag{3.8}$$

In order to obtain the representations of sl(1,3) in the factor spaces one has to replace in (2.30)–(2.35) all basis vectors from the maximal invariant subspace by zero (II, Corollary) or, which is the same, to project the relations (2.30)–(2.35) on the orthogonal complement to $\bar{I}([m]_4)$, which is isomorphic to $\overline{W}([m]_4)$:

$$\overline{W}([m]_4) = \overline{V}([m]_4)/\bar{I}([m]_4) = \overline{V}([m]_4)\ominus\bar{I}([m]_4). \tag{3.9}$$

From now on we assume that $\bar{I}([m]_4) = 0$ in the typical case. Then $\overline{W}([m]_4) = \overline{V}([m]_4)$ and the projection on $\overline{W}([m]_4)$ does not change the transformation properties of the basis, i.e., it preserves (2.30)–(2.35) in the typical modules. To be more precise we formulate the following proposition.

Proposition 3.1: Let $f_1, f_2,..., f_n$ be the I-basis in $\overline{V}([m]_4)$, which, let us assume for definiteness, is transformed under the action of the sl(1,3) generators $E_1,...,E_{15}$ as follows:

$$E_k f_i = \sum_{j=1}^{n} A_{ji} f_j, \quad k=1,...,15, \quad i=1,...,n. \tag{3.10}$$

Denote by P the projection operator of $\overline{V}([m]_4)$ on $\overline{W}([m]_4)$ [see (3.9)]. Then the algebra sl(1,3) transforms the basis of the factor space according to

$$E_k Pf_i = \sum_{j=1}^{n} A_{ji} Pf_j, \quad k=1,...,15, \quad i=1,...,n. \tag{3.11}$$

The proof follows from the observation that $Pf_i = 0$ if $f_i \in \bar{I}([m]_4)$ and $Pf_i = f_i$ for $f_i \in \overline{W}([m]_4)$, i.e., (3.11) means that in (3.10) one is replacing all basis vectors from $\bar{I}([m]_4)$ by zero. The nonzero vectors Pf_i, $i=1,...,n$, constitute an orthonormed basis in the factor space.

Proposition 3.2: The linear operator P, defined everywhere on $\overline{V}([m]_4)$ with the relations

$$P\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \left[1-\sum_{k=1}^{3}\delta(m_{k4}-k+1)\delta(\xi-\xi_k-1)\right]\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \tag{3.12a}$$

or, equivalently,

$$P\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \left[1-\sum_{k=1}^{3}\delta(m_{k4}-k+1)\delta(m_{k3}-k+\xi)\right]\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \tag{3.12b}$$

is a projection operator of $\overline{V}([m]_4)$ on $\overline{W}([m]_4)$.

The proof is straightforward. If $\overline{V}([m]_4)$ is a typical module, $m_{k4}-k+1\neq 0$ $\forall k=1,2,3$, and, therefore, $P=1$. In the nontypical case only one of the terms in the sum survives, so that

$$P\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \begin{cases} 0, & \text{if the vector is from } \bar{I}([m]_4), \\ \left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle, & \text{if the vector is from } \overline{W}([m]_4). \end{cases} \tag{3.13}$$

The relation (3.12b) indicates that, when $m_{k4}-k+1=0$, the I-patterns with $m_{k3}-k+\xi=0$ are annihilated by P and, therefore, these vectors do not belong to $\overline{W}([m]_4)$. Thus, we come to the following conclusion. In order to obtain the (orthonormed) basis in $\overline{W}([m]_4)$ one has to remove from the I-basis all those I-patterns, for which simultaneously $m_{k4}=k-1$ and $m_{k3}=k-\xi$ hold (an equivalent statement: the I-pattern is a basis vector in the factor space iff, whenever $m_{k4}=k-1$, then $\xi_k=\xi$). This justifies the definition below.

Definition: The table of complex numbers

$$\left|\begin{matrix} m_{14},m_{24},m_{34} \\ m_{13},m_{23},m_{33} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle \equiv \left|\begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix}\right\rangle \tag{3.14}$$

will be called a GZ pattern for sl(1,3), if the entries m_{ij} satisfy the conditions ($\mathbf{Z}_+$ = all non-negative integers)

$$\begin{aligned} &(1) \quad m_{14}-m_{24},m_{24}-m_{34}\in\mathbf{Z}_+, \\ &(2) \quad m_{13}-m_{12},m_{12}-m_{23},m_{23}-m_{22},m_{22}-m_{33}, \quad m_{12}-m_{11},m_{11}-m_{22}\in\mathbf{Z}_+, \\ &(3) \quad \frac{1}{2}\sum_{i=1}^{3}(m_{i4}-m_{i3})\in\mathbf{Z}_+, \\ &(4) \quad m_{i4}-m_{i3}=0,1,2, \quad i=1,2,3, \\ &(5) \quad |(m_{i4}-m_{i3})-(m_{j4}-m_{j3})|\leqslant 1, \quad i,j=1,2,3, \\ &(6) \quad \text{if } m_{k4}=k-1, \quad \text{then} \quad m_{k4}-m_{k3}=\sum_{i\neq k=1}^{3}(m_{i4}-m_{i3}), \quad k=1,2,3. \end{aligned} \tag{3.15}$$

All GZ patterns with a fixed upper row $[m]_4 \equiv [m_{14},m_{24},m_{34}]$ constitute an orthonormal basis in the irreducible finite-dimensional sl(1,3) module $\overline{W}([m]_4)$. We call this basis a GZ basis.

In order to obtain the transformation of the GZ basis under the action of the odd generators we use Proposition 3.2. We project all I-patterns that appear on the left-hand side and on the right-hand side of (2.30)–(2.35) on $\overline{W}([m]_4)$. After some calculations we obtain

$$\begin{aligned} e_{10}\left|\begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix}\right\rangle = &\sum_{i=1}^{3}\sum_{j=1}^{2}\theta(3\xi_i-2\xi)\theta(2-\xi)S(i,j)S(j,1) \\ &\times(-1)^{(\xi_i-\xi_{i-1})\xi}\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2} \\ &\times\left[1-\sum_{k=1}^{3}\delta(l_{k4}+1)\delta_{ki}\right]\left|\begin{matrix} [m]_4 \\ [m-1]_3^i \\ [m-1]_2^j \\ m_{11} \end{matrix}\right\rangle, \end{aligned} \tag{3.16}$$

$$\begin{aligned} e_{20}\left|\begin{matrix} [m]_4 \\ [m]_3 \\ [m]_2 \\ m_{11} \end{matrix}\right\rangle = &\sum_{i=1}^{3}\sum_{j=1}^{2}\theta(3\xi_i-2\xi)\theta(2-\xi)S(i,j) \\ &\times(-1)^{(\xi_i-\xi_{i-1})\xi}\left|\frac{(l_{j2}-l_{11}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2} \\ &\times\left[1-\sum_{k=1}^{3}\delta(l_{k4}+1)\delta_{ki}\right]\left|\begin{matrix} [m]_4 \\ [m-1]_3^i \\ [m-1]_2^j \\ m_{11}-1 \end{matrix}\right\rangle, \end{aligned} \tag{3.17}$$

$$e_{30}\begin{vmatrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\theta(3\xi_i-2\xi)\theta(2-\xi)(-1)^{(\xi_i-\xi_{i-1})\xi}$$

$$\times\left[1-\sum_{k=1}^{3}\delta(l_{k4}+1)\delta_{ki}\right]\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m-1]_3^i\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle. \quad (3.18)$$

The relations (2.33)–(2.35) remain unaltered. The reason for this is that $\overline{W}([m]_4)$ is invariant with respect to e_{01}, e_{02}, and e_{03}. For completeness we also write these relations here as

$$e_{01}\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\theta(2\xi-3\xi_i)\theta(\xi-1)(-1)^{(\xi_i-\xi_{i+1})(\xi-1)}S(i,j)S(j,1)$$

$$\times(l_{i4}+1)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{i3})\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m+1]_3^{-i}\\ [m+1]_2^{-j}\\ m_{11}\end{matrix}\right\rangle, \quad (3.19)$$

$$e_{02}\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\theta(2\xi-3\xi_i)\theta(\xi-1)(-1)^{(\xi_i-\xi_{i+1})(\xi-1)}S(i,j)$$

$$\times(l_{i4}+1)\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m+1]_3^{-i}\\ [m+1]_2^{-j}\\ m_{11}+1\end{matrix}\right\rangle, \quad (3.20)$$

$$e_{03}\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\theta(2\xi-3\xi_i)\theta(\xi-1)(-1)^{(\xi_i-\xi_{i+1})(\xi-1)}(l_{i4}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{\Pi_{k\neq i=1}^{3}(l_{k4}-l_{i4})}\right|^{1/2}\left|\begin{matrix}[m]_4\\ [m+1]_3^{-i}\\ [m+1]_2\\ m_{11}+1\end{matrix}\right\rangle. \quad (3.21)$$

The relations (3.16)–(3.21) describe all irreducible finite-dimensional representations of the LS sl(1,3). We have not written the transformations corresponding to the even generators here. They easily can be obtained from the anticommutators

$$E_{ij}=\{e_{i0},e_{0j}\},\quad i,j=1,2,3, \quad (3.22)$$

and have been given in Ref. 1, Eq. (3.22).

Observe that the finite-dimensional irreducible representations of sl(1,3) are enumerated by all complex triplets $[m_{14},m_{24},m_{34}]$, such that

$$m_{14}-m_{24},m_{24}-m_{34}\in\mathbb{Z}_+. \quad (3.23)$$

In general, two different triplets $[m]_4\neq[m']_4$ describe inequivalent representations.

Since the Cartan subalgebra of sl(1,3), namely,

$$H=\text{lin env}\{E_{11},E_{22},E_{33}\}, \quad (3.24)$$

is the Cartan subalgebra of the even part gl(3), it is clear that each GZ pattern is an eigenvector of H, i.e., the GZ basis consists of weight vectors. The highest weight vector x_Λ is

$$x_\Lambda=\left|\begin{matrix}m_{14},m_{24},m_{34}\\ m_{14},m_{24},m_{34}\\ m_{14},m_{24}\\ m_{14}\end{matrix}\right\rangle. \quad (3.25)$$

The numbers m_{14},m_{24},m_{34} are the coordinates of the highest weight Λ in the dual to E_{11},E_{22},E_{33} basis, i.e.,

$$\Lambda(E_{ii})=m_{i4},\quad i=1,2,3, \quad (3.26)$$

and, therefore

$$E_{ii}x_\Lambda=m_{i4}x_\Lambda. \quad (3.27)$$

The interpretation of any GZ pattern

$$x=\left|\begin{matrix}[m]_4\\ [m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \quad (3.28)$$

is the same as in the case of Lie algebras. It indicates that x belongs (and is defined up to a multiple) to the nondecreasing chain (the flag)

$$x \in V(m_{11}) \subset V([m]_2) \subset V([m]_3) \subset \overline{W}([m]_4) \qquad (3.29)$$

of irreducible $gl(1) \subset gl(2) \subset gl(3) \subset sl(1,3)$ modules, correspondingly.

IV. CONCLUDING REMARKS

In Refs. 1 and 2 and in the present paper all finite-dimensional irreducible representations of the basic Lie superalgebra $sl(1,3)$ have been constructed. In Ref. 2 we needed 62 relations in order to turn the linear spaces $\overline{W}([m]_4)$ into irreducible modules over $sl(1,3)$. There we treated the typical representations separately and also treated each one from the three classes of nontypical representations separately. Here, on the ground of appropriate notations for the GZ basis, we succeeded in expressing the transformation properties of the basis in terms of six relations (3.16)–(3.21) simultaneously for the typical and the nontypical cases. One can go even farther and unify (3.16)–(3.21) in only two expressions (one for e_{k0} and one for e_{0k}, $k = 1,2,3$) but this is not of great advantage. To our mind, it is more important that the form of the GZ patterns and the expressions (3.16)–(3.21) clearly indicate the direction in which one can try to generalize the results for the Lie superalgebras $sl(1,n)$ and $sl(m,n)$.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for the kind hospitality at the International Centre for Theoretical Physics, Trieste.

He is also thankful to Professor P. Budinich and to the International School for Advanced Studies for financial support.

[1] T. D. Palev, J. Math. Phys. **26**, 1640 (1985).

[2] T. D. Palev, "Finite-dimensional representations of the Lie superalgebra sl(1,3) in a Gel'fand basis. II. Nontypical representations," preprint INRNE, Sofia, 1985.

[3] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).

[4] V. G. Kac, Lect. Notes Math. **626**, 597 (1978).

[5] V. G. Kac, Adv. Math. **26**, 8 (1977).

The representation of Lie groups by bundle maps

R. N. Sen
Department of Mathematics and Computer Science, Ben Gurion University of the Negev, 84105 Beer Sheva, Israel

(Received 15 January 1986; accepted for publication 18 March 1986)

This paper deals with representations of connected Lie groups by bundle maps of fiber bundles. It is pointed out that a large class of such representations can be obtained from the bundle structure theroem, and explicit constructions are given, first on principal bundles and then on associated bundles. Examples are provided to show that, for bundle representations, the theorem of full reducibility breaks down even for compact Lie groups. Finally, a general construction is given for obtaining representations of a Lie group on an arbitrary principal bundle. However, it is not known whether this exhausts all possibilities.

I. INTRODUCTION

In physics, one commonly considers group actions on two kinds of spaces: (1) linear spaces (which gives rise to the notion of *linear representations* of groups), and (2) topological spaces, usually with the additional structure of a manifold or a metric space. In view of the significance that vector bundles have assumed, it would seem reasonable to study *group representations on vector bundles*. Group representations on Hilbert bundles—which are generally equivalent to the product—have already been studied in some detail and found useful in physical applications.[1,2] Moreover, the inducing construction in infinite-dimensional group representations may be considered as based upon group representations on Hilbert bundles.[3] However, to the best of our knowledge, there has been no discussion of group actions on vector bundles that are *not necessarily equivalent* to the product. In the present article we shall endeavor to make a beginning in this direction.

In the theory of fiber bundles, the notion of *principal bundles* has come to occupy a pivotal role. Every other bundle is *associated* with a principal bundle in a well-defined manner. Therefore, if we are able to define group actions on principal bundles, we should also be able to transfer this action to associated bundles without much fuss. This is indeed the case. Now, among principal bundles, there is a very important class given by the *bundle structure theorem*; and (although it has seldom been noticed) *the theorem also proves* that left translations provide a representation of the group under consideration! We are therefore able to obtain a great deal of information from a detailed study of this theorem and its consequences. Armed with this information, it becomes relatively easy to attack the problem of representing any given group by bundle maps of a given bundle.

In our case, as elsewhere, the categories of topological spaces and topological groups are too broad to work with, and some restrictions would be required. We shall therefore make the following assumptions in the main part of this paper.

(1) The groups that we are trying to represent are connected Lie groups.

(2) The base spaces of our bundles are connected, paracompact, and Hausdorff, and the structure groups are Lie groups.

It is possible to relax these assumptions somewhat, without too much effort.[4]

Finally, in keeping with the practice in linear representation theory, we shall represent only the algebraic and topological structures of the group, and forget about the differentiable structure. Therefore we consider bundles *only in the topological category*, and disregard the geometry.

We shall be working with a fiber bundle with a structure group, according to the definition given by Steenrod.[4] This definition is more useful for our purposes than the equivalent coordinate-free definition of Ehresmann,[5] which is used, for instance, in the standard work of Kobayashi and Nomizu,[6] as well as in the majority of works on physical applications that have appeared in recent years. We assume that the reader is familiar with this definition, as well as the basic results in fiber bundle theory. A summary of these is given in Appendix A. Details may be found in the book of Steenrod.[4] Our terminology is either standard or self-evident. Our notations follow Steenrod's book, with a few departures. These are also explained in Appendix A.

The plan of this paper is as follows. In Sec. II we give an exhaustive discussion of bundle representations that may be obtained from the bundle structure theorem. In the last subsection we show, by means of examples, that the theorem of full reducibility fails for bundle representations even for compact Lie groups. This, in our opinion, is the feature that may hold promise of physical applications for such representations. Finally, in Sec. III we give a fairly general construction of bundle representations on arbitrary principal bundles. However, it has not been proved that all representations can be obtained by this method.

II. BUNDLE REPRESENTATIONS FROM THE BUNDLE STRUCTURE THEOREM

In this section we shall first define group representations on fiber bundles, or *bundle representations*[7] in the general case. Then we shall state the bundle structure theorem of Whitney and Steenrod.[8] It will become obvious at once that a large class of bundle representations of Lie groups is furnished by the bundle structure theorem. We shall work out

 0022-2488/86/082002-07$02.50

the details and obtain some explicit formulas in the rest of the section.

A. Definitions

Let $\{B,X,p,G,G\}$ be a principal bundle such that X is paracompact and Hausdorff and G is Lie, and let A be a Lie group.

Definition 1: A representation of A upon the bundle B is a continuous homomorphism

$$h\colon A\to\mathscr{A}(B) \tag{1}$$

of A into a topological group $\mathscr{A}(B)$ of bundle automorphisms of B. For computational purposes, it is convenient to replace the above by the following equivalent, but more explicit definition.

Definition 2: A representation A upon the bundle B is a map

$$H\colon A\times B\to B \tag{2}$$

with the following properties.

(i) H is continuous.

(ii) H preserves the fibers, i.e., there exists a map $\bar{H}\colon A\times X\to X$ such that

$$\bar{H}\pi=pH.$$

Here $\pi\colon A\times B\to A\times X$ is the obvious projection, defined by

$$\pi(a,b)=(a,p(b)),$$

where $a\in A$, $b\in B$, and $p(b)\in X$.

(iii) For each $a\in A$, the map $h(a)\colon B\to B$, defined by

$$h(a)b=H(a,b),$$

is a bundle map.

(iv) The collection of bundle maps $\{h(a)\}$, $a\in A$, has the standard representation properties, which follow from the associativity of multiplication in the group:

$$h(a)h(a')b=h(aa')b,\quad \forall a,a'\in A,\quad b\in B;$$

$$h(e)b=b,\quad \forall b\in B.$$

Here e is the identity in A.

It follows from the above that the bundle maps $h(a)$ are invertible and that

$$h(a)^{-1}=h(a^{-1}).$$

Note also the following, which is easily verified:

$$\bar{H}(a,x)=\bar{h}(a)x,\quad \forall a\in A,\quad x\in X,$$

where $\bar{h}(a)$ is defined by

$$\bar{h}(a)p=ph(a).$$

B. The bundle structure theorem

We state below a simplified form of the bundle structure theorem. This form applies to a group B and a closed subgroup G of B such that G does not have any proper normal subgroups. Here B/G denotes the space of left cosets with the quotient topology. The reader is referred to Steenrod[4] for the statement, proof, and discussion of the complete theorem.

Theorem: (a) Let B be a topological group and G be a closed subgroup that admits a local cross section in B, and let $p\colon B\to B/G$ be the natural projection. Then (the space) B is a principal bundle with total space B, base B/G, projection p, and group and fiber G, the group acting upon the fiber by left translations.

(b) The left translations of the group B (by elements of B) are bundle maps of this bundle upon itself.

Proof: See Appendix B.

Now let $H\colon B\times B\to B$ be the map defined by the left translations. Then conditions (ii) and (iii) of Definition 2 are proven in part II of the bundle structure theorem. Condition (iv) expresses the group property of left translations. There remains condition (i), or the continuity of H. *But this is exactly the continuity of multiplication in the topological group B*! We have therefore established the following theorem.

Theorem: Let G be a closed subgroup of the Lie group B. The left translations of B provide a representation of B on the principal bundle $\{B,B/G,p,G,G\}$.

The fact that B is a Lie group is used to guarantee the existence of a local cross section of G in B. This is a fundamental theorem of Chevalley. For a proof, see Varadarajan.[9]

C. An explicit formula for left translations

We shall obtain an explicit formula for left translations as bundle maps.

Let b denote a point in the bundle B, with $p(b)=x_0$. Let a_1 and a_2 denote elements of the group B that act upon the bundle B by left translations. Set $a_1x_0=x_1$ and $a_2x_1=x_2$. Finally, denote by η_c a local cross section over V_c, which is some coordinate neighborhood containing the point x_0. As in the proof of the bundle structure theorem, define the coordinate function $\phi_c\colon V_c\times G\to p^{-1}(V_c)$ by the equation

$$b=\phi_c(x_0,g)=\eta_c(x_0)g. \tag{3}$$

Assume now that $x_1=a_1x_0\in V_c$. We want an explicit formula for a_1b,

$$a_1b=\phi_c(x',g'),$$

i.e., we wish to determine x' and g'. Since $p\phi_c(x',g')=x'$ and $p(a,b)=x_1$, we have $x'=x_1$. To calculate g', proceed as follows:

$$a_1b=\phi_c(x_1,g')=\eta_c(x_1)g'$$

or

$$g'=\eta_c(x_1)^{-1}a_1b.$$

Substitute for b from Eq. (3) to obtain

$$g'=\eta_c(a_1x_0)^{-1}a_1\eta_c(x_0)g.$$

The factor that multiplies g on the right is familiar to physicists as a "Wigner rotation" or a "Mackey cocycle." Finally,

$$a_1b=\phi_c(x_1,\eta_c(x_1)^{-1}a_1\eta_c(x_0)g). \tag{4}$$

Formula (4) solves the problem when x_0 and x_1 lie in the same coordinate neighborhood. When the initial and final points on the base space do *not* lie in the same coordinate neighborhood, we have to "glue together" several pieces to obtain the final result. So let V_d be a coordinate neighborhood that contains x_2 and x_1, but not x_0, and let us try to obtain an explicit formula for a_2a_1b, where $p(a_2a_1b)=x_2$. We first write the point a_1b in the coordinates of $p^{-1}(V_d)$.

The formula is

$$a_1 b = \phi_d(x_1, g_{dc}(x_1) g'),$$

where g_{dc}: $V_c \cap V_d \to G$ is given by

$$g_{dc}(x_1) = \eta_d(x_1)^{-1} \eta_c(x_1).$$

Here η_d is the local cross section over V_d and ϕ_d is the coordinate function over $V_d \times G$. These functions are defined exactly as in Steenrod's proof of the bundle structure theorem. For the convenience of the reader who is not familiar with this proof, a summary of the notations is given in Appendix B.

Using the explicit expression for $g_{dc}(x_1)$, we may rewrite the formula for $a_1 b$ as

$$a_1 b = \phi_d(x_1, \eta_d(x_1)^{-1} a_1 \eta_c(x_0) g).$$

Now, by exactly the same arguments which led to Eq. (4), we obtain

$$a_2 a_1 b = \phi_d(x_2, \eta_d(x_2)^{-1} a_2 \eta_d(x_1) \cdot \eta_d(x_1)^{-1} a_1 \eta_c(x_0) g)$$
$$= \phi_d(x_2, \eta_d(x_2)^{-1} a_2 a_1 \eta_c(x_0) g). \quad (5)$$

Observe the cancellation of $\eta_d(x_1) \cdot \eta_d(x_1)^{-1}$ in the middle!

Finally, let a be any element of A, and x any point in B/G. The points x and ax may not lie in any one coordinate neighborhood on B/G, but must surely lie in the union of a finite number of such coordinate neighborhoods.[10] The following formula may easily be obtained by iterating the steps leading to formulas (4) and (5). Here

$$ab \equiv h(a)b = \phi'(ax, \eta'(ax)^{-1} a \eta(x) g), \quad (6a)$$

where g is defined by

$$b = \phi(x, g), \quad (6b)$$

V and V' are coordinate neighborhoods containing the points x and ax, respectively, η and η' the local cross sections over V and V', respectively, and ϕ and ϕ' the corresponding coordinate neighborhoods; they need not be the special choices used in proving the bundle structure theorem. This last statement may be proved by a little straightforward computation.

D. Group representations on associated bundles

Let $\{E, X, \pi, G, Y\}$ be a bundle associated with the principal bundle $\{B, X, p, G, G\}$. Denote by $\{V_j\}$, $j \in J$, a family of coordinate neighborhoods on X, and by $\{\theta_j\}$, $j \in J$, the corresponding local cross sections of the bundle E:

$$\theta_j: \ V_j \to E,$$
$$\pi \theta_j(x) = x, \quad \forall x \in V_j.$$

Furthermore, let ξ_j: $V_j \times Y \to \pi^{-1}(V_j)$ be the local trivialization of $\pi^{-1}(V_j)$. Then the coordinate transformations g_{ji}: $V_i \cap V_j \to G$ are given by

$$g_{ji}(x) = \xi_{j,x}^{-1} \xi_{i,x},$$

and these coincide with the corresponding functions in B. Finally, denote by D a continuous and effective action of G on Y, i.e., the image of $y \in Y$ under $g \in G$ is given by $D(g)y$. Then, if $z \in E$, the analogs of formulas (6) are the following:

$$z = \xi_i(x, y), \quad (7a)$$
$$az = \xi_j(ax, D\{\theta_j(ax)^{-1} a \theta_i(x)\} y). \quad (7b)$$

Here $z \in \pi^{-1}(V_i)$ and $az \in \pi^{-1}(V_j)$. The only new element is the representation D of G on Y. The choice of this representation is arbitrary.

E. The effect of reducing the group of the bundle

Let us briefly recall what is meant by *reducing the group of the bundle*. Let $\{E, X, \pi, G, Y; (V_\sigma, \gamma_\sigma), \sigma \in \Sigma\}$ be a coordinate bundle. The coordinate transformations of this bundle are given by

$$g_{\sigma\rho}(x) = \phi_{\sigma,x}^{-1} \phi_{\rho,x}, \quad x \in V_\sigma \cap V_\rho,$$

and $g_{\sigma\rho}(x) \in g$. Now suppose that there exists another bundle atlas $\{(V_j, \phi_j)\}$, $j \in J$, compatible with the atlas $\{(V_\sigma, \phi_\sigma)\}$, $\sigma \in \Sigma$, but with the additional property that

$$g_{ji}(x) = \phi_{j,x}^{-1} \phi_{i,x} \in K, \quad x \in V_i \cap V_j, \quad (8)$$

where K is a proper closed subgroup of G. We may therefore consider the equivalence class of all J-atlases on E. These define a bundle

$$\{E, X, \pi, K, Y\},$$

the structure group of which is a proper closed subgroup of the group G of the original bundle $\{E, X, \pi, G, Y\}$. This passage from $\{E, X, \pi, G, Y\}$ to $\{E, X, \pi, K, Y\}$ is known as "reducing the group of the bundle."

The possibility of reducing the group G of a bundle depends upon the topology of G. For example, if G is a connected Lie group, then it is homeomorphic[11] to $\mathbb{R}^n \times K$, where n is a positive integer and K is a maximal compact subgroup of G. The factor $\mathbb{R}^n$ is homotopically trivial, and may therefore be shrunk away[12] (if the base space of the bundle is paracompact), and the group of the bundle reduced to K.

We now return to the principal bundle $\{B, X, p, G, G\}$ of the Lie group B. Since G is closed in B, G is also a Lie group, and therefore the group of the bundle can be reduced to K, a maximal compact subgroup of G. This gives us the bundle $\{B, X, p, K, G\}$, which is no longer *a principal bundle*. However, we may construct the principal bundle $\{F, X, q, K, K\}$ associated with it. The total space F of this bundle *is no longer a group!*

Nevertheless, the representation (6) on the bundle B may be transferred to the bundle F. The details are as follows. We work with the open cover $\{V_j\}$, $j \in J$, of X, which effects the reduction of the group of the bundle B. Let

$$\omega j: \ V_j \to q^{-1}(V_j)$$

be local cross sections over V_j of the bundle F. Since F is a principal bundle, the group K acts freely on the fibers from the right.[13] We may therefore define the local trivializations ψ_j: $V_j \times K \to q^{-1}(V_j)$ as follows:

$$\psi_j(x, k) = w_j(x) \cdot k, \quad \forall x \in V_j, \quad k \in K.$$

Then the coordinate transformations may be written as

$$k_{ji}(x) = \psi_{j,x}^{-1} \psi_{i,x} = \omega_j(x)^{-1} \omega_i(x), \quad (9)$$

where $k_{ji}(x) \in K$. The fact that the bundle F is obtained after reducing the group of the bundle B means that the local cross sections ω_i may be so chosen that the k_{ji} of (9) agree with the g_{ji} of (8):

$$k_{ji}(x) = g_{ji}(x), \quad \forall x \in V_i \cap V_j.$$

The representation of the group A upon the bundle F may now be written down immediately by comparison with Eq. (6). Let $\beta \in F$, and denote by $a\beta$ the image of β under the action of $a \in A$. Then, if $q(\beta) \in V_i$, $q(a\beta) \in V_j$, we have

$$\beta = \psi_i(x,k), \tag{10a}$$

$$a\beta = \psi_j(ax, \omega_j(ax)^{-1} a \omega_i(x) k). \tag{10b}$$

Finally, if $\{C,X,\pi,K,Y\}$ is a bundle associated with the principal bundle $\{F,X,q,K,K\}$, then we may write the action of A on it as

$$\gamma = \zeta_i(x,k), \tag{11a}$$

$$a\gamma = \zeta_j(ax, \Delta\{\omega_j(ax)^{-1} a \omega_i(x)\} y). \tag{11b}$$

Here $\gamma \in C$, the $\{\zeta_j\}$, $j \in J$, are coordinate functions for C, $y \in Y$, and $\{D(k)\}$ is the representation of K on Y.

F. The effect of enlarging the group of the bundle

If K is a closed subgroup of G, then a coordinate transformation on X with values in K is trivially one with values in G. Suppose now that the action of K on Y may be extended continuously, and effectively, to an action of G on Y, and suppose that we are given a coordinate bundle $\{E,X,\pi,K,Y;$ $(V_j,\phi_j), j \in J\}$. We may then enlarge the J-atlas (V_j,ϕ_j) to another, call it a Σ-atlas (V_σ,ϕ_σ), in which the coordinate transformations take their values in G. We thus obtain the coordinate bundle $\{E,X,\pi,G,Y;\ (V_\sigma,\phi_\sigma), \sigma \in \Sigma\}$. Passing to equivalence classes, we obtain the fiber bundle $\{E,X,\pi,G,Y\}$. This process is called *enlarging the group of the bundle* $\{E,X,\pi,K,Y\}$.

It is always possible to enlarge the group of a principal bundle to obtain a new principal bundle. Let G be a proper closed subgroup of the Lie group L, and consider the principal bundle $\{B,X,p,G,G\}$. The group G, being a subgroup of L, acts continuously and effectively upon L by left translations, and therefore one may obtain the bundle $\{C,X,r,G,L\}$ by the Steenrod construction. The associated principal bundle $\{C,X,r,L,L\}$ certainly exists. Correspondingly, the action of the group B upon the bundle $\{B,X,p,G,G\}$ by left translations may always be transferred to the bundle $\{C,X,r,L,L\}$. The steps are as follows.

Since G is a proper subgroup of L, there exists a natural injection of $V_b \times G$ into $V_b \times L$. Denote this by i. Next, let $\{\eta_b\}$, $\{\phi_b\}$, $b \in B$, be local cross sections and coordinate functions for the bundle B, given (say) by the formulas (B2) and (B3) of Appendix B. Let ψ_b be a coordinate function on the bundle $\{C,X,r,G,L\}$. Then there exists a continuous fiber-preserving map

$$\sigma\colon\ B \to C,$$

the restriction of which to $p^{-1}(V_b)$ is given by the commutative diagram of Fig. 1. We use σ to define a local cross section ω_b over V_b in $\{C,X,r,G,L\}$ by requiring commutativity of the diagram of Fig. 2. Finally, to obtain the principal bundle $\{C,X,r,L,L\}$ from $\{C,X,r,G,L\}$ we have to pass from the B-atlas, $\{V_b,\psi_b\}$, $b \in B$, to an enlarged atlas $\{V_\gamma,\psi_\gamma\}$, $\gamma \in \Gamma$, in which the coordinate transformations take their values in L. The formulas for the action of B on the bundle $\{C,X,r,L,L\}$ are

$$b = \psi_\gamma(x,l), \tag{12a}$$

$$a \cdot b = \psi_\delta(ax, \omega_\delta(ax)^{-1} a \omega(x) \cdot l), \tag{12b}$$

where $x \in V_\gamma$, $ax \in V_\delta$, $\omega_\gamma\colon\ V_\gamma \to C$ is a local cross section, and ω_γ agrees with ω_c on the intersection $V_\gamma \cap V_c$.

Observe that the representation of B on a principal bundle given by formulas (12) cannot be obtained by applying the bundle structure theorem to B.

$$\begin{tikzcd} V_b \times G \arrow[r, "i"] \arrow[d, "\phi_b"'] & V_b \times L \arrow[d, "\psi_b"] \\ p^{-1}(V_b) \arrow[r, "\sigma"] & r^{-1}(V_b) \end{tikzcd}$$

FIG. 1.

$$\begin{tikzcd} & V_b \arrow[dl, "\eta_b"'] \arrow[dr, "\omega_b"] & \\ p^{-1}(V_b) \arrow[rr, "\sigma"] & & r^{-1}(V_b) \end{tikzcd}$$

FIG. 2.

G. Examples

We shall give a class of examples to illustrate phenomena that cannot occur in representations on bundles equivalent to the product.

Denote by O_n the real orthogonal group in n dimensions. Then

$$\mathrm{O}_n / \mathrm{O}_{n-1} = S^{n-1},$$

where S^n is the unit n-sphere $x_1^2 + \cdots + x_n^2 + x_{n+1}^2 = 1$. The group O_n acts transitively on S^{n-1}, and is a principal bundle over S^{n-1} with fiber and base O_{n-1}: $\{\mathrm{O}_n, S^{n-1}, p, \mathrm{O}_{n-1}, \mathrm{O}_{n-1}\}$.

The group of the tangent bundle TS^{n-1} of S^{n-1} is $\mathrm{GL}(n-1,\mathbb{R})$. However, as discussed in Sec. II E, the group may always be reduced to a maximal compact subgroup, which in the present case is O_{n-1}. That is, the bundle $\{TS^{n-1}, S^{n-1}, p, \mathrm{O}_{n-1}, \mathbb{R}^{n-1}\}$ is associated with $\{\mathrm{O}_n, S^{n-1}, p, \mathrm{O}_{n-1}, \mathrm{O}_{n-1}\}$.

The bundle structure theorem gives a representation of O_n upon $\{\mathrm{O}_n, S^{n-1}, p, \mathrm{O}_{n-1}, \mathrm{O}_{n-1}\}$. By the procedure of Sec. II D this gives rise to an action upon $\{TS^{n-1}, S^{n-1}, p, \mathrm{O}^{n-1}, \mathbb{R}^{n-1}\}$ for any choice of a linear representation of O_{n-1} upon $\mathbb{R}^{n-1}$ [cf. Eqs. (7a) and (7b)].

We shall obtain the examples to illustrate the phenomena we have in mind by letting O_{n-1} act *identically* upon

TABLE I. Vector fields on spheres.

n	2	3	4	5	6	7	8
$\rho(n)$	2	1	4	1	2	1	8
No. of vector fields on S^{n-1}	1	0	3	0	1	0	7

$\mathbb{R}^{n-1}$. However, we first need a digression.

The number of linearly independent vector fields[14] on S^{n-1} equals $\rho(n) - 1$, where $\rho(n)$ is the Hurwitz–Radon–Eckmann number, defined as follows. Write $n = (2a+1)2^b$, where n, a, and b are non-negative integers, let $b = c + 4d$, where c and d are non-negative integers, and $0 \leqslant c \leqslant 3$. Then

$$\rho(n) = 2^c + 8d.$$

For $n = 2$ to 8, the values of $\rho(n)$ and the number of independent vector fields on S^{n-1} are shown in Table I. That is, S^1, S^3, and S^7 admit one, three, and seven vector fields respectively; S^2, S^4, and S^6 admit none; and S^5 admits one.

A vector field on a manifold is a (nowhere zero) cross section of its tangent bundle. If an n-dimensional vector bundle has n everywhere linearly independent cross sections, then it is equivalent to the product[15] (its base space is *parallelizable*). In this case the entire bundle can be expressed as the Whitney sum of n one-dimensional subbundles. If, however, the bundle has only m independent cross sections, $m < n$, then one can identify only m one-dimensional subbundles. If these are split off, the rest is no longer a subbundle.

Now we may return to the bundle representations obtained by letting the group O_{n-1} act identically upon the fiber $\mathbb{R}^{n-1}$, $gx = x$, $\forall g \in O_{n-1}$, $x \in \mathbb{R}^{n-1}$. Then we have the following.

(i) $n = 2,4,8$. The spheres S^1, S^3, and S^7 are parallelizable, i.e., TS^1, TS^3, and TS^7 are equivalent to the product. The representation splits into a Whitney sum of exactly n irreducible subrepresentations.

(ii) $n = 3,5,7$. There are no vector fields at all on S^2, S^4, and S^6, and therefore *no irreducible one-dimensional subrepresentations whatsoever*.

(iii) $n = 6$. S^5 admits exactly one vector field. The five-dimensional representation admits only one irreducible one-dimensional subrepresentation.

Recall that we are dealing with representations of the *compact* Lie groups, which are the best-behaved groups of all. Phenomena analogous to (ii) and (iii) do not exist among linear representations, or bundle representations on product bundles, and are topological.

III. GROUP ACTIONS ON PRINCIPAL BUNDLES

Let $\{B,M,p,G,G\}$ be a principal bundle, $\{V_i\}_{i\in J}$ an open cover of M, and

$$\phi_i\colon\ V_i \times G \to p^{-1}(V_i), \quad i \in J,$$

the corresponding local trivializations. The corresponding coordinate transformations on M with values in G:

$$g_{ji}\colon\ V_i \cap V_j \to G, \quad V_i \cap V_j \neq \varnothing,$$

are

$$g_{ji}(x) = \phi_{j,x}^{-1} \circ \phi_{i,x}, \quad x \in V_i \cap V_j,$$

where the homeomorphism $\phi_{i,x}\colon\ G \to p^{-1}(x)$ is defined by

$$\phi_i(x,g) = \phi_{i,x}(g).$$

Finally, if $x \in V_i \cap V_j$, and

$$b = \phi_i(x,g), \tag{13}$$

then

$$b = \phi_j(x, g_{ji}(x)\cdot g). \tag{14}$$

Formulas (13) and (14) show how the local trivializations are glued together with the help of the group G to form the bundle.

Now suppose that the Lie group A has an action on B, say from the left. This action would give rise to an action of A on M. Under the latter, M would split into a collection of pairwise disjoint A-orbits. The bundle B itself would split into a collection of pairwise disjoint A-invariant subbundles, each of which has to be considered separately. Therefore, there is no loss of generality in assuming that M itself is an A-orbit. That is, there exists a closed subgroup K of A such that M and A/K are in one-to-one correspondence. There is no effective loss of generality in assuming that they are also topologically the same, i.e., in identifying M with A/K. Then A is a principal bundle over M with group and fiber K: $\{A,M,\pi,K,K\}$.

Thus the space M is furnished with *two* G-structures: one with group K, arising from A via the bundle structure theorem, and the other, with group G, being the one used to construct the bundle B.

Let us now consider the action of A on the bundle B. Let $x \in V_i$ and $a \in A$ such that also $ax \in V_i$. Finally, let $b \in B$, $g \in G$ such that

$$b = \phi_i(x,g).$$

Then

$$ab = \phi_i(ax, \delta_i(a,x)g). \tag{15}$$

The formal properties of the objects δ are fairly obvious! Let $a' \in A$ such that $a'ax \in V_i$. Then

$$\delta_i(e,x)g = g, \tag{16a}$$

$$\delta_i(a',ax)\delta_i(a,x)g = \delta_i(a'a,x)g. \tag{16b}$$

These are, of course, precisely the "cocycle conditions" acting on g. We know that if

$$\eta_i\colon\ V_i \to A$$

is a family of local cross-sections of the principal bundle $\{A,M,\pi,K,K\}$, then the cocycles

$$s_i(a,x) = \eta_i(ax)^{-1} a \eta_i(x) \tag{17}$$

satisfy the conditions

$$s_i(e,x) = e, \tag{18a}$$

$$s_i(a',ax)s_i(a,x) = s_i(a'a,x). \tag{18b}$$

Comparing (16) and (18), we see that a solution to our problem is provided by *any continuous homomorphism*

$$l\colon\ K \to G. \tag{19}$$

We simply set

$$\delta_i(a,x) = l(s_i(a,x)), \quad i\in J. \tag{20}$$

Thus the elementary bundle representation formula (15) becomes

$$ab = \phi_i(ax, l(s_i(a,x))g). \tag{21}$$

Next, suppose that $ax\in\phi_j$. Then

$$ab = \phi_j(ax, g_{ji}(x)\cdot l(s_i(a,x))\cdot g).$$

Finally, let $a'a\in V_j$. Then

$$a'ab = \phi_j(a'ax, (s_j(a',ax))g_{ji}(x)(s_i(a,x))g). \tag{22}$$

This process may be continued. However, unlike formula (5), no cancellation occurs in the middle, and the formula becomes longer at each step.

Thus, to determine an action of the group A upon the bundle B, the only new ingredient which is required is a continuous homomorphism l of K into G, where K is the subgroup of A such that A/K and M are homeomorphisms. Any such homomorphism determines one such action.

This completes our construction.

APPENDIX A: FIBER BUNDLES AND BUNDLE MAPS

The bundle with total space B, base space X, projection p (p: $B\to X$), group G, and fiber Y will be denoted by $\{B,X,p,G,Y\}$, or by B for brevity when no confusion is likely to result. The group G is topologized and is assumed to act continuously and effectively on Y. For definiteness, we shall assume that the group acts from the left, and write the image of $y\in Y$ under $g\in G$ as gy. A *bundle atlas* or coordinate system for the bundle B is an open cover $\{V_j\}, j\in J$, of X (here J is an indexing set), and for each V_j, a homeomorphism

$$\phi_j:\ V_j\times Y\to p^{-1}(V_j), \tag{A1a}$$

which satisfies

$$p\phi_j(x,y) = x, \quad \forall x\in V_j, \quad y\in Y. \tag{A1b}$$

The ϕ_j are also called *local trivializations*. Next, a map

$$\phi_{j,x}:\ Y\to p^{-1}(x), \quad x\in V_j, \tag{A2a}$$

is defined by setting

$$\phi_{j,x}(y) = \phi_j(x,y), \quad x\in V_j, \quad y\in Y. \tag{A2b}$$

Then, for each pair $i,j\in J$ and each $x\in V_i\cap V_j$, the map

$$\phi_{j,x}^{-1}\phi_{i,x}:\ Y\to Y, \quad x\in V_i\cap V_j,$$

is a homeomorphism. This homeomorphism is required to coincide with the action of an element of G upon Y, thus defining a continuous map

$$g_{ji}:\ V_i\cap V_j\to G, \tag{A3a}$$

where

$$g_{ji}(x) = \phi_{j,x}^{-1}\phi_{i,x}. \tag{A3b}$$

The maps g_{ji} are called, by Steenrod, *coordinate transformations on X with values in G*. They are also known as *transition maps* in the literature. The open cover $\{V_j\}, j\in J$, of X, together with a family of maps (A3a) and (A3b) constitutes a *G-structure* on X. The bundle B is uniquely defined by the base X, the G-structure upon it, and the fiber Y upon which G acts continuously and effectively. This fundamental theorem is sometimes known as the *Steenrod recognition principle*.

If the fiber Y is the same as the group G itself (considered as a topological space), the bundle is called a *principal bundle*. The principal bundle is sometimes regarded as the fundamental object in the class of all bundles with a given base and given *G-structure* upon the base.

Let $\{B,X,p,G,Y;\ (V_j,\phi_j), j\in j\}$ be a fiber bundle with a given coordinate system, in an obvious notation. A map

$$h:\ B\to B$$

is called a *bundle map* of B upon itself if it satisfies the following conditions.

(1) h is continuous.

(2) h *preserves fibers*, i.e., $p(b) = p(b')\Rightarrow ph(b) = ph(b')$, where $b,b'\in B$. Thus h induces a *base map*

$$\bar{h}:\ X\to X, \tag{A4a}$$

which satisfies

$$\bar{h}p = ph \tag{A4b}$$

(3) If $x\in V_j\cap\bar{h}^{-1}(V_k)$, and h_x: $p^{-1}(x)\to p^{-1}(x')$—where $x' = h(x)$—is the *fiber map* induced by h, then the map

$$\bar{g}_{kj}(x) = \phi_{k,x'}^{-1}h_x\phi_{j,x} \tag{A5a}$$

of Y into Y coincides with the action of an element G of Y.

(4) The map

$$\bar{g}_{kj}:\ V_j\cap\bar{h}^{-1}(V_k)\to G \tag{A5b}$$

so induced is continuous.

An invertible bundle map with a continuous inverse will be called a *bundle automorphism*.

APPENDIX B: THE BUNDLE STRUCTURE THEOREM

Theorem: (a) Let B be a topological group and G a closed subgroup that admits a local cross section in B, and let p: $B\to B/G$ be the natural projection. Then the space B is a principal bundle with total space B, base B/G, projection p, and group and fiber G, the group acting on the fiber by left translations.

(b) The left translations of the group B by elements of B itself are bundle maps of this bundle upon itself.

Proof (outline): (1) Choose the coordinate system as follows. The indexing set J is the set B itself. Let the local cross section f be defined over the open set $V\subset B/G$, f: $V\to B$, $pf(x) = x$, $\forall x\in V$. Define V_b to be the left translation of V by $b\in B$:

$$V_b = b\cdot V. \tag{B1}$$

(2) Define f_b: $V_b\to B$ by

$$f_b(x) = bf(b^{-1}x). \tag{B2}$$

Then f_b is continuous and $pf_b(x) = x$.

(3) Define the local trivializations ϕ_b: $V_b\times G\to p^{-1}(V_b)$ by

$$\phi_b(x,g) = f_b(x)\cdot g. \tag{B3}$$

Then ϕ_b is continuous, and $p\phi_b(x,g) = x$. Next, define

$$p_b:\ p^{-1}(V_b)\to G \tag{B4a}$$

by

$$p_b(z) = [f_b(p(z))]^{-1}z, \quad \forall z \in p^{-1}(V_b). \tag{B4b}$$

It follows quickly that

$$p_b \phi_b(x,g) = g, \tag{B5a}$$

$$\phi_b(p(z), p_b(z)) = z, \tag{B5b}$$

and that p_b is continuous. Therefore ϕ_b is a homeomorphism.

(4) Let $x \in V_b \cap V_c$. Then

$$p_c \phi_b(x,g) = f_c(x)^{-1}[f_b(x)g]$$
$$= [f_c(x)^{-1} f_b(x)]g.$$

Since

$$g_{cb}(x) = f_c(x)^{-1} f_b(x) \in G, \tag{B6}$$

it follows that $p_c \phi_b(x,g)$ is the left translation of g by the element $g_{cb}(x) \in G$. The continuity of $g_{cb}(x)$ follows from the continuity of f. This completes the construction of the coordinate bundle.

(5) It is proved easily that two different local cross sections give rise to strictly equivalent coordinate bundles. This completes the proof of part (a).

(6) To prove part (b), we have to verify properties (1)–(4) in the definition of bundle maps as given in Appendix A. Of these, (1) and (2) are immediate, and (3) and (4) are verified as follows. Let $x \in V_b$, $b_1 \in B$, and $x' = b_1 \in V_c$. Then the functions $\bar{g}_{cb}(x)$ of eqs. (A5) are given by

$$\bar{g}_{cb}(x) \cdot g = \phi_{c,x}^{-1}(b_1 \phi_{b,x}(g))$$
$$= f_c(x')^{-1} b_1 f_b f_b(x) g,$$

i.e.,

$$\bar{g}_{cb}(x) = f_c(b_1 x)^{-1} b_1 f_b(x), \tag{B7}$$

which will be recognized as a local "Wigner rotation" of B into G. Thus $\bar{g}_{cb}(x)$ takes its values in G, and is continuous in x for all $x \in V_b \cap V_c$. This completes the proof of part (b).

[1]H. J. Borchers and R. N. Sen, Commun. Math. Phys. **42**, 101 (1975).

[2]R. N. Sen, Physica (Utrecht) **94A**, 39, 55 (1978).

[3]The Hilbert space is the space of square-integrable Borel sections of the bundle with respect to an invariant or quasi-invariant measure on the base. See G. W. Mackey, *Unitary Group Representations in Physics, Probability, and Number Theory* (Oxford lectures) (Benjamin/Cummings, Reading, MA, 1978).

[4]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1956).

[5]C. Ehresmann, *Colloque international de topologie algébrique* (C.N.R.S., Paris, 1949), pp. 3–15.

[6]S.Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, (Interscience, New York, 1963), Vol. I.

[7]In earlier works,[1,2] the term *bundle representations* was used to denote representations on product bundles only. This was extended in an invariant manner, to representations on all bundles in R. N. Sen, J. Math. Phys. **22**, 290 (1981). The invariance of the earlier definition was proved in this work.

[8]N. Steenrod,[4] p. 30.

[9]V. S. Varadarajan, *Lie Groups, Lie Algebras and their Representations* (Prentice–Hall, Englewood Cliffs, NJ, 1974), p. 79.

[10]This is an immediate consequence of the fact that B/G is locally Euclidean.

[11]This is a theorem of E. Cartan. For a proof, see G. D. Mostow, Bull. Am. Math. Soc. **55**, 969 (1949).

[12]See Steenrod,[4] §12.

[13]See Kobayashi and Nomizu,[6] pp. 50ff.

[14]This is the famous "vector field problem" which was solved by J. F. Adams. See J. F. Adams, *Algebraic Topology: A Student's Guide* (Cambridge U.P., Cambridge, 1972), p. 239.

[15]See, for example, J. W. Milnor and J. D. Stasheff, *Characteristic Classes, Annals of Mathematics Studies*, No. 76 (Princeton U.P., Princeton, NJ, 1974), p. 18.

SU(m/n) weight systems and superprojection matrices

Chang-Ho Kim, Kee Yong Kim, W. S. I'Yi,[a] and Yongduk Kim
Physics Department, Sogang University, C.P.O. Box 1142, Seoul, Korea

Young-Jai Park
Physics Department, Kangwon National University, Chunchon, Korea

(Received 25 November 1985; accepted for publication 18 March 1986)

Representations of the SU(m/n) superalgebra are studied in terms of the Kac–Dynkin weight systems. Superprojection matrices are introduced for the possible branching patterns.

I. INTRODUCTION

There has been considerable interest in superalgebras[1] related to superunifications,[2] nuclear physics,[3] supergravities,[4]and superstring theories.[5] Especially, superalgebras allow more systematic analyses for the construction of supermultiplets in supergravities.[4,6]

Recently we have shown that the Dynkin weight technique and projection matrices are very useful for the exploration of grand unified theories.[7] In this spirit, we investigate representations of SU(m/n) superalgebras in the Kac–Dynkin basis.[6,8] Branching patterns of SU(m/n) (see Refs. 8 and 9) are elegantly analyzed introducing superprojection matrices. Our new superprojection matrices are extremely powerful for the construction of supergravities.

In Sec. II we summarize the algebraic structure of SU(m/n). In Sec. III we present an easy method of generating the full weight system of the SU(m/n) superalgebra. In Sec. IV, explicit constructions of weight systems are presented with some examples of SU(2/3) and SU(2/2). Branching rules and superprojection matrices are shown in Sec. V.

II. ALGEBRAIC STRUCTURE OF SU(m/n)

SU(m/n) is classified as a classical superalgebra of the type $A(m-1,n-1)$ according to Kac's convention.[1,8] SU(m/n) ($m \neq n$) consists of the even(bosonic) part, the subalgebra $SU(m) \otimes SU(n) \otimes U(1)$, and the odd(fermionic) part, which transforms as $[(\underset{\sim}{\overline{m}},\underset{\sim}{n}) \oplus (\underset{\sim}{m},\underset{\sim}{\overline{n}})]$ representations of the even part. Thus generators of SU(m/n) are

$$(\underbrace{m^2-1}_{\sim},\underset{\sim}{1}) + (\underset{\sim}{1},\underbrace{n^2-1}_{\sim}) + (\underset{\sim}{1},\underset{\sim}{1}) + (\underset{\sim}{m},\underset{\sim}{\overline{n}})_F + (\underset{\sim}{\overline{m}},\underset{\sim}{n})_F, \tag{2.1}$$

where the subscript F indicates the odd part. For $m = n$, the even part of SU(n/n) is $SU(n) \otimes SU(n)$ without the U(1) factor,

$$(\underbrace{n^2-1}_{\sim},\underset{\sim}{1}) + (\underset{\sim}{1},\underbrace{n^2-1}_{\sim}) + (\underset{\sim}{n},\underset{\sim}{\overline{n}})_F + (\underset{\sim}{\overline{n}},\underset{\sim}{n})_F. \tag{2.2}$$

The Cartan subalgebra consists of mutually commuting generators H_i ($i = 1,\ldots,m+n-1$).

Let $\Delta = \{\alpha_1,\ldots,\alpha_{m+n-1}\}$ be a fixed set of simple roots of SU(m/n) and let C be the graded Cartan matrix associated with Δ. Simple roots α_i^+,α_i^- correspond to raising and lowering operators E_i^+,E_i^-, respectively.

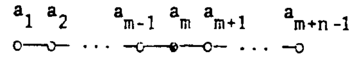
We use the conventions introduced by Kac[1,8] (Chevalley basis)

$$[H_i,E_j^{\pm}] = \pm (C)_{ij} E_j^{\pm}, \quad i,j = 1,\ldots,m+n-1, \tag{2.3}$$

where the C_{ij} are the ij elements of the graded Cartan matrix of SU(m/n):

$$C = \left[\begin{array}{ccccc:c:cccc}
\multicolumn{5}{c}{\overbrace{\hphantom{2\ -1\ 0\ \cdots\ \cdots}}^{m-1}} & & & & & \\
2 & -1 & 0 & \cdots & & & & & & \\
-1 & 2 & -1 & \cdots & & & & & & \\
 & & \ddots & \ddots & & & & 0 & & \\
 & & & \ddots & -1 & & & & & \\
 & & & -1 & 2 & -1 & 0 & & & \\
\hdashline
 & & & 0 & -1 & 0 & 1 & 0 & & \\
\hdashline
 & & & & 0 & -1 & 2 & -1 & & \\
 & & & & & 0 & -1 & \ddots & & \\
 & 0 & & & & & & & \ddots & \\
 & & & & & & & & & 2 \\
 & & & & & & \multicolumn{4}{c}{\underbrace{\hphantom{2\ -1\ \cdots\ 2}}_{n-1}}
\end{array}\right]. \tag{2.4}$$

Note that $C_{mm} = 0$, $C_{m,m-1} = -1$, and $C_{m,m+1} = 1$. As Lemire and Patera[10] pointed out, the graded Cartan matrix for a simple superalgebra depends on the choice of simple roots in the root system. Throughout this paper, we adopt the special sets of simple roots established by Kac.[1] The following Kac–Dynkin diagram contains the same information as the graded Cartan matrix (2.4):



The set $(a_1,a_2,\ldots,a_{m+n-1})$ characterizes the highest weight Λ of a representation; the a_i's are non-negative integers for $i \neq m$ and a_m can be any complex number.[8]

III. CONSTRUCTIONS OF SU(m/n) IRREDUCIBLE REPRESENTATIONS

Classifying finite-dimensional irreducible representations of the simple Lie superalgebras, Kac[1] pointed out the existence of "typical" and "atypical" irreducible representations.

The atypicality conditions[1,8] are expressed as

$$a_m = \sum_{k=m+1}^{j} a_k - \sum_{k=i}^{m-1} a_k - 2m + i + j,$$
$$1 \leqslant i \leqslant m, \quad m \leqslant j \leqslant m+n-1. \tag{3.1}$$

If a_m does not satisfy the condition (3.1), the representation Λ is called typical and otherwise atypical. In other words, the typical cases are direct summands in any representation

[a] Permanent address: Physics Department, Chungbuk National University, Cheongju, Korea.

 0022-2488/86/082009-07$02.50

in which they appear.[8,10] The atypical representations exist due to the fact that finite-dimensional representations of simple Lie superalgebras are not completely reducible in general.[10]

According to the Ramond's classification,[11] "typical" means that a representation consists of Bose and Fermi systems of the same dimensions (i.e., same degrees of freedom). Ramond's typical representations are important in supergravity because Bose and Fermi systems should have the same degrees of freedom.

One can obtain the full weight systems associated with the highest weight Λ by applying lowering operators; under the action of even generators the weight system Λ moves inside multiplets of $SU(m)\otimes SU(n)$ and the odd generators change a weight system into other systems with different spins.[8,11] The odd generators[8] $\beta_j^{i\pm}$ $(i=1,\ldots,m,\ j=m,\ldots,m+n-1)$ reside in $(\bar{\underset{\sim}{m}},\underset{\sim}{n})$ and $(\underset{\sim}{m},\bar{\underset{\sim}{n}})$:

$$\beta_j^{i+}\in(\underset{\sim}{m},\bar{\underset{\sim}{n}}),\quad \beta_j^{i-}\in(\bar{\underset{\sim}{m}},\underset{\sim}{n}). \tag{3.2}$$

Let D be a dimension of a typical representation of $SU(m/n)$, and d the dimension of the weight system of $SU(m)\otimes SU(n)$ related to the highest weight Λ. Then

$$D=d\cdot 2^{mn}. \tag{3.3}$$

This equation implies the dimensions of bosons and fermions are equally

$$D/2=d\cdot 2^{mn-1}. \tag{3.4}$$

There are $m\cdot n$ β_j^{i-}'s and the dimensions associated with the antisymmetric combinations of the β_j^{i-}'s coincide with the binomial coefficients of $(1+1)^{mn}$. Actually there exist only the antisymmetric combinations of negative odd generators,[8] since

$$\{\beta_j^{i-},\beta_l^{k-}\}=0,\quad \forall i,j,k,l. \tag{3.5}$$

By acting β_j^{i-}'s on a state $|\lambda\rangle$, the transformed state alternates between bosonic and fermionic states,[11] i.e.,

$$\begin{aligned}\beta_j^{i-}|\lambda\rangle_{\rm Bose}&=|\lambda'\rangle_{\rm Fermi},\\ \beta_l^{k-}|\lambda\rangle_{\rm Fermi}&=|\lambda'\rangle_{\rm Bose}.\end{aligned} \tag{3.6}$$

According to Hurni and Morel,[8] the operation of $\beta_j^{i\pm}$ causes a floor changing, i.e., from the ground floor (GND) to the mnth floor. For the highest weight Λ, the full weight system is

floor	weight system
GND	Λ
1st	$\Lambda\otimes(\underset{\sim}{m},\underset{\sim}{n})$
2nd	$\Lambda\otimes[(\underset{\sim}{m},\underset{\sim}{n})\otimes(\underset{\sim}{m},\underset{\sim}{n})]_A$
$\vdots$	$\vdots$
mnth	$\Lambda\otimes$ [tensor products of $m\cdot n$ times of $(\underset{\sim}{m},\underset{\sim}{n})]_A=\Lambda$,

(3.7)

where the subscript A means antisymmetrization. In general, the antisymmetrization of $[(\underset{\sim}{m},\underset{\sim}{n})\otimes(\underset{\sim}{m},\underset{\sim}{n})\otimes\cdots]$ is difficult to perform. This problem is easily solved using pertinent projection matrices (see Table I).

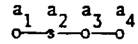
TABLE I. Generalized projection matrix $P(SU(mn)\to SU(m)\otimes SU(n))$ for searching the multiplets of $[\underset{\sim}{m},\underset{\sim}{n})\otimes(\underset{\sim}{m},\underset{\sim}{n})\otimes\cdots]_A$. If one wants the antisymmetrized product of k times $(\underset{\sim}{m},\underset{\sim}{n})$, one acts P on the right of the weight system $(0\cdots 0\,1\,0\cdots 0)$ of $SU(mn)$, where only the kth element is 1 (see Refs. 7 and 12).

$$P=\left(\begin{array}{cccc:cccc}
 & & & 1 & 1 & 0 & 0 & 0\\
 & 0 & 1 & & 2 & \cdot & \cdot & \\
 & \cdot & & & 3 & \cdot & \cdot & 0\\
\cdot & & & 0 & \cdot & \cdot & \cdot & \\
1 & & & & (m-1) & & & \\ \hdashline
0 & 0 & \cdots & 0 & m & 0 & \cdots & \\ \hdashline
 & & & 1 & (m-1) & 1 & & \\
 & 0 & \cdot & & \cdot & 2 & & 0\\
 & \cdot & & & \cdot & \cdot & & \\
\cdot & & & 0 & \cdot & \cdot & & \\
1 & & & & 1 & \cdot & & \\ \hdashline
0 & 0 & \cdots & 0 & 0 & m & 0 & \cdots\\ \hdashline
 & & & & 0 & (m-1) & 1 & \\
\text{Repeated} & & & & \cdot & \cdot & 2 & \\
 & & & & \cdot & \cdot & \cdot & \\
 & & & & \cdot & \cdot & \cdot & \\
 & & & & \cdot & 2 & \cdot & 0\\
 & & & & \cdot & 1 & (m-1) & \\
 & & & & \cdot & 0 & m & \\
 & & & & \cdot & \cdot & (m-1) & 1\\
 & & & & \cdot & \cdot & \cdot & \cdot\\
 & & & & \cdot & \cdot & \cdot & \cdot\\
 & & & & 0 & 0 & \cdot & \cdot
\end{array}\right)$$

m-1 (columns of left block), n-1 (columns of right block), m-1 (rows of first block), mn-1 (rows)

IV. EXPLICIT CONSTRUCTIONS OF WEIGHT SYSTEMS

We explicitly construct the full weight system for $SU(2/3)$ and $SU(2/2)$ as follows.

Example 1: $SU(2/3)$: 

. The graded Cartan matrix is

$$\left(\begin{array}{c:c:cc} 2 & -1 & 0 & 0\\ \hdashline -1 & 0 & 1 & 0\\ \hdashline 0 & -1 & 2 & -1\\ 0 & 0 & -1 & 2\end{array}\right). \tag{4.1}$$

$SU(2/3)$ has 12 even roots and 12 odd ones. Simple even roots corresponding to the ith column of the graded Cartan matrix[7] (4.1) are denoted by $\alpha_i^\pm$ $(i\neq m)$, and the second column of (4.1) gives simple odd roots $\beta_2^{2\pm}$. Then

$$\begin{aligned}&\alpha_1^\pm=(\pm 2\ \mp 1\ \ 0\ \ 0),\quad \alpha_3^\pm=(0\ \ \pm 1\ \pm 2\ \mp 1),\\ &\alpha_4^\pm=(0\ \ 0\ \ \mp 1\ \pm 2),\ \beta_2^{2\pm}=(\mp 1\ \ 0\ \ \mp 1\ \ 0).\end{aligned} \tag{4.2}$$

Note that β_2^{2+} corresponds to the lowest weight of $(\underset{\sim}{2},\bar{\underset{\sim}{3}})$ and β_2^{2-} to the highest weight of $(\underset{\sim}{2},\underset{\sim}{3})$,

$$(\underset{\sim}{2},\bar{\underset{\sim}{3}})=(1,0\,1),\quad(\underset{\sim}{2},\underset{\sim}{3})=(1,1\,0). \tag{4.3}$$

Other odd roots are

$$\begin{aligned}\beta_2^{1-}&=[\alpha_1^-,\beta_2^{2-}]=(-1\ \ 1\ \ 1\ \ 0),\\ \beta_3^{2-}&=[\beta_2^{2-},\alpha_3^-]=(\ \ 1\ -1\ -1\ \ 1),\\ \beta_3^{1-}&=[\beta_2^{1-},\alpha_3^-]=(-1\ \ 0\ -1\ \ 1),\\ \beta_4^{1-}&=[\beta_3^{1-},\alpha_4^-]=(-1\ \ 0\ \ 0\ -1),\\ \beta_4^{2-}&=[\beta_3^{2-},\alpha_4^-]=(\ \ 1\ -1\ \ 0\ -1).\end{aligned} \tag{4.4}$$

All the β_j^{i-}'s belong to the $(\underset{\sim}{2},\underset{\sim}{3})$ weight system.

From the highest weight $\underset{\sim}{\Lambda}$, the whole weight system is generated by $\Lambda \otimes [(1,1\,0) \otimes \cdots]_A$. To carry out antisymmetrization we embed the $(1,1\,0)$ weight system to the fundamental representation of SU(6). The projection matrix $P(\mathrm{SU}(6) \rightarrow \mathrm{SU}(2) \otimes \mathrm{SU}(3))$ is obtained from Table I:

$$\begin{array}{c} \qquad \mathrm{SU}(2) \qquad\quad \mathrm{SU}(3) \\ P = \left(\begin{array}{c:cc} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \\ 1 & 0 & 1 \end{array}\right). \end{array} \tag{4.5}$$

For instance, the contents of $[(\underset{\sim}{2},\underset{\sim}{3}) \otimes (\underset{\sim}{2},\underset{\sim}{3})]_A$, or equivalently $[(1,1\,0) \otimes (1,1\,0)]_A$, are those of $\underset{\sim}{15}$ when SU(6) reduces to $\mathrm{SU}(2) \otimes \mathrm{SU}(3)$. The desired results are derived by acting $P(\mathrm{SU}(6) \rightarrow \mathrm{SU}(2) \otimes \mathrm{SU}(3))$ to the right of the $(0\,1\,0\,0\,0)$ weight system,

$$\boxed{}\!\!\!\!\boxed{} = \underset{\sim}{15} = (0\,1\,0\,0\,0). \tag{4.6}$$

Also $[3 \text{ times}]_A$ coincides with $(0\,0\,1\,0\,0) \cdot P$, and so on. The results are summarized below:

$$\begin{aligned} (\underset{\sim}{2},\underset{\sim}{3}) &= (1,1\,0) \\ [(\underset{\sim}{2},\underset{\sim}{3}) \otimes (\underset{\sim}{2},\underset{\sim}{3})]_A &= (0,2\,0) \oplus (2,0\,1), \\ [3 \text{ times}]_A &= (3,0\,0) \oplus (1,1\,1), \\ [4 \text{ times}]_A &= (2,1\,0) \oplus (0,0\,2), \\ [5 \text{ times}]_A &= (1,0\,1), \\ [6 \text{ times}]_A &= (0,0\,0). \end{aligned} \tag{4.7}$$

Using the above relation (4.7), the typical representation of the highest weight Λ of SU(2/3) is

$$\begin{array}{lll} \text{GND:} & \text{Bose} & \Lambda \otimes (0,0\,0) = \Lambda, \\ \text{1st:} & \text{Fermi} & \Lambda \otimes (1,1\,0), \\ \text{2nd:} & \text{Bose} & \Lambda \otimes [(0,2\,0) \oplus (2,0\,1)], \\ \text{3rd:} & \text{Fermi} & \Lambda \otimes [(3,0\,0) \oplus (1,1\,1)], \\ \text{4th:} & \text{Bose} & \Lambda \otimes [(2,1\,0) \oplus (0,0\,2)], \\ \text{5th:} & \text{Fermi} & \Lambda \otimes (1,0\,1), \\ \text{6th:} & \text{Bose} & \Lambda \otimes (0,0\,0) = \Lambda, \end{array} \tag{4.8}$$

where the ground floor is assumed to be a Bose system for convenience.[11] As expected, the numbers of bosons and fermions are all equal to $\dim(\Lambda) \cdot 2^5$.

Although the weight system (4.8) is constructed, the a_2 component is still missing, which is crucial to determine the typicality. As in the usual Lie algebra, each floor is connected with β_j^{i-} strings of (4.4). In this way one gets the a_2 component consistently. The atypical condition (3.1) determines whether an odd root string is decoupled or not. This atypical condition comes from the fact that the U(1) sector of $\mathrm{SU}(m) \otimes \mathrm{SU}(n) \otimes \mathrm{U}(1)$, which is reduced from $\mathrm{SU}(m/n)$, should be supertraceless for any irreducible representations.

Given a highest weight $(0\,a_2\,0\,0)$ the atypical condition is

$$a_2 = -4 + i + j, \quad 1 \leqslant i \leqslant 2, \quad 2 \leqslant j \leqslant 4. \tag{4.9}$$

If $i = j = 2$, a_2 is zero. This implies the β_2^{2-} string is terminated when $a_2 = 0$. Hence the representation $(0\,0\,0\,0)$ is the singlet composed of the ground floor alone. The representation $(0\,0\,0\,0)$ satisfies the supertraceless condition, which makes one sure that β_2^{2-} and all other odd roots are decoupled. In general, the weight system $(0\,0 \cdots 0)$ of $\mathrm{SU}(m/n)$ is singlet because the β_m^{m-} $(= (\overline{\underset{\sim}{m}},\underset{\sim}{n}))$ string is decoupled when all the weight components are zero.

The representation $(0\,a_2\,0\,0)$ is typical when $a_2 \neq -1$, 1, 0, and 2. This system is resolved into $\underset{\sim}{32}_{\mathrm{B}} \oplus \underset{\sim}{32}_{\mathrm{F}}$. The subscripts B and F denote Bose and Fermi systems, respectively.

The fundamental representation of SU(2/3) is $\underset{\sim}{2}_{\mathrm{B}} \oplus \underset{\sim}{3}_{\mathrm{F}}$ (Ref. 11), which coincides with the weight system $(1\,0\,0\,0)$, and its complex conjugate representation is $\underset{\sim}{\bar{3}}_{\mathrm{F}} \otimes \underset{\sim}{\bar{2}}_{\mathrm{B}}$, that is, $(0\,0\,0\,1)$. Their weight systems are

$$\begin{array}{llll} \text{GND : Bose} & (1\;0\;0\;0) & & \\ & \big\Downarrow \beta_2^- \quad \searrow^{\alpha_1^-} & & \\ & & (-1\;1\;0\;0) & \\ \text{1st : Fermi} & (0\;1\;1\;0) & & \\ & \quad \searrow^{\alpha_3^-} & & \\ & & (0\;0\;-1\;1) & \\ & & \quad \searrow^{\alpha_4^-} & \\ & & & (0\;0\;0\;-1) \end{array} \tag{4.10}$$

and

$$\begin{array}{llll} \text{GND : Fermi} & (0\;0\;0\;1) & & \\ & \big\Downarrow \beta_4^{2-} \quad \searrow^{\alpha_4^-} & & \\ & & (0\;0\;1\;-1) & \\ & & \quad \searrow^{\alpha_3^-} & \\ & & & (0\;-1\;-1\;0) \quad \cdot \\ \text{1st : Bose} & (1\;-1\;0\;0) & & \\ & \quad \searrow^{\alpha_1^-} & & \\ & (-1\;0\;0\;0). & & \end{array} \tag{4.11}$$

The representations $(1\,0\,0\,0)$ and $(0\,0\,0\,1)$ are supertraceless [cf. Sec. V, Eq. (5.8)].

The fundamental representation[8,11] of $\mathrm{SU}(m/n)$ $(n \neq 1)$ is generally $(1\,0 \cdots 0)$ $[= \underset{\sim}{m}_{\mathrm{B}} \oplus \underset{\sim}{n}_{\mathrm{F}}]$ and its conjugate representation is $(0 \cdots 0\,1)$ $[= \overline{\underset{\sim}{n}}_{\mathrm{F}} \oplus \overline{\underset{\sim}{m}}_{\mathrm{B}}]$. For $\mathrm{SU}(m/1)$, the fundamental is $(1\,0 \cdots 0)$ $[= \underset{\sim}{m}_{\mathrm{B}} \oplus \underset{\sim}{1}_{\mathrm{F}}]$ but its conjugate[6] is $(0 \cdots 0\;-1)$ $[= \underset{\sim}{\bar{1}}_{\mathrm{F}} \oplus \overline{\underset{\sim}{m}}_{\mathrm{B}}]$.

The tensor product of the fundamental representation and its conjugate representation of $\mathrm{SU}(m/n)$ is

$$\begin{aligned} &(1\,0 \cdots 0) \otimes (0\,0 \cdots 0\,1) \\ &\quad = (1\,0 \cdots 0\,1) \oplus (0 \cdots 0) \\ &\qquad (m \neq n, \quad n \neq 1), \end{aligned} \tag{4.12}$$

$$\begin{aligned} &(1\,0 \cdots 0) \otimes (0 \cdots 0\;-1) \\ &\quad = (1\,0 \cdots 0\;-1) \oplus (0 \cdots 0) \\ &\qquad (m \neq n, \quad n = 1), \end{aligned} \tag{4.13}$$

$$\begin{aligned} &(1\,0 \cdots 0) \otimes (0 \cdots 0\,1) \\ &\quad = (1\,0 \cdots 0\,1) \oplus (0 \cdots 0) \oplus (0 \cdots 0) \\ &\qquad (m = n). \end{aligned} \tag{4.14}$$

Similar to the usual Lie algebra $\mathrm{SU}(N)$, $(1\,0 \cdots 0\,1)$ is the adjoint representation of $\mathrm{SU}(m/n)$ for $n \neq 1$. The adjoint representation of $\mathrm{SU}(m/1)$ is $(1\,0 \cdots 0\;-1)$. The positive odd roots β_j^{i+} reside in the ground floor of the adjoint representation, and the even roots are in the first floor. The second floor consists of β_j^{i-}'s.

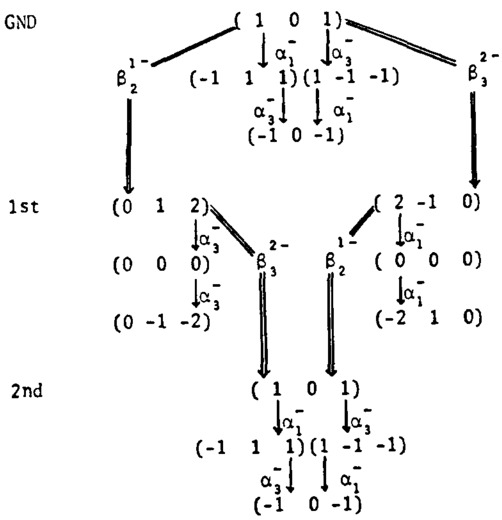
Example 2: SU(2/2): $\overset{a_1}{\circ}\!-\!\overset{a_2}{\bullet}\!-\!\overset{a_3}{\circ}$. As the simplest case for $m = n$, we consider SU(2/2). The graded Cartan matrix is

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}. \tag{4.15}$$

Then one has simple roots

$$\alpha_1^{\pm} = (\pm 2 \mp 1\, 0), \quad \alpha_3^{\pm} = (0 \pm 1 \pm 2),$$
$$\beta_2^{2\pm} = (\mp 1\, 0 \mp 1). \tag{4.16}$$

Here (1 0 1) is the adjoint representation of SU(2/2):



. (4.17)

The odd roots β_2^{2-} and β_3^{1-} are decoupled since $a_2 = 0$. The Cartan subalgebra $H_2 = (0\,0\,0)$ does not appear in the adjoint representation of SU(2/2). The dimension of this weight system is $\underline{14}$. Hence the even part of SU(2/2) is SU(2) ⊗ SU(2). Obviously the ground, first, and second floors consist of the positive odd roots, the even roots and the Cartan subalgebra, and negative odd roots, respectively.

V. BRANCHING RULES AND SUPERPROJECTION MATRICES

It is extremely important to know branching rules[9,13] of superalgebras in order to study low energy phenomenologies of superunification and supergravity models. We approach this problem in the context of superprojection matrices. The superalgebra $SU(m/n)$ has several branching types such as

$$SU(m/n) \to SU(m) \otimes SU(n) \otimes U(1), \tag{5.1}$$

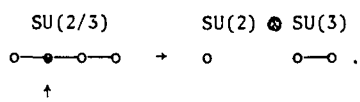
$$SU(m_1 + m_2/n_1 + n_2) \to SU(m_1/n_1) \otimes SU(m_2/n_2) \otimes U(1), \tag{5.2}$$

$$SU(m_1 m_2 + n_1 n_2/m_1 n_2 + m_2 n_1) \to SU(m_1/n_1) \otimes SU(m_2/n_2), \tag{5.3}$$

$$SU(m/n) \to OSp(m/n). \tag{5.4}$$

The first two branching patterns (5.1) and (5.2) are classified as regular branchings. The patterns (5.3) and (5.4) are special branchings.[7]

As an example of regular branchings, we analyze the SU(2/3) case. If one eliminates the second node (corresponding to the simple odd root) from the Kac–Dynkin diagram, SU(2/3) is broken down to the subgroup SU(2) ⊗ SU(3) ⊗ U(1),



The corresponding superprojection matrix of SU(2/3) → SU(2) ⊗ SU(3) ⊗ U(1) is

$$P = \begin{array}{c} \begin{array}{ccc} SU(2) & SU(3) & U(1) \end{array} \\ \left(\begin{array}{c:cc:c} 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & -2 \end{array}\right) \end{array} \tag{5.5}$$

Note that the matrix (5.5) is exactly the same as the projection matrix of SU(5) → SU(2) ⊗ SU(3) ⊗ U(1) except for the minus signs in the U(1) column.[7] The minus signs come from the supertraceless condition.

Let W be any weight system of SU(2/3). The SU(2) ⊗ SU(3) ⊗ U(1) contents W' are obtained by acting (5.5) on the right-hand side of W:

$$W \cdot P = W'. \tag{5.6}$$

Branching of the fundamental representation (1 0 0 0) is

$$\begin{array}{l} \text{floor} \\ \text{GND:} \\ \\ \\ \text{1st:} \\ \\ \\ \end{array}
\begin{array}{c} SU(2/3) \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ \hdashline 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{array}\right] \end{array} \cdot P =
\begin{array}{c} \begin{array}{ccc} SU(2) & SU(3) & U(1) \end{array} \\ \left[\begin{array}{c:cc:c} 1 & 0 & 0 & 3 \\ -1 & 0 & 0 & 3 \\ \hdashline 0 & 1 & 0 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & 0 & -1 & 2 \end{array}\right] \end{array}; \tag{5.7}$$

(1 0 0 0) decomposes into $[(1)(0\,0)_3 \oplus (0)(1\,0)_2]$, where the subscripts denote U(1) values. The generalized superprojection matrix of $SU(m/n) \to SU(m) \otimes SU(n) \otimes U(1)$ is

$$P=\begin{array}{c}\begin{array}{ccc}\mathrm{SU}(m) & \mathrm{SU}(n) & \mathrm{U}(1)\end{array}\\ \left[\begin{array}{c:c:c} & & n\\ & & 2n\\ I_{m-1} & 0 & \vdots\\ & & (m-1)\cdot n\\ \hdashline 0 & 0 & mn\\ \hdashline 0 & I_{n-1} & -(n-1)\cdot m\\ & & \vdots\\ 0 & I_{n-1} & -2m\\ & & -m\end{array}\right]\end{array}, \tag{5.8}$$

where I_m is the $m\times m$ identity matrix. Note that the U(1) column of the superprojection matrix (5.8) only gives some constant value times I_m when $\mathrm{SU}(m/m)\rightarrow\mathrm{SU}(m)\otimes\mathrm{SU}(m)$. The branching pattern (5.1) corresponds to elimination of the mth node from the Kac–Dynkin diagram. As a matter of fact, this pattern is just a special case ($m_2=n_1=0$) of (5.2).

The second branching pattern (5.2) is defined as

$$\begin{aligned}&\mathrm{SU}(m_1+m_2/n_1+n_2)\rightarrow\mathrm{SU}(m_1/n_1)\otimes\mathrm{SU}(m_2/n_2)\otimes\mathrm{U}(1),\\ &(1\,0\cdots0)\rightarrow(1\,0\cdots0)(0\cdots0)_a\oplus(0\cdots0)(1\,0\cdots0)_b.\end{aligned} \tag{5.9}$$

The subscripts a and b are U(1) values that satisfy the supertraceless condition

$$a(m_1-n_1)+b(m_2-n_2)=0. \tag{5.10}$$

This pattern is divided into four cases:

(i) $n_1=0$: $\mathrm{SU}(m_1+m_2/n)\rightarrow\mathrm{SU}(m_1)\otimes\mathrm{SU}(m_2/n)\otimes\mathrm{U}(1)$,

(ii) $m_2=0$: $\mathrm{SU}(m/n_1+n_2)\rightarrow\mathrm{SU}(m/n_1)\otimes\mathrm{SU}(n_2)\otimes\mathrm{U}(1)$,

(iii) $m_2=n_1=0$: $\mathrm{SU}(m/n)\rightarrow\mathrm{SU}(m)\otimes\mathrm{SU}(n)\otimes\mathrm{U}(1)$,

(iv) $m_1, m_2, n_1, n_2\neq0$.

Cases (i) and (ii) correspond to the elimination of the m_1th and $(m+n_1)$th node from the Kac–Dynkin diagram, respectively. As mentioned above, case (iii) coincides with (5.1). Case (iv) is not obtained by this simple method. For case (i), the generalized superprojection matrix is

$$\begin{array}{cl}\begin{array}{ccc}\mathrm{SU}(m_1) & \mathrm{SU}(m_2/n) & \mathrm{U}(1)\end{array} & \\ \left[\begin{array}{c:c:c} & & 1(n-m_2)\\ & & 2(n-m_2)\\ I_{m_1-1} & 0 & \vdots\\ & & (m_1-1)(n-m_2)\\ \hdashline 0 & 0 & m_1(n-m_2)\\ \hdashline & & m_1(n-m_2+1)\\ & & m_1(n-m_2+2)\\ & & \vdots\\ & & m_1(n-1)\\ 0 & I_{m_2+n-1} & m_1n\\ & & -m_1(n-1)\\ & & -m_1(n-2)\\ & & \vdots\\ & & -m_1\end{array}\right] & \begin{array}{l}\left.\begin{array}{c}\\ \\ \\ \\ \end{array}\right\}m_1-1\\ \quad m_1\text{th}\\ \left.\begin{array}{c}\\ \\ \\ \\ \end{array}\right\}m_2-1\\ \quad(m_1+m_2)\text{th}\\ \left.\begin{array}{c}\\ \\ \\ \\ \end{array}\right\}n-1\end{array}\end{array}. \tag{5.11}$$

The generalized superprojection matrix of the case (ii) is

$$
\begin{array}{ccc|l}
\mathrm{SU}(m/n_1) & \mathrm{SU}(n_2) & U(1) & \\
\left[\begin{array}{c:c:c}
 & & n_2 \\
 & & 2n_2 \\
 & & \vdots \\
 & & (m-1)n_2 \\
I_{m+n_1-1} & 0 & mn_2 \\
 & & -(m-1)n_2 \\
 & & -(m-2)n_2 \\
 & & \vdots \\
 & & -(m-n_1+1)n_2 \\
\hdashline
0 & 0 & -(m-n_1)n_2 \\
\hdashline
 & & -(m-n_1)(n_2-1) \\
0 & I_{n_2-1} & \vdots \\
 & & -(m-n_1)2 \\
 & & -(m-n_1)1
\end{array}\right]
& & &
\begin{array}{l}
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} m-1 \\
m\mathrm{th} \\
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} n_1-1 \\
(m+n_1)\mathrm{th} \\
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} n_2-1
\end{array}
\end{array}
\quad . \qquad (5.12)
$$

Now we consider the special patterns (5.3) and (5.4). Let $\Delta_\alpha = \{\alpha_i\}$ be a set of even roots and $\Delta_\beta = \{\beta_j\}$ a set of odd roots of $\mathrm{SU}(m/n)$. The commutation relations

$$[\alpha_i,\alpha_j] \in \Delta_\alpha, \quad [\alpha_i,\beta_j] \in \Delta_\beta, \quad \{\beta_i,\beta_j\} \in \Delta_\alpha \qquad (5.13)$$

allow the special branchings

$$
\begin{array}{ccc}
\mathrm{SU}(m_1m_2+n_1n_2/m_1n_2+m_2n_1) & \to & \mathrm{SU}(m_1/n_1) \otimes \mathrm{SU}(m_2/n_2) \\
(1\,0\cdots 0) & \to & (1\,0\cdots 0)(1\,0\cdots 0).
\end{array}
\qquad (5.14)
$$

As a simple example of this type, $\mathrm{SU}(2) \otimes \mathrm{SU}(1/2)$ can be embedded in $\mathrm{SU}(2.4)$ with either $n_1 = 0$ or $n_2 = 0$:

$$
P(\mathrm{SU}(2/4) \to \mathrm{SU}(2) \otimes \mathrm{SU}(1/2))
$$
$$
\begin{array}{ccc}
\mathrm{SU}(2) & \mathrm{SU}(1/2) & \\
\end{array}
$$
$$
= \begin{pmatrix}
1 & 1 & 0 \\
0 & 2 & 0 \\
1 & -1 & 1 \\
0 & 0 & 2 \\
1 & 0 & 1
\end{pmatrix}. \qquad (5.15)
$$

There exist special branching types[7]

$$\mathrm{SU}(N) \to \mathrm{SO}(N), \quad \mathrm{SU}(2n) \to \mathrm{Sp}(2n). \qquad (5.16)$$

The special branching $\mathrm{SU}(m/n) \to \mathrm{OSp}(m/n)$ (see Ref. 13) is based on (5.16).

VI. CONCLUDING REMARKS

We presented an easy systematic method of generating Kac–Dynkin weight systems. The Kac–Dynkin weight technique is extremely powerful in particle physics, especially in constructing supergravity models.[14] It is essential for Fermi and Bose systems to have the same degrees of freedom in supergravity. The typical representations of $\mathrm{SU}(m/n)$ may be nice candidates. Some of the atypical representations also have the same degrees of freedom[6] for bosons and fermions. For instance, the adjoint representation of $\mathrm{SU}(m-1/m)$ or $\mathrm{SU}(m/m-1)$ $(m \geqslant 2)$ has dim(Bose) = dim(Fermi) = $2m(m-1)$. Also the fundamental representations of $\mathrm{SU}(m/m)$ are the systems $\underset{\sim}{m}_{\mathrm{B}} \oplus \underset{\sim}{m}_{\mathrm{F}}$.

ACKNOWLEDGMENTS

We thank H. W. Lee and J. H. Yim for helpful discussions.

This research is supported in part by the Korea Science and Engineering Foundation, a Sogang University Research Grant, and by a Chungbuk National University Research Grant.

[1]P. Ramond, Phys. Rev. D **3**, 2415 (1971); V. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976); M. Scheunert, W. Nahm, and V. Rittenberg, *ibid.* **17**, 1626,1640 (1976).

[2]Y. Ne'eman, Phys. Lett. B **81**, 190 (1979); D. B. Fairlie, *ibid.* **82**, 97 (1979); J. G. Taylor, *ibid.* **83**, 331 (1979); R. E. Ecclestone, *ibid.* **116**, 21 (1982).

[3]F. Iachello, Phys. Rev. Lett. **44**, 772 (1980); A. B. Balantekin, I. Bars, and F. Iachello, *ibid.* **47**, 19 (1981); Nucl. Phys. A **370**, 284 (1981).

[4]S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976); D. Z. Freedman, P. Van Nieuwenhuizen, and S. Ferara, Phys. Rev. D **13**, 3214 (1976); B. Morel, Nucl. Phys. B **252**, 663 (1985).

[5]A. Neveu and J. H. Schwarz, Nucl. Phys. B **31**, 86 (1971).

[6]K. Y. Kim, C. H. Kim, Y. D. Kim, and Y. J. Park, J. Kor. Phys. Soc. (to be published).

[7]R. Slansky, Phys. Rep. **79**, 1 (1981); Y. D. Kim, K. Y. Kim, I. G. Koh, Y. J. Park, W. S. I'Yi, Y. S. Kim, C. H. Kim, and H. T. Park, *Tables of All Irreducible Representations for All Classical Groups: Complete Weight System Branching Matrices with U*(1) *and Anomaly Table* (Sogang U.P., Seoul, Korea, 1981); J. E. Kim, Phys. Rev. D **26**, 674 (1982); C. H. Kim,

I. G. Koh, Y. J. Park, K. Y. Kim, and Y. D. Kim, *ibid.* **27**, 1932 (1983).
[8]J. P. Hurni and B. Morel, J. Math. Phys. **24**, 157 (1983); I. Bars, B. Morel, and H. Ruegg, *ibid.* **24**, 2253 (1983).
[9]I. Bars, Yale University preprint YTP 82-25, 1982.
[10]F. W. Lemire and J. Patera, J. Math. Phys. **23**, 1409 (1982).
[11]P. Ramond, University of Florida Theoretical Physics preprint 84-1, 1984.
[12]A. N. Schellekens, I. G. Koh, and K. Kang, J. Math. Phys. **23**, 2244 (1982).
[13]B. G. Wybourne, Talk Given at the XIV ICGTMP, Seoul, 1985.
[14]C. H. Kim, K. Y. Kim, Y. D. Kim, W. S. l'Yi, and Y. J. Park, Sogang University preprint, Sogang-HEP 120 and 121, 1985.

The Hirota conditions

Alan C. Newell and Zeng Yunbo[a)]
Department of Mathematics, University of Arizona, Tucson, Arizona 85721

(Received 13 March 1986; accepted for publication 9 April 1986)

The condition on the polynomial P for a Hirota equation $P\tau\cdot\tau=0$ to have an N-soliton solution for arbitrary N is examined and simplified.

I. INTRODUCTION

While the role of affine Lie algebras in explaining many of the miracles of soliton mathematics is understood,[1–3] the Hirota conditions have so far eluded interpretation. These relations express the condition under which a given partial differential equation, when expressed in Hirota or quadratic (homogeneous) form, has an N-soliton or N-phase rational solution. It is generally agreed, although not rigorously proved, that, if these conditions hold for arbitrary N, the evolution equation is completely integrable and belongs to a commuting family, each of whose members is also a completely integrable soliton equation. The goal of this paper is to simplify the Hirota conditions and to express them in a way that may lead to an algebraic interpretation. In particular, we build on the idea, first expressed by one of the authors in Ref. 4, that the *phase shift function* plays a central role in identifying the members of a particular family of soliton equations. This function is common to each of the equations in the commuting family and measures the phase shift experienced by two colliding solitons. The fact that the same phase shift, which is a function of the two-soliton amplitudes, is shared by each of the members of the family is a simple consequence of the commutability of the flows.

The Hirota formalism homogenizes the partial differential equation by converting it into a bilinear, and in some cases a quadratic, equation. For example, the transformation

$$q(x,t)=2\frac{\partial^2}{\partial x^2}\ln\tau \tag{1.1}$$

converts the Korteweg–de Vries equation

$$q_t+6qq_x+q_{xxx}=0 \tag{1.2}$$

into the form

$$\tau\tau_{xt}-\tau_x\tau_t+\tau\tau_{xxxx}-4\tau_x\tau_{xxx}+3\tau_{xx}^2=0. \tag{1.3}$$

Hirota developed a very neat way of writing this equation by introducing a derivative operator D_{x_j}, which acts on ordered pairs of functions as follows:

$$D_x\sigma(x)\cdot\tau(x)=\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\sigma(x+\epsilon)\tau(x-\epsilon) \tag{1.4}$$

and in general

$$D_{x_1}^{\alpha_1}\cdots D_{x_n}^{\alpha_n}\sigma(x_r)\cdot\tau(x_r)=\prod_{r=1}^{n}\lim_{\epsilon_r\to 0}\frac{\partial^{\alpha_r}}{\partial\epsilon_r^{\alpha_r}}\sigma(x_r+\epsilon_r)\tau(x_r-\epsilon_r). \tag{1.5}$$

The right-hand sides of (1.4) and (1.5) are exactly the same as the Leibnitz formula for derivatives of products except for certain sign changes. Using this notation, Eq. (1.3) may be written (call $t=t_3$)

$$(D_xD_{t_3}+D_x^4)\tau\cdot\tau=0. \tag{1.6}$$

Associated with the Korteweg–de Vries equation is the polynomial $x_1x_3+x_1^4$ ($x_1=x$, $x_3=t_3$). Each member of the Korteweg–de Vries (KdV) family of equations may be written in quadratic form. The next member in the family, designated KdV 5, is

$$q_{t_5}-q_{xxxxx}-20q_xq_{xx}-10qq_{xxx}-30q^2q_x=0, \tag{1.7}$$

and, using (1.5), this may be written

$$(D_xD_{t_5}-D_x^6+\tfrac{5}{3}(D_xD_{t_3}+D_{t_3}^2))\tau\cdot\tau=0. \tag{1.8}$$

Notice that in order to write KdV 5 in quadratic form, one needs to include the KdV 3 time variable t_3 in addition to the time t_5 that appears in (1.7). The Hirota equations for KdV 5 are the pair of equations (1.6) and (1.8). We also observe that these two examples of Hirota equations are even and homogeneous under the weight assignment $W(D_{t_{2k+1}})=2k+1$.

Other well-known equations that have Hirota form are the Sawada–Kotera equation, $t_1=x$,

$$(D_{t_1}^6+9D_{t_1}D_{t_5})\tau\cdot\tau=0; \tag{1.9}$$

the Ramani equation

$$(D_{t_1}^6-5D_{t_1}^3D_{t_3}-5D_{t_3}^2)\tau\cdot\tau=0; \tag{1.10}$$

the Ito equation

$$(D_{t_3}^2+2D_{t_3}D_{t_1}^3)\tau\cdot\tau=0; \tag{1.11}$$

and the Kadomtsev–Petviashvili equation

$$(u_t+uu_x+u_{xxx})_x+u_{yy}=0, \tag{1.12}$$

which is transformed by (1.1) into

$$(\tfrac{3}{4}D_y^2-D_xD_{t_3}+\tfrac{1}{4}D_x^4)\tau\cdot\tau=0. \tag{1.13}$$

One of the advantages of the Hirota formalism is that it is relatively easy to find expressions for the multisoliton solution. The reason for this is that the N-phase multisoliton solution, which for the KdV family is given by

[a)] Present address: Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China.

 0022-2488/86/082016-06$02.50

$$\tau(x=t_1,t_3,t_5,\dots)$$
$$= \sum_{\mu_j,\mu_l=0,1} \exp\left(\sum_{j=1}^{N}\mu_j\theta_j + \sum_{1<j<l<N} A_{jl}\,\mu_j\,\mu_l\right), \tag{1.14}$$

consists of sums of exponentials and the Hirota operator D acts in a simple way on ordered pairs of exponentials, e.g.,

$$D_x^m e^{k_1x}\cdot e^{k_2x} = (k_1-k_2)^m e^{k_1x+k_2x}. \tag{1.15}$$

In (1.14),

$$\theta_j = \sum_0^\infty (-1)^n k_j^{2n+1} t_{2n+1}, \quad t_1 = x, \tag{1.16}$$

The phase shift A_{jl} is given by

$$e^{A_{jl}} = ((k_j-k_l)/(k_j+k_l))^2, \tag{1.17}$$

and the first sum is taken over all configurations of the μ_j, $j=1,\dots,N$, each choice being either a zero or a one.

We emphasize that (1.14) provides the common N-soliton solution for all members of the KdV family. As we have mentioned, they all share the same phase shift, a property that can be deduced readily from the fact that the flows $q_{t_{2r+1}}$, $r=0,1,\dots$, commute. The general formula analogous to (1.15) is

$$P(D_{t_1},D_{t_3},\dots)e^{\theta_j}\cdot e^{\theta_l} = P(\mathbf{k}_j-\mathbf{k}_l)e^{\theta_j+\theta_l}, \tag{1.18}$$

where

$$P(\mathbf{k}_j-\mathbf{k}_l) = P(k_j-k_l,\dots,(-1)^r(k_j^{2r+1}-k_l^{2r+1}),\dots). \tag{1.19}$$

We now ask a natural question. Given an even homogeneous polynomial $P_{2L}(D_{t_1},D_{t_3},\dots,D_{t_{2k+1}})$ of weight $2L$, under what conditions does the corresponding Hirota equation

$$P_{2L}(D_{t_1},D_{t_3},\dots,D_{t_{2k+1}})\tau\cdot\tau = 0 \tag{1.20}$$

having an N-soliton solution for arbitrary N? The one-soliton form

$$\tau = 1+e^{\theta}, \tag{1.21}$$

with

$$\theta = \sum_0^\infty k^{(2r+1)} t_{2r+1}, \quad t_1 = x, \tag{1.22}$$

is a solution provided that the vector $\{k^{(2r+1)}\}_0^\infty$ lies on the manifold (which we call the dispersion relation)

$$P_{2L}(k^{(1)},k^{(3)},\dots,k^{(2r+1)},\dots) = 0. \tag{1.23}$$

We are going to confine ourselves in this paper to the class of Hirota equation for which (1.23) is satisfied by $k^{(2r+1)}$ being a power of a single parameter k:

$$k^{(1)} = k, \quad k^{(2r+1)} = (-1)^r k^{2r+1}. \tag{1.24}$$

This corresponds to evolution equations like the Korteweg–de Vries equation, which describe how a function of $x=t_1$ evolves with respect to a sequence of times t_{2r+1}. The Kadomtsev–Petviashvili (KP) equation, on the other hand, is part of a family for which the equations describe the evolution in times $t_3,t_4,\dots$ of a function $q(x=t_1,y=t_2,t_3,t_4,\dots)$ of $x=t_1$ and $y=t_2$. The dispersion relative for (1.13) is satisfied by expressing each $k^{(r)}$ (here $\theta=\Sigma k^{(r)}t_r$) as a function of two parameters

$$k^{(1)} = u-v, \quad k^{(2)} = u^2-v^2, \quad k^{(3)} = u^3-v^3,\dots .$$

These equations are associated with the Lie algebra $\mathrm{gl}(\infty)$ corresponding to the infinite-dimensional linear group. On the other hand, the KdV hierarchy, which is recovered from the KP hierarchy by setting $v=-u=k/2$ and writing t_{2r+1} as $(-1)^r 2^{r+1} t_{2r+1}$, is associated with a subalgebra of $\mathrm{gl}(\infty)$, namely the Kac–Moody algebra $A_1^{(1)}$ associated with $\mathrm{sl}(2)$.

The two-soliton solution

$$\tau = 1+e^{\theta_1}+e^{\theta_2}+e^{A_{12}+\theta_1+\theta_2} \tag{1.25}$$

is a solution of (1.20) with

$$\theta_j = \sum_0^\infty (-1)^r k_j^{2r+1} t_{2r+1}$$

provided the phase shift is chosen as

$$e^{A_{12}} = -P_{2L}(\mathbf{k}_1-\mathbf{k}_2)/P_{2L}(\mathbf{k}_1+\mathbf{k}_2), \tag{1.26}$$

where $P(\mathbf{k}_1\pm\mathbf{k}_2)$ is defined by (1.19). The coefficients of $e^{2\theta_1}$ and $e^{2\theta_2}$ are zero because of (1.18) and the fact that $P_{2L}(\mathbf{0})$ is zero. The coefficient of $e^{2\theta_1+\theta_2}$ is zero because $P_{2L}(\mathbf{k}_1) = P_{2L}(k_1,-k_1^3,k_1^5,\dots)$ is zero. Thus, Hirota equations in quadratic form always have a two-soliton solution. For a three-soliton solution, there is an additional constraint, obtained by demanding that the coefficient of $e^{\theta_1+\theta_2+\theta_3}$ in the expression

$$P(D_{t_1},D_{t_3},\dots)\tau\cdot\tau$$

be zero. This condition can be written

$$\begin{aligned}
&p_{123}\{P_{2L}(\mathbf{k}_2-\mathbf{k}_3)P_{2L}(\mathbf{k}_3+\mathbf{k}_1)\\
&\quad\times P_{2L}(\mathbf{k}_2+\mathbf{k}_1)P_{2L}(\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}_3)\}\\
&\quad+P_{2L}(\mathbf{k}_2-\mathbf{k}_3)P_{2L}(\mathbf{k}_3-\mathbf{k}_1)\\
&\quad\times P_{2L}(\mathbf{k}_2-\mathbf{k}_1)P(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) = 0,
\end{aligned} \tag{1.27}$$

where

$$\begin{aligned}
&P(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\\
&\quad= P(k_1+k_2+k_3,-k_1^3-k_2^3-k_3^3,\dots,\\
&\qquad\times(-1)^r(k_1^r+k_2^r+k_3^r),\dots),
\end{aligned}$$

and p_{123} is the permutation over 1,2,3. For an N-soliton solution, the condition, originally derived by Hirota,[5] is

$$\sum_{\mu_j} P\left(\sum_1^N \mu_j\mathbf{k}_j\right)\prod_{m>l} P(\mu_m\mathbf{k}_m-\mu_l\mathbf{k}_l)\mu_l\mu_m = 0. \tag{1.28}$$

The summation in (1.28) is over all sequences $(\mu_1,\mu_2,\dots,\mu_N)$, where $\mu_j=\pm1$, $j=1,\dots,N$. In each term of the summation,

$$P\left(\sum_1^N \mu_j\mathbf{k}_j\right)\prod_{m>l} P(\mu_m\mathbf{k}_m-\mu_l\mathbf{k}_l)\mu_l\mu_m,$$

all the μ's are determined once a particular choice of the sequence $(\mu_1,\mu_2,\dots,\mu_N)$ of plus and minus ones is made. Equation (1.28) is known as the Hirota condition and we call a homogeneous polynomial P of even degree that satisfies this condition for all N *a Hirota polynomial.* It is the expression (1.28) that we aim to simplify. In particular, we would like to find an algorithm to determine all polynomials of weight $2L$ that satisfy it.

II. DISCUSSION OF RESULTS

The first curious fact about (1.28) is that it is not, on the surface, linear. And it should be, because integrable evolution equations come in families and therefore linear combinations of these flows should also be integrable and satisfy (1.28) for every N. However, recall that all the members of a commuting family share the same phase shift

$$e^{A_{ml}} = -\frac{P(\mathbf{k}_m - \mathbf{k}_l)}{P(\mathbf{k}_m + \mathbf{k}_l)} = -\frac{P_{2M}(\mathbf{k}_m - \mathbf{k}_l)}{P_{2M}(\mathbf{k}_m + \mathbf{k}_l)}, \tag{2.1}$$

where $2M$ is the lowest weight of any number of the integrable family. For example, the lowest weight of the KdV family (1.9) is that of the KdV equation itself, namely 4. The Sawada–Kotera family of integrable equations begins at level 6. Therefore in (1.28), we can replace the second P, which contains differences on two k's only, with P_{2M}, because dividing (1.28) across by $P_{2L}(\mathbf{k}_l + \mathbf{k}_m)$ gives an equation linear in $P_{2L}(\Sigma_1^N \mu_l \mathbf{k}_l)$ with coefficients of functions of the phase shifts, which are the same for every L in the commuting family of Hirota equations. With this observation, the Hirota condition for a given P_{2L} can now be written

$$\begin{aligned} Q(k_1,\dots,k_N) &= \sum P_{2L}\left(\sum_1^N \mu_j \mathbf{k}_j\right) \\ &\quad\times \prod_{m>l} P_{2M}(\mu_m \mathbf{k}_m - \mu_l \mathbf{k}_l)\mu_l \mu_m = 0. \end{aligned} \tag{2.2}$$

What we will show is that if $Q(k_1,\dots,k_s) = 0$ for $s \leqslant N-1$, then (2.2) has a factor

$$k_1^{N+1} \cdots k_N^{N+1} \prod_{m>l} (k_m^2 - k_l^2)^2$$

of degree $3N^2 - N$. But, from (2.2), a straightforward count shows that $Q(k_1,\dots,k_N)$ has degree $2L + MN(N-1)$. Thus if

$$3N^2 - N > 2L + MN(N-1), \tag{2.3}$$

$Q(k_1,\dots,k_N)$ must be identically zero. For cases in which $M = 2$ or 3, that is, in those cases for which the lowest weight member of the integrable sequence is 4 or 6, this condition is nontrivial and tells us that after one establishes that P_{2L} has an r-soliton solution, $3 \leqslant r \leqslant N_0$, where N_0 is the maximum integer for which (2.3) holds, then it has an N-soliton solution for arbitrary N. For $M = 2$, $N_0 = [(-1 + \sqrt{1+8L})/2]$ and for $M = 3$, $N_0 = L$. In actual fact one simply has to establish that P_{2L} has an N_0-soliton solution because, by simply allowing several soliton amplitudes to decay or their locations to move to infinity, the fact that P_{2L} has an N_0-soliton solution implies that it has an r-soliton solution $3 \leqslant r < N_0$.

Let us look at several consequences of this result. Denote by $P_{2L}^{(2M)}$ the polynomial weight $2L$, which has the phase shift function given by (2.1). Then we have the following.

(i) If $M = L = 2$, (2.3) holds for every $N \geqslant 2$. It follows that

$$P_4(D_{t_1}, D_{t_3})\tau \cdot \tau = 0$$

has an N-soliton solution for all N. This is the well-known result that the KdV 3 equation has an N-soliton solution.

(ii) If $M = L = 3$, then (2.3) is satisfied for any $N > 3$. This implies that if

$$P_6(D_{t_1}, D_{t_3}, D_{t_5})\tau \cdot \tau = 0$$

has a three-soliton solution, then it has an N-soliton solution for arbitrary N.

(iii) In the case $M = 3$, when $N > L$, (2.3) holds. Therefore, if

$$P_{2L}^{(6)}(D_{t_1}, D_{t_3}, \dots)\tau \cdot \tau = 0$$

has an L-soliton solution, then it has an N-soliton solution for arbitrary N.

(iv) In the case $M = 2$, (2.3) is satisfied provided

$$N > [(-1 + \sqrt{1+8L})/2] = N_0.$$

Therefore, if

$$P_{2L}^{(4)}(D_{t_1}, D_{t_3}, \dots)\tau \cdot \tau = 0 \tag{2.4}$$

has a N_0-soliton solution, then it has an N-soliton solution for arbitrary N. For instance, when $L = 3,4,5$, (2.4) has an N-soliton solution for arbitrary N; when $L = 6,7,8,9$, if (2.4) has a three-soliton solution, then it has an N-soliton solution for arbitrary N.

Because we have assumed the dispersion relation (1.23) is satisfied by (1.24), this last result refers to members of the KdV family. It shows that, contrary to the conjecture stated by the first author in Ref. 4, the Hirota polynomials (that is, the polynomials that have N-soliton solutions for arbitrary N) are not completely determined by the phase shift function. Namely, just because P_{2L} satisfies (2.1) with $M = 2$ is not sufficient to guarantee it is a Hirota polynomial. As we have just mentioned, it is sufficient for $L = 3,4,5$ that is, for polynomials of weights 6, 8, and 10. For polynomials of weights 12–18, P_{2L} needs also to satisfy the three-soliton condition. For a polynomial of general weight $2L$, P_{2L} must satisfy (2.2) for all N up to $[(-1 + \sqrt{1+8L})/2]$.

Since the general form of the polynomial at any weight level is a linear combination of all products of odd weights that add to $2L$, these constraints leads to a set of homogeneous linear algebraic equations on the W_L coefficients, where W_L is the number of ways an even number $2L$ can be decomposed into a sum of odd numbers less than $2L$. It is reasonable to conjecture that these equations will contain information about the underlying algebraic structure of the equation family whose phase shift function is given by (2.1).

III. PROOF OF MAIN RESULT

Consider the equation in Hirota form

$$P_{2L}(D_{t_1}, D_{t_3}, \dots)\tau \cdot \tau = 0, \tag{3.1}$$

with the phase shift function given by

$$e^{A_{12}} = -\frac{P_{2L}(\mathbf{k}_1 - \mathbf{k}_2)}{P_{2L}(\mathbf{k}_1 + \mathbf{k}_2)} = -\frac{P_{2M}(\mathbf{k}_1 - \mathbf{k}_2)}{P_{2M}(\mathbf{k}_1 + \mathbf{k}_2)}, \quad M \leqslant L, \tag{3.2}$$

where P_{2L} and P_{2M} satisfy the conditions

$$P_{2i}(-D_{t_1}, -D_{t_3},\ldots) = P_{2i}(D_{t_1}, D_{t_3},\ldots), \tag{3.3}$$

$$P_{2i}(0,0,\ldots) = 0, \tag{3.4}$$

$$P_{2i}(\mathbf{k}) = P_{2i}(k, -k^3, k^5,\ldots) = 0, \tag{3.5}$$

and we define

$$P_{2i}(\mathbf{k}_1 \pm \mathbf{k}_2) = P_{2i}(k_1 \pm k_2, -(k_1^3 \pm k_2^3),\ldots, \times(-1)^r(k_1^{2r+1} \pm k_2^{2r+1}),\ldots). \tag{3.6}$$

The condition that (3.1) has an N-soliton solution is

$$\sum P_{2L}\left(\sum_1^N \mu_i \mathbf{k}_i\right)\prod_{j>i} P_{2L}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)\mu_i\mu_j = 0,$$

which is equivalent to [using (3.2)]

$$\sum P_{2L}\left(\sum_1^N \mu_i \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)\mu_i\mu_j = 0, \tag{3.7}$$

where the sum is taken over all sequences $(\mu_i)_{i=1}^N$ of plus and minus ones. It is easy to see that (3.7) can be rewritten as

$$\begin{aligned} Q_N &= Q(k_1,\ldots,k_N) \\ &= \sum_{\mu_j=-1,1} P_{2L}\left(\sum_1^N \mu_i \mathbf{k}_i\right) \\ &\quad \times \prod_{j>i} P_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)\prod_i \mu_i^{N-1} = 0. \end{aligned} \tag{3.8}$$

In the following proof, we follow closely the ideas used by Hirota[6] for proving that the KdV equation has an N-soliton solution. The Q_N has the following properties.

(i) When N is odd, Q_N is even in the k_i; when N is even, Q_N is odd in each of the k_i, i.e.,

$$Q(k_1,\ldots,-k_i,\ldots,k_N) = (-1)^{N-1}Q(k_1,\ldots,k_i,\ldots,k_N). \tag{3.9}$$

(ii) Q_N is a homogeneous symmetric polynomial in the k_i's, i.e.,

$$Q(k_1,\ldots,k_i,\ldots,k_j,\ldots,k_N) = Q(k_1,\ldots,k_j,\ldots,k_i,\ldots,k_N). \tag{3.10}$$

The result (3.9) is easily seen by replacing k_i by $-k_i$ and μ_i by $-\mu_i$ (dummy index) in (3.8). Also, (3.10) can be verified by interchanging k_i, k_j, and μ_i,μ_j.

It is clear from (3.5) and (3.8) that

$$Q(k_1) = P_{2L}(\mathbf{k}_1) = 0, \tag{3.11}$$

$$\begin{aligned} Q(k_1,k_2) &= P_{2L}(\mathbf{k}_1 + \mathbf{k}_2)P_{2M}(\mathbf{k}_1 - \mathbf{k}_2) \\ &\quad - P_{2L}(\mathbf{k}_1 - \mathbf{k}_2)P_{2M}(\mathbf{k}_1 + \mathbf{k}_2) = 0. \end{aligned} \tag{3.12}$$

Theorem: Provided

$$Q(k_1,\ldots,k_i) = 0, \quad i<N-1, \tag{3.13}$$

then

$$Q(k_1,\ldots,k_N) = k_1^{N+1}\cdots k_N^{N+1}\prod_{j>i}^N (k_j^2 - k_i^2)^2\tilde{Q}(k_1,\ldots,k_N), \tag{3.14}$$

and if N,L,M satisfy

$$3N^2 - N > 2L + MN(N-1), \tag{3.15}$$

then

$$Q(k_1,\ldots,k_N) = 0. \tag{3.16}$$

Proof: (3.5) implies

$$P_{2M}(\mu_j \mathbf{k}_j) = 0, \tag{3.17}$$

which yields

$$P_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i) = k_j k_i \overline{P}_{2M}(\mu_j,\mu_i,k_j,k_i).$$

Hence we obtain from (3.8) that

$$Q(k_1,\ldots,k_N) = k_1^{N-1}\cdots k_N^{N-1}\overline{Q}(k_1,\ldots,k_N). \tag{3.18}$$

By using (3.13) and (3.17) and noting that (3.3) implies

$$P(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i) = P(\mathbf{k}_j - \mu_i\mu_j \mathbf{k}_i), \tag{3.19}$$

and

$$\begin{aligned} \left.\frac{dP_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)}{dk_i}\right|_{k_i=0} &= \left.\frac{dP_{2M}(\mathbf{k}_j - \mu_i\mu_j \mathbf{k}_i)}{dk_i}\right|_{k_i=0} \\ &= \mu_i\mu_j \left.\frac{dP_{2M}(\mathbf{k}_j - \mathbf{k}_i)}{dk_i}\right|_{k_i=0}, \end{aligned} \tag{3.20}$$

we find

$$\begin{aligned} &\left.\frac{d^{N-1}Q(k_1,\ldots,k_N)}{dk_1^{N-1}}\right|_{k_1=0} \\ &= \sum_{\mu_j=-1,1} P_{2L}\left(\sum_{i=2}^N \mu_i \mathbf{k}_i\right)\prod_{j>i\geq 2} P_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)\prod_{i=1}^N \mu_i^{N-1}\prod_{j=2}^N \left.\frac{dP_{2M}(\mu_j \mathbf{k}_j - \mu_1 \mathbf{k}_1)}{dk_1}\right|_{k_1=0} \\ &= 2\prod_{j=2}^N \left.\frac{dP_{2M}(\mathbf{k}_j - \mathbf{k}_1)}{dk_1}\right|_{k_1=0} \sum_{\substack{\mu_j=-1,1 \\ j>1}} P_{2L}\left(\sum_{i=2}^N \mu_i \mathbf{k}_i\right)\prod_{j>i\geq 2} P_{2M}(\mu_j \mathbf{k}_j - \mu_i \mathbf{k}_i)\prod_{i=2}^N \mu_i^{N-2} \\ &= 2\prod_{j=2}^N \left.\frac{dP_{2M}(\mathbf{k}_j - \mathbf{k}_1)}{dk_1}\right|_{k_1=0} Q(k_2,\ldots,k_N) = 0. \end{aligned} \tag{3.21}$$

According to (3.9) and (3.18), Q_N can be written as

$$Q(k_1,\ldots,k_N) = k_1^{N-1}R_1(k_2,\ldots,k_N) + k_1^{N+1}R_2(k_2,\ldots,k_N) + k_1^{N+3}R_3(k_2,\ldots,k_N) + \cdots.$$

Using (3.21), we obtain

$$R_1(k_2,\ldots,k_N) = 0$$

and

$$Q(k_1,\ldots,k_N) = k_1^{N+1}R_2(k_2,\ldots,k_N) + k_1^{N+3}R_3(k_2,\ldots,k_N) + \cdots = k_1^{N+1}R(k_1,k_2,\ldots,k_N).$$

Therefore, from the properties (3.9) and (3.10), it follows that

$$Q(k_1,\ldots,k_N) = k_1^{N+1}\cdots k_N^{N+1}\hat{Q}(k_1,\ldots,k_N), \tag{3.22}$$

where the polynomial $\hat{Q}(k_1,\ldots,k_N)$ is even and symmetric in the k_i's.

Next, evaluate Q_N when $k_1 = k_2$:

$$\begin{aligned}
Q(k_1,k_1,k_3,\ldots,k_N) &= \sum_{\substack{\mu_2=\mu_1\\ \mu_j=-1,1}} P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\prod_{i=3}^N \mu_i^{N-1}\prod_{\substack{j>i\\ j>2}} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)P_{2M}(\mu_1\mathbf{k}_2-\mu_1\mathbf{k}_1)\Big|_{k_1=k_2}\\
&\quad+\sum_{\substack{\mu_2=-\mu_1\\ \mu_j=-1,1}} P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\prod_{i=3}^N \mu_i^{N-1}(-1)^{N-1}\prod_{j>i\geq 3} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\\
&\quad\times\prod_{j\geq 3}[P_{2M}(\mu_j\mathbf{k}_j+\mu_1\mathbf{k}_2)P_{2M}(\mu_j\mathbf{k}_j-\mu_1\mathbf{k}_1)]P_{2M}(\mu_1\mathbf{k}_2+\mu_1\mathbf{k}_1)\Big|_{k_1=k_2}\\
&= \sum_{\substack{\mu_2=-\mu_1\\ \mu_j=-1,1}} P_{2L}\left(\sum_{i=3}^N \mu_i\mathbf{k}_i\right)\prod_{i=3}^N \mu_i^{N-1}\prod_{j>i\geq 3} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\\
&\quad\times(-1)^{N-1}\prod_{j\geq 3}[P_{2M}(\mathbf{k}_j+\mu_1\mu_j\mathbf{k}_1)P_{2M}(\mathbf{k}_j-\mu_1\mu_j\mathbf{k}_1)]P_{2M}(\mu_1\mathbf{k}_1+\mu_1\mathbf{k}_1)\\
&= (-1)^{N-1}2\prod_{j\geq 3}[P_{2M}(\mathbf{k}_j+\mathbf{k}_1)P_{2M}(\mathbf{k}_j-\mathbf{k}_1)]P_{2M}(\mathbf{k}_1+\mathbf{k}_1)Q(k_3,\ldots,k_N)=0.
\end{aligned}\tag{3.23}$$

Since Q_N is a symmetric polynomial in the k_i's, (3.23) implies that for any i,j,

$$Q(k_1,\ldots,k_N)|_{k_i=k_j}=0,$$

and from (3.22) this yields

$$\hat{Q}(k_1,\ldots,k_N)|_{k_i=k_j}=0.$$

Hence $\hat{Q}_N$ is certainly factorized by (k_i-k_j) and therefore $\hat{Q}_N$, as a symmetric polynomial in the k_i's, must be factorized by

$$\prod_{\substack{i,j=1\\ i\neq j}}(k_i-k_j) \quad\text{or}\quad \prod_{j>l}(k_j-k_i)^2.$$

But since $\hat{Q}_N$ is even in the k_i, $\hat{Q}_N$ must be factorized by

$$\prod_{j>l}(k_i^2-k_j^2)^2.$$

This implies that (3.14) holds. So the order of Q_N must be at least $3N^2-N$. However the order of P_{2L} is $2L$ and the order of the polynomial product

$$\prod_{j>i}^N P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)$$

is $MN(N-1)$; hence according to (3.8) the order of Q_N must be at most $2L+MN(N-1)$. Clearly, if $Q_N\neq 0$, it must satisfy that

$$3N^2-N\leq\text{order}(Q_N)\leq 2L+MN(N-1).$$

Therefore, if $3N^2-N>2L+MN(N-1)$, there is a contradiction and we must conclude that $Q(k_1,\ldots,k_N)=0$.

IV. FURTHER SIMPLIFICATION AND EXAMPLES

The previous section has pointed out that in order to see whether the equations

$$P_{2L}^{(4)}(D_{t_1},D_{t_3},\ldots)\tau\cdot\tau=0 \tag{4.1}$$

or

$$P_{2L}^{(6)}(D_{t_1},D_{t_3},\ldots)\tau\cdot\tau=0 \tag{4.2}$$

has an N-soliton solution for arbitrary N, we only need to check whether it has an r-soliton solution $r\leq N_0$. However, it is not trivial to check the condition (3.8) for some r and therefore it is useful to simplify it further.

If $H(k_1,\ldots,k_N)$ is a polynomial of $k_1,\ldots,k_N$,

$$\begin{aligned}
H(k_1,\ldots,k_N) &= \sum a(l_1,\ldots,l_N)k_1^{2l_1}\cdots k_N^{2l_N}\\
&\quad+\sum b(j_1,\ldots,j_N)k_1^{2j_1+1}\cdots k_N^{2j_N+1}\\
&\quad+\sum c(i_1,\ldots,i_N)k_1^{i_1}\cdots k_N^{i_N},
\end{aligned}$$

where some of the $i_1,\ldots,i_N$ in the last sum are even and some of them are odd.

Define the operators L_e and L_o as follows:

$$L_eH(k_1,\ldots,k_N) = \sum a(l_1,\ldots,l_N)k_1^{2l_1}\cdots k_N^{2l_N},$$

$$L_oH(k_1,\ldots,k_N) = \sum b(j_1,\ldots,j_N)k_1^{2j_1+1}\cdots k_N^{2j_N+1}.$$

Proposition: The condition (3.8) that (3.1) has an N-soliton solution is equivalent to

$$L_e\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right]=0, \quad \text{when } N \text{ is odd,} \tag{4.3}$$

$$L_o\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right]=0, \quad \text{when } N \text{ is even.} \tag{4.4}$$

Proof: It is easy to see from the definition for the operators L_e and L_o that

$$L_e\left[P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\right] = L_e\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right]$$

and

$$L_o\left[P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\prod_{j>1} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\right] = \prod_{i=1}^N \mu_i L_o\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right].$$

Since Q_N is even in k_i when N is odd we have, for N odd,

$$\begin{aligned}
Q(k_1,\ldots,k_N) &= L_e Q(k_1,\ldots,k_N)\\
&= \sum_{\mu=-1,1} L_e\left[P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\right]\\
&= 2^N L_e\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P(\mathbf{k}_j-\mathbf{k}_i)\right].
\end{aligned}$$

When N is even, Q_N is odd in k_i, hence we get

$$\begin{aligned}
Q(k_1,\ldots,k_N) &= L_o Q(k_1,\ldots,k_N)\\
&= \sum_{\mu_j=-1,1} L_o\left[P_{2L}\left(\sum_1^N \mu_i\mathbf{k}_i\right)\right.\\
&\qquad\left.\times\prod_{j>i} P_{2M}(\mu_j\mathbf{k}_j-\mu_i\mathbf{k}_i)\prod_{i=1}^N \mu_i\right]\\
&= \sum_{\mu_j=-1,1} L_o\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right]\prod_{i=1}^N \mu_i^2\\
&= 2^N L_o\left[P_{2L}\left(\sum_1^N \mathbf{k}_i\right)\prod_{j>i} P_{2M}(\mathbf{k}_j-\mathbf{k}_i)\right].
\end{aligned}$$

As an example, we will use these simplifications to identify all equations with weight level 6 with the Hirota property. The most general form of P_6 is

$$D_{t_1}D_{t_5}+aD_{t_3}^2+bD_{t_1}^3D_{t_3}+cD_{t_1}^6. \tag{4.5}$$

From (3.5), a,b,c, must satisfy

$$1+a-b+c=0. \tag{4.6}$$

The theorem given in the previous section told us that (4.5) has an N-soliton solution for arbitrary N if it has a three-soliton solution. Therefore the condition that (4.5) has an N-soliton solution reads

$$\begin{aligned}
&L_e[P_6(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)P_6(\mathbf{k}_3-\mathbf{k}_2)\\
&\quad\times P_6(\mathbf{k}_3-\mathbf{k}_1)P_6(\mathbf{k}_2-\mathbf{k}_1)]=0.
\end{aligned} \tag{4.7}$$

Notice that this expression is considerably simpler than (1.7). Using (4.6) and (4.7), a little calculation shows

$$\begin{aligned}
&(3a+6c+1)(9c-1)\\
&\quad\times[a^2+(7c+2)a+c^2+2c+1)]=0.
\end{aligned} \tag{4.8}$$

This implies that all the Hirota equations at weight level 6 are the following equations: (i) KdV equation,

$$\begin{aligned}
&(D_{t_1}D_{t_5}+(-\tfrac{1}{3}-2c)D_{t_3}^2\\
&\quad+(\tfrac{2}{3}-c)D_{t_1}^3D_{t_3}+cD_{t_1}^6)\tau\cdot\tau=0;
\end{aligned} \tag{4.9}$$

and

$$\begin{aligned}
&\text{(ii)}\ (D_{t_1}D_{t_5}+aD_{t_3}^2+(\tfrac{10}{9}+a)\\
&\qquad\times D_{t_1}^3D_{t_3}+\tfrac{1}{9}D_{t_1}^6)\tau\cdot\tau=0.
\end{aligned} \tag{4.10}$$

Taking $a\to\infty$, we get Ito's equation from (4.10),

$$(D_{t_3}^2+2D_{t_1}^3D_{t_3})\tau\cdot\tau=0.$$

$$\begin{aligned}
&\text{(iii)}\ (D_{t_1}D_{t_5}+(-\tfrac{7}{2}c-1\pm\tfrac{1}{2}\sqrt{45c^2+20c})D_{t_3}^2\\
&\qquad+(-\tfrac{5}{2}c\pm\tfrac{1}{2}\sqrt{45c^2+20c})D_{t_1}^3D_{t_3}\\
&\qquad+cD_{t_1}^6)\tau\cdot\tau=0.
\end{aligned} \tag{4.11}$$

Taking $c=-1$, (4.11) yields the Sawada–Kotera equation after rescaling the variables

$$(D_{t_1}^6+9D_{t_1}D_{t_5})\tau\cdot\tau=0.$$

We obtain the Ramani equation by taking $c\to\infty$ in (4.11),

$$(D_{t_1}^6-5D_{t_1}^3D_{t_3}-5D_{t_3}^2)\tau\cdot\tau=0. \tag{4.12}$$

We emphasize that this is a complete list of all Hirota polynomials of weight 6 that satisfy the conditions (3.3)–(3.5).

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant No. NSF-DMS-8403187, Air Force Office of Scientific Research Grant No. AFOSR-83-0227, Army Grant No. DAAG-29-85-K0091, Office of Naval Research Grant No. Physics-N00014-84-K-0420, and the Science Fund of the Chinese Academy of Sciences.

[1] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, "Transformation groups for soliton equations," *Proceedings of the RIMS Symposium on Nonlinear Integrable Systems—Classical and Quantum Theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).

[2] H. Flaschka, A. C. Newell, and T. Ratiu, "Kac–Moody algebras and soliton equations. II and III," Physica D 9, 300, 324 (1983).

[3] J. Lepowsky and R. L. Wilson, "Construction of the affine Lie algebra $A_1^{(1)}$," Commun. Math. Phys. **62**, 43 (1978).

[4] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).

[5] R. Hirota, "Direct Methods in soliton theory," in *Topics in Current Physics*, Vol. 17, edited by R. Bullough and P. Caudrey (Springer, New York, 1980), pp. 157–176.

[6] R. Hirota, "Exact solution of the Korteweg–de Vries equation for multiple collisions of solitons," Phys. Rev. Lett. **27**, 1192 (1972).

Spinor structures on spheres and projective spaces

Ludwik Dabrowski[a] and Andrzej Trautman[b]
Scuola Internazionale Superiore di Studi Avanzati, 34014 Trieste, Italy

(Received 5 March 1986; accepted for publication 2 April 1986)

An explicit construction of spinor structures on real, complex, and quaternionic projective spaces is given for all cases when they exist. The construction is based on a theorem describing the bundle of orthonormal frames of a homogeneous Riemannian manifold. This research is motivated by a remarkable coincidence of spinor connections on low-dimensional spheres with simple, topologically nontrivial gauge configurations.

I. INTRODUCTION

Spinors—and structures associated with them—are indispensable in physics and important in geometry. They have become an essential tool in theoretical physics of particles and nuclei; they are also useful in the study of gravitation.[1] A proper treatment of spinors on manifolds, with an account of their topology, is relatively recent.[2] It has led to the deep idea of spin cobordism,[3] to a study of harmonic spinors[4] and of the index theorem for the Dirac operator.[5]

In physics, spinors have recently acquired a new significance through the twistor program,[6] work on supersymmetry and unified theories based on higher-dimensional geometries of the Kaluza–Klein type. There are interesting ideas on the possible physical relevance of inequivalent spinor structures.[7] The Feynman method of quantization based on sums over classical histories applied to gravity coupled to fermions requires an analysis of nontrivial spinor configurations.

A somewhat unexpected link between spinors and another part of physics consists in the recognition of coincidences between spinor connections on low-dimensional spheres and simple, topologically nontrivial gauge configurations.[8] Indeed, any sphere S_n of dimension $n \geqslant 2$ has a unique spinor structure. The Levi-Civita connection corresponding to the standard Riemannian metric on S_n lifts to a spinor connection, which may be interpreted as a "gauge configuration" for the group $\mathrm{Spin}(n)$. This configuration is invariant under the action of $\mathrm{Spin}(n+1)$ and satisfies the Yang–Mills equations on S_n. For example, the cases $n = 2,3$, and 4 correspond to the Dirac magnetic pole of lowest strength, the meron solution, and the instanton-cum-anti-instanton system, respectively. Landi[9] has shown that the spinor connection on S_8 concides with a gauge configuration described recently by Grossman, Kephart, and Stasheff (GKS).[10] Rawnsley[11] generalized the duality properties of the instanton and of the GKS solution to the gauge field obtained from the spinor connection on any $4k$-dimensional sphere. The local, differential-geometric properties of the spinor gauge fields and Riemannian curvature tensors of spheres are the same, but their global properties are different; in particular, they have different values of "topological charge." For example, the Levi-Civita connection on S_2 corresponds to a magnetic pole of strength twice the lowest, Dirac value. The meron charge is related to the Chern–Simons[12] conformal invariants.

These considerations have led us to study spinor structures on projective spaces, which are, after spheres, the simplest homogeneous manifolds. The natural spinor connections on these spaces also may be interpreted as simple gauge configurations, but we postpone their description to another work. In this paper, we restrict ourselves to the construction of the spinor structures themselves.

In order to find the spinor structures on a Riemannian manifold it is convenient to know its bundle of orthonormal frames. For a "generic" manifold without isometries there is not much one can say about this bundle: it is, for example, a parallelizable manifold, but this does not help much in constructing spinor structures. If, however, the manifold is homogeneous, i.e., admits a transitive Lie group G of isometries, then its bundle of frames can be explicitly described in terms of G and its subgroups. Moreover, the bundle of orthonormal frames can be restricted to a subgroup of the full orthogonal group. Such a restriction is convenient because it allows one to work with a bundle of lower dimension than that of the bundle of all orthonormal frames.

If a Riemannian n-manifold M is orientable, then its bundle of frames can be restricted to $\mathrm{SO}(n)$. This group admits a unique, nontrivial (for $n > 1$) double covering by the spin group, $\mathrm{Spin}(n) \rightarrow \mathrm{SO}(n)$. A spin structure on M is a "prolongation" of the bundle of frames that "agrees" with this covering. (The precise definition is recalled in Sec. II.) It exists if, and only if, the second Stiefel–Whitney class of M is zero. In the nonorientable case the situation is somewhat more complicated (Whiston[13]). The full orthogonal group $\mathrm{O}(n)$ has, in general, several inequivalent double coverings. For example, for $n = 1$, $\mathrm{Spin}(1) = \mathbb{Z}_2$ and $\mathrm{SO}(1) = 1$, but the orthogonal group $\mathrm{O}(1) = \mathbb{Z}_2$ has two different coverings:

$$\rho_+ : \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_2 \quad \text{and} \quad \rho_- : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2 .$$

In any dimension n, two such coverings ρ_+ and ρ_- can be obtained from Clifford algebras of $\mathbb{R}^n$ equipped, respectively, with quadratic forms ϕ and $-\phi$, where

$$\phi(x) = x_1^2 + \cdots + x_n^2 .$$

The topological obstructions to the existence of prolongations of the bundle of frames associated with ρ_+ and ρ_- are different from each other. We use the term "pin$^{\pm}$ structure"

[a] Permanent address: Instytut Fizyki Teoretycznej, Uniwersytet Wrocławski, Cybulskiego 36, 50-205 Wrocław, Poland.

[b] Permanent address: Instytut Fizyki Teoretycznej, Uniwersytet Warszawski, Hoża 69, 00-681 Warszawa, Poland.

 0022-2488/86/082022-07$02.50

for such a prolongation corresponding to $\rho^{\pm}$. In Sec. III we show that a real projective space of dimension $2k$ admits two inequivalent pin$^+$ or pin$^-$ structures depending on whether k is even or odd. We also give an explicit description of the spin structures on odd-dimensional complex projective spaces in terms of metaunitary groups.

II. PRELIMINARIES: SPINOR GROUPS AND STRUCTURES

This chapter contains a brief summary of the basic definitions and results related to spinor groups and structures that are needed in the sequel. We follow rather closely the articles by Atiyah, Bott, and Shapiro,[14] Atiyah and Bott,[15] and Karoubi,[16] but we adapt the notation and terminology to our needs.[17]

Let (e_i), $i = 1,\ldots,n$, be the standard orthonormal frame in $\mathbb{R}^n$. We denote by $C^+(n)$ and $C^-(n)$ the two related Clifford algebras generated by the e's subject to the relations

$$e_ie_j + e_je_i = +2\delta_{ij} \quad \text{and} \quad -2\delta_{ij} \quad (i,j=1,\ldots,n),$$

respectively. In any of the Clifford algebras we have the main involution α and the main anti-involution β. The groups $\mathrm{Pin}^+(n)$ and $\mathrm{Pin}^-(n)$ consist of all invertible elements of $C^+(n)$ and $C^-(n)$, respectively, which preserve the underlying vector space $\mathbb{R}^n$ under the twisted adjoint representation $\rho_\pm$,

$$\rho_\pm(s)v = \alpha(s)vs^{-1}, \quad \text{where} \quad v\in\mathbb{R}^n, \quad s\in C^{\pm}(n),$$

and are normalized by $|\beta(s)s| = 1$. [The last condition is meaningful because the previous ones imply $\beta(s)s\in\mathbb{R}$.] In general, the groups $\mathrm{Pin}^+(n)$ and $\mathrm{Pin}^-(n)$ are nonisomorphic. The connected components of the identity of these groups consist of even elements and are isomorphic to each other; they are both denoted by $\mathrm{Spin}(n)$. The sequences

$$1\to\mathbb{Z}_2\to\mathrm{Pin}^{\pm}(n)\xrightarrow{\rho_\pm}\mathrm{O}(n)\to 1$$

and

$$1\to\mathbb{Z}_2\to\mathrm{Spin}(n)\xrightarrow{\rho}\mathrm{SO}(n)\to 1, \quad \text{where} \quad \rho=\rho_\pm|\mathrm{Spin}(n),$$

are exact. We use the generic term "spinor group" to denote one of the groups $\mathrm{Spin}(n)$, $\mathrm{Pin}^+(n)$, or $\mathrm{Pin}^-(n)$, $n = 1,2,\ldots$. The centers of these groups are as shown in Table I. Here $\mathbb{Z}_2 = \{1,-1\}$, $\mathbb{Z}_2^+ = \{1,\epsilon\}$, $\mathbb{Z}_2^- = \{1,-\epsilon\}$, $\mathbb{Z}_4 = \{1,\epsilon,-1,-\epsilon\}$, and $\epsilon = e_1e_2\cdots e_n$ is the "volume element." The products occurring in Table I are direct. Note also that if $\epsilon\in\mathrm{Pin}^{\pm}(n)$, then

$$\epsilon^2 = (\pm 1)^n(-1)^{n(n-1)/2}.$$

TABLE I. Centers of spinor groups.

n	$\mathrm{Spin}(n)$	$\mathrm{Pin}^+(n)$	$\mathrm{Pin}^-(n)$
$4k$	$\mathbb{Z}_2^+\times\mathbb{Z}_2^-$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
$4k+1$	$\mathbb{Z}_2$	$\mathbb{Z}_2^+\times\mathbb{Z}_2^-$	$\mathbb{Z}_4$
$4k+2$	$\mathbb{Z}_4$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
$4k+3$	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_2^+\times\mathbb{Z}_2^-$

The existence of spinor structures on projective spaces depends crucially on the structure of the center of an appropriate spinor group. It is convenient to define

$$\mathrm{Pin}(n) = \begin{cases}\mathrm{Pin}^+(n), & \text{for} \quad n\equiv 0,1 \bmod 4,\\ \mathrm{Pin}^-(n), & \text{for} \quad n\equiv 2,3 \bmod 4,\end{cases}$$

and make a corresponding notational convention for the covering map ρ. If Σ is one of the groups $\mathrm{Spin}(4k)$ or $\mathrm{Pin}(2k+1)$, $k=1,2,\ldots$, then $[s]_\pm$ denotes the coset $s\mathbb{Z}_2^{\pm}\in\Sigma/\mathbb{Z}_2^{\pm}$, i.e., if $s,t\in\Sigma$, then

$$[s]_\pm = [t]_\pm \quad \text{iff} \quad s=t \text{ or } s=\pm t\epsilon.$$

We inject $\mathbb{R}^n$ into $\mathbb{R}^{n+1}$ by sending $(x_1,\ldots,x_n)$ into $(x_1,\ldots,x_n,0)$ and extend this to injections of the corresponding Clifford algebras and spinor groups.

In order to adapt to our purposes the classical definition of a spinor structure on the tangent bundle of Riemannian manifold M (see Haefliger and Milnor in Ref. 2), consider the following. Let M be n dimensional with a positive-definite metric tensor g. Let Ω be a closed subgroup of $\mathrm{O}(n)$ and $\Sigma = \rho_\pm^{-1}(\Omega)\subset\mathrm{Pin}^{\pm}(n)$. Assume that $F\subset F_g$ is a restriction to Ω of the bundle F_g of all linear frames on M that are orthonormal with respect to g. A spinor structure on M is defined by giving a prolongation of $\pi\colon F\to M$ to the group Σ, i.e., principal Σ-bundle $\sigma\colon S\to M$ and a morphism of bundles $\eta\colon S\to F$ such that there is a commutative diagram

$$\begin{tikzcd} S\times\Sigma \arrow[r] \arrow[d, "\eta\times\rho_\pm"'] & S \arrow[d, "\eta"] \arrow[dr, "\sigma"] & \\ F\times\Omega \arrow[r] & F \arrow[r, "\pi"'] & M, \end{tikzcd}$$

where the horizontal arrows denote the action maps. If $\Omega\subset\mathrm{SO}(n)$, then M is orientable and one has a spin structure. If M is nonorientable, then F_g is connected, but Ω is not, and one has a pin$^{\pm}$ structure. The two structures pin$^+$ and pin$^-$ corresponding to the two covering maps ρ^+ and ρ^- are, in general, inequivalent. In some cases one exists whereas the other does not as may be seen from the topological conditions for their existence[16]: $w_2=0$ for pin$^+$, $w_1^2+w_2=0$ for pin$^-$, and $w_1=0$, $w_2=0$ for spin. (Here w_1 and w_2 denote, respectively, the first and second Stiefel–Whitney classes of the tangent bundle of M.) We sometimes say "pin structure" when we mean one of the two and we use the generic term "spinor structure" to denote a pin or spin structure.

It is clear that, given the bundles described above, one can always extend their structure groups Ω and Σ to $\mathrm{O}(n)$ and $\mathrm{Pin}^{\pm}(n)$, respectively. The extended bundles provide a classical description of pin structure. Conversely, given such a classical pin structure on M, say

$$\mathrm{Pin}^{\pm}(n)\to P\xrightarrow{f}F_g\to M,$$

and a restriction F of F_g to $\Omega\subset\mathrm{O}(n)$, one can restrict the structure group $\mathrm{Pin}^{\pm}(n)$ of P to Σ by taking the induced bundle $S=f^{-1}(F)$. Similar remarks apply to spin structures.

The classical definition of equivalence of spinor structures can be easily adapted to our considerations. Let, for simplicity, $\eta_a\colon S_a\to F$ $(a=1,2)$ be two spin structures, where each S_a is the total space of a principal Σ-bundle over M. They are equivalent if there is a based isomorphism

i: $S_1 \to S_2$ of principal Σ-bundles such that $\eta_2 \circ i = \eta_1$. The bundles $\pi \circ \eta_a$: $S_a \to M$ may be isomorphic, as principal Σ-bundles, without defining equivalent spin structures. The equivalence classes of isomorphic spinor structures are in a bijective correspondence with the elements of $H^1(M,\mathbb{Z}_2)$, the first cohomology group of M with coefficients in $\mathbb{Z}_2$ (see Milnor[2] and Whiston[13]).

III. FRAME BUNDLES OF HOMOGENEOUS SPACES

We restrict our study to very regular situations: all manifolds and maps are *smooth*, Lie groups and other spaces are *compact*, and subgroups are *closed*. All Riemannian manifolds are *proper*, i.e., their metric tensors are positive definite. The italicized adjectives will be omitted from now on.

Let M be a manifold admitting a transitive left action γ: $G \times M \to M$ of a Lie group G. Denoting $\gamma_a(x) = \gamma(a,x)$ one has $\gamma_a \circ \gamma_b = \gamma_{ab}$, for any $a,b \in G$, and $\gamma_1 = \mathrm{id}$, where 1 is the unit of G. Let $H = \{a \in G: \gamma_a(x) = x\}$ be the *stability* ("little") group at $x \in M$. The homogeneous space M is canonically diffeomorphic to the quotient space G/H. The diffeomorphism h: $G/H \to M$, mapping the coset bH, $b \in G$, into $\gamma_b(x)$, intertwines the actions of G in G/H and M, $h \circ \gamma_a = \gamma_a \circ h$ for all $a \in G$ (cf., for example, Bredon[18]). We shall often identify G/H with M and, by doing so, make h disappear.

Let γ'_a denote the tangent map to γ_a at x. For any $a \in H$, this map is a linear automorphism of the tangent space $T_x M$ to M at x, and

$$\gamma': H \to \mathrm{GL}(T_x M)$$

is a homomorphism of groups. Its kernel N is a normal subgroup of H and, therefore, also a subgroup—but not normal, in general—of G. According to the general theory of fiber bundles (Steenrod[19]) these data define a principal H/N bundle

$$F = G/N \to G/H = M, \tag{1}$$

where the action of H/N in F is given by $(aN)(bN) = abN$, $a \in G$ and $b \in H$.

Let now M be an n-dimensional Riemannian manifold with a metric tensor g admitting a group G of isometries acting transitively on M. The preceding construction leads to the following theorem.

Theorem: The bundle π: $F \to M$, defined by (1), is a restriction to the group H/N of the bundle π_g: $F_g \to M$ of all linear frames on M, orthonormal with respect to the metric tensor g.

To prove the theorem, it suffices to give an injective immersion i: $F \to F_g$ and a monomorphism of Lie groups j: $H/N \to \mathrm{O}(n)$ such that

$$i((aN)(bN)) = i(aN)\, j(bN) \quad \text{and} \quad \pi_g \circ i = \pi. \tag{2}$$

Recall that an orthonormal frame in an n-dimensional vector space V may be identified with an isometry from $\mathbb{R}^n$ to V. Let f be a frame at x, orthonormal with respect to g. This frame is unchanged by γ'_a, $a \in H$, if and only if $a \in N$. For any $a \in G$, the composition $\gamma'_a \circ f$ is a frame at $\gamma_a(x)$, also orthonormal with respect to g. The maps i and j are now defined by

$$i(aN) = \gamma'_a \circ f, \quad a \in G,$$

and

$$j(bN) = f^{-1} \circ \gamma'_b \circ f, \quad b \in H,$$

where

$$f^{-1}: T_x M \to \mathbb{R}^n$$

is the inverse, or "dual," frame with respect to f. The morphism properties (2) are now easy to verify.

Example 1: The $(n-1)$-dimensional sphere S_{n-1} with its standard Riemannian metric admits $\mathrm{SO}(n)$ as a group of isometries. The action γ of $\mathrm{SO}(n)$ on S_{n-1} is transitive and the stability group of any point is isomorphic to $\mathrm{SO}(n-1)$, whereas $N = \ker \gamma'$ reduces to the identity. The $\mathrm{SO}(n-1)$-bundle,

$$\mathrm{SO}(n) \to \mathrm{SO}(n)/\mathrm{SO}(n-1) = S_{n-1},$$

is simply the bundle of orthonormal frames of S_{n-1} with coherent orientation. For n even, $n = 2k$, the group $\mathrm{SO}(2k)$ contains a subgroup $\mathrm{U}(k)$, which also acts transitively on S_{2k-1}. The stability group is $\mathrm{U}(k-1)$ and

$$\mathrm{U}(k) \to \mathrm{U}(k)/\mathrm{U}(k-1) = S_{2k-1}$$

is the bundle of "unitary frames." Similarly, for $n = 4l$, there is the bundle of "symplectic frames"

$$\mathrm{SP}(l) \to \mathrm{Sp}(l)/\mathrm{Sp}(l-1) = S_{4l-1}.$$

Example 2: Let K denote one of the three number fields: $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. The set K^{n+1} is a right module (a vector space if $K = \mathbb{R}$ or $\mathbb{C}$) over K: if $q = (q_\alpha) \in K^{n+1}$, $\alpha = 1,\dots,n+1$, and $\lambda \in K$, then

$$q\lambda = (q_\alpha \lambda) \in K^{n+1},$$

so that

$$(q\lambda)\mu = q(\lambda\mu), \quad q(\lambda + \mu) = q\lambda + q\mu, \text{ etc.},$$

for any $\lambda, \mu \in K$. If $q \in K^{n+1}$ and $q \neq 0$, then the direction of q is the set

$$\mathrm{dir}\, q = \{q\lambda: 0 \neq \lambda \in K\}$$

and the set of all such directions is the projective n-dimensional space over K,

$$KP_n = \{\mathrm{dir}\, q: 0 \neq q \in K^{n+1}\}.$$

The module K^n has a natural, positive-definite quadratic form ϕ given by

$$\phi(q) = \sum_{\alpha=1}^{n} \bar{q}_\alpha q_\alpha,$$

where $\bar{\lambda} = \lambda$ for $K = \mathbb{R}$ and $\bar{\lambda}$ is the conjugate of λ otherwise. Let $\mathrm{U}(n,K)$ be the set of all maps a: $K^n \to K^n$ such that $\phi \circ a = \phi$, $a(q\lambda) = a(q)\lambda$, and $a(q+q') = a(q) + a(q')$ for any $\lambda \in K$ and $q,q' \in K^n$. With respect to composition of maps this set is a group, namely

$$\mathrm{U}(n,K) = \begin{cases} \mathrm{O}(n), \\ \mathrm{U}(n), \\ \mathrm{Sp}(n), \end{cases} \quad \text{for } K = \begin{cases} \mathbb{R}, \\ \mathbb{C}, \\ \mathbb{H}. \end{cases}$$

The action of $\mathrm{U}(n+1,K)$ in KP_n given by

$$\gamma_a(\mathrm{dir}\, q) = \mathrm{dir}\, a(q)$$

is transitive. Let e_α denote the element of K^{n+1} consisting of 1 at the αth place and zeros elsewhere so that

$$q = \sum_{\alpha=1}^{n+1} e_\alpha q_\alpha .$$

The stability group H of $x = \operatorname{dir} e_{n+1} \in KP_n$ may be computed as follows; let

$$a(e_\beta) = \sum_\alpha e_\alpha a_{\alpha\beta} ,$$

where $a_{\alpha\beta} \in K$ and $\alpha,\beta = 1,...,n+1$. The condition $\operatorname{dir} a(e_{n+1}) = \operatorname{dir} e_{n+1}$ implies $a_{\alpha,n+1} = 0$ for $\alpha = 1,...,n$. Since $\phi \circ a = \phi$ is equivalent to

$$\sum_\gamma \bar{a}_{\gamma\alpha} a_{\gamma\beta} = \delta_{\alpha\beta} ,$$

we obtain also $a_{n+1,\alpha} = 0$, for $\alpha = 1,...,n$, so that H is isomorphic to $U(1,K) \times U(n,K)$. The isomorphism is realized as follows: if $\lambda \in U(1,K)$ and $b \in U(n,K)$, then the corresponding element of H is represented by the matrix

$$a = \begin{pmatrix} b & 0 \\ 0 & \lambda \end{pmatrix} . \tag{3}$$

Let y: $\mathbb{R} \to KP_n$ be a curve through x, $y(0) = x$. For sufficiently small $|t|$ one can write

$$y(t) = \operatorname{dir}(e_{n+1} + q(t)) ,$$

where

$$q(t) = \sum_{\alpha=1}^{n} e_\alpha q_\alpha(t) , \quad q(0) = 0 ,$$

is a curve in K^n. The tangent vector to y at $t = 0$ is represented by

$$\dot{q}(0) = v = \sum_{\alpha=1}^{n} e_\alpha v_\alpha .$$

If $a \in H$ is as in (3), then the tangent vector to the curve

$$\begin{aligned} t &\to \gamma_a(y(t)) \\ &= \operatorname{dir} a(e_{n+1} + q(t)) = \operatorname{dir}(e_{n+1}\lambda + bq(t)) \\ &= \operatorname{dir}(e_{n+1} + bq(t)\lambda^{-1}) \end{aligned}$$

is represented by

$$\gamma'_a(v) = bv\lambda^{-1} \in K^n .$$

Therefore, the kernel N of γ' consists of all matrices of the form (3) such that $bv = v\lambda$ for any $v \in K^n$. It follows that b is λ times the unit automorphism of K^n and λ belongs to the center of K. The group N is thus isomorphic with the center of $U(n+1,K)$,

$$N = \begin{cases} \mathbb{Z}_2 , \\ U(1) , \\ \mathbb{Z}_2 , \end{cases} \quad \text{for } K = \begin{cases} \mathbb{R} , \\ \mathbb{C} , \\ \mathbb{H} . \end{cases}$$

Taking into account

$$U(1,K) = \begin{cases} \mathbb{Z}_2 , \\ U(1) , \\ Sp(1) , \end{cases} \quad \text{for } K = \begin{cases} \mathbb{R} , \\ \mathbb{C} , \\ \mathbb{H} , \end{cases}$$

we can compute the structure groups

$$H/N = \begin{cases} O(n) , \\ U(n) , \\ (Sp(n) \times Sp(1))/\mathbb{Z}_2 , \end{cases} \quad \text{for } \begin{cases} \mathbb{R}P_n , \\ \mathbb{C}P_n , \\ \mathbb{H}P_n , \end{cases}$$

and the reduced bundles of orthonormal frames

$$G/N = \begin{cases} O(n+1)/\mathbb{Z}_2 , \\ U(n+1)/U(1) , \\ Sp(n+1)/\mathbb{Z}_2 \end{cases} \quad \text{for } \begin{cases} \mathbb{R}P_n , \\ \mathbb{C}P_n , \\ \mathbb{H}P_n . \end{cases}$$

For n even, $n = 2k$, the quotient $O(2k+1)/\mathbb{Z}_2$ can be identified with $SO(2k+1)$: the bundle of all orthonormal frames of $\mathbb{R}P_{2k}$ is connected, i.e., $\mathbb{R}P_{2k}$ is not orientable. Note that the quotient $O(2k+1)/\mathbb{Z}_2$ may also be represented as $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^+ \times \mathbb{Z}_2^-$. For n odd, $n = 2k-1$, the quotient $O(2k)/\mathbb{Z}_2$ is the disjoint sum of two copies of $SO(2k)/\mathbb{Z}_2$. Therefore, $\mathbb{R}P_{2k-1}$ is orientable and its bundle of orthonormal frames of coherent orientation is diffeomorphic to $SO(2k)/\mathbb{Z}_2$. The bundle $U(n+1)/U(1)$ is diffeomorphic to the quotient $SU(n+1)/Z_{n+1}$ of the group $SU(n+1)$ by its center.

Example 3: Consider a Lie group G with a bi-invariant metric; e.g., if G is semisimple, then such a metric is obtained from the Killing form. In this case, the manifold of G is a homogeneous Riemannian space with respect to the action of $G \times G$ given by

$$\gamma_{(a,b)}(c) = acb^{-1} ,$$

for any $a,b,c \in G$. The stability group at the unit element of G is isomorphic to G embedded diagonally in $G \times G$. For any $a \in G$, the map $\gamma'_{(a,a)}$ coincides with Ad_a, where

$$\mathrm{Ad}: G \to GL(\mathfrak{g})$$

is the adjoint representation of G in its Lie algebra $\mathfrak{g} = T_1 G$. The kernel of Ad is the center $Z(G)$ of G and the reduced bundle of orthonormal frames is a $G/Z(G)$-bundle

$$(G \times G)/Z(G) \to G . \tag{4}$$

Note that, unless G is Abelian, the total space of the bundle (4) is "larger" than that obtained by considering G as a homogeneous space with respect to left translations.

IV. SPINOR STRUCTURES ON SPHERES AND PROJECTIVE SPACES

In this section we use our description of the restricted bundle of orthonormal frames to construct spinor structures on spheres and on the projective spaces: $\mathbb{R}P_n$ ($n = 1$ or $n > 1$ and $\neq 1 \bmod 4$), $\mathbb{C}P_{2k-1}$ ($k = 1,2,...$), and $\mathbb{H}P_n$ ($n = 1,2,...$). The case of spheres is easy and well known. For an orientable projective space over K, the crucial information is contained in the structure of the center $Z(n,K)$ of the group $\Sigma(n,K) = \rho^{-1}(U(n,K) \cap SO(m))$, where ρ: $\mathrm{Spin}(m) \to SO(m)$ is the covering map and $m = n \dim_{\mathbb{R}} K$. If the center is a direct product of the form $\mathbb{Z}_2 \times \Lambda(n,K)$, then there is a spin structure on KP_{n-1} given by the sequence

$$\Sigma(n,K)/\Lambda(n,K) \to \Sigma(n,K)/Z(n,K) \to KP_{n-1} . \tag{5}$$

A similar statement applies to the nonorientable space $\mathbb{R}P_{2k}$; here the relevant groups are $\mathrm{Pin}(2k+1)$ and its center. Each real projective space other than $\mathbb{R}P_{4l+1}$ ($l = 1,2,...$) has two pin or spin structures. We construct them both and show that they are inequivalent.

A. Spheres

The circle S_1 has two inequivalent spin structures (Milnor[2]). Since both S_1 and $SO(2)$ can be identified with $U(1)$, and $\mathrm{Spin}(1)=\mathbb{Z}_2$, these structures are given by the maps

$$\mathrm{U}(1)\times\mathbb{Z}_2\xrightarrow[\mathrm{pr}_1]{}\mathrm{U}(1)\xrightarrow[\mathrm{id}]{}\mathrm{U}(1)$$

and

$$\mathrm{U}(1)\xrightarrow[\text{square}]{}\mathrm{U}(1)\xrightarrow[\mathrm{id}]{}\mathrm{U}(1)\,.$$

For any $n\geqslant 2$, there is a unique spinor structure given by

$$\mathrm{Spin}(n+1)\to\mathrm{SO}(n+1)\to S_n\,.$$

For $n=4l-1$ $(l=1,2,...)$ one can restrict the bundle of frames to $\Omega=\mathrm{Sp}(l-1)$ and the spinor bundle to $\Sigma=\mathrm{Sp}(l-1)\times\mathbb{Z}_2$. The (restricted) spinor structure is

$$\mathrm{Sp}(l)\times\mathbb{Z}_2\to\mathrm{Sp}(l)\to S_{4l-1}\,.$$

There is an analogous restriction of the spinor bundle of S_{2k-1} to the metaunitary group $\mathrm{MU}(k)\subset\mathrm{Spin}(2k)$, cf. Sec. IV C.

B. Real projective spaces

(i) Consider first the case of odd dimension. The one-dimensional real projective space is diffeomorphic to the circle S_1; its spin structures have already been given. Let now the dimension $n=2k-1$ be greater than 1. The space $\mathbb{R}P_{2k-1}$ is orientable and the fundamental group Π_1 of its bundle of frames $\mathrm{SO}(2k)/\mathbb{Z}_2$ may be computed by considering three curves in $\mathrm{Spin}(2k)$ joining 1 to ϵ, $-\epsilon$, and -1, respectively. Each of these curves projects to a loop in $\mathrm{SO}(2k)/\mathbb{Z}_2$ and defines a nontrival element of Π_1. No two of these elements coincide and, since $\epsilon^2=(-1)^k$, one has $\Pi_1=\mathbb{Z}_4$ for k odd and $\Pi_1=\mathbb{Z}_2\times\mathbb{Z}_2$ for k even. The group $\mathrm{SO}(2k-1)$ is the fiber of

$$\mathrm{SO}(2k)/\mathbb{Z}_2\to\mathbb{R}P_{2k-1}$$

and its fundamental group $\mathbb{Z}_2$ is embedded in Π_1 as follows:

if k is odd, $k\geqslant 3$,

then $\mathbb{Z}_2\to\mathbb{Z}_4$ is given by 1 mod 2$\mapsto$2 mod 4;

if k is even,

then $\mathbb{Z}_2\to\mathbb{Z}_2\times\mathbb{Z}_2$ is the diagonal map.

To check this for odd k one can consider the projection to $\mathrm{SO}(2k)/\mathbb{Z}_2$ of the curve in $\mathrm{Spin}(2k)$ joining 1 and -1. This projection is the square of the loop obtained by projecting the curve joining 1 and ϵ. The square is represented by 2 mod 4 in $\mathbb{Z}_4$ and, since it is noncontractible, it is homotopic to a nontrivial loop in $\mathrm{SO}(2k-1)$. It is now clear that $\mathbb{R}P_{4l+1}$ $(l=1,2,...)$ has no spinor structure: a noncontractible loop in a fiber of its bundle of frames ("rotation by 360°") can be continuously deformed into the square of a loop in the bundle (Clarke[20]). This result is, of course, well known: $\mathbb{R}P_{4l+1}$ has $w_1=0$ and $w_2\neq 0$ for $l=1,2,...$.

The space $\mathbb{R}P_{4l-1}$ has $w_1=0$, $w_2=0$, and $\pi_1=\mathbb{Z}_2$. There are, therefore, two inequivalent spin structures on $\mathbb{R}P_{4l-1}$ $(l=1,2,...)$. They are

$$\pi^\pm:\ \mathrm{Spin}(4l)/\mathbb{Z}_2^\pm\to\mathrm{SO}(4l)/\mathbb{Z}_2\,,$$

where the $\pi^\pm$ are obvious projections and the action of $\mathrm{Spin}(4l-1)$ in $\mathrm{Spin}(4l)/\mathbb{Z}_2^\pm$ is obtained from the natural action of $\mathrm{Spin}(4l-1)$ in $\mathrm{Spin}(4l)$ by passing to the quotient. To see that π^+ and π^- define inequivalent spin structures consider a curve in $\mathrm{Spin}(4l)$ connecting 1 with ϵ. Its projection to $\mathrm{SO}(4l)/\mathbb{Z}_2$ is a loop. There are exactly two lifts of this loop to $\mathrm{Spin}(4l)/\mathbb{Z}_2^+$ and they are both closed curves (loops). There are also exactly two lifts of this loop to $\mathrm{Spin}(4l)/\mathbb{Z}_2^-$ and neither of them is closed. This contradicts the existence of a bundle isomorphism h: $\mathrm{Spin}(4l)/\mathbb{Z}_2^+\to\mathrm{Spin}(4l)/\mathbb{Z}_2^-$ such that $\pi^-\circ h=\pi^+$.

There is, however, an orientation–reversing isometry

$$j:\ \mathbb{R}P_{4l-1}\to\mathbb{R}P_{4l-1},$$
$$\mathrm{dir}(\rho(a)e_{4l})\mapsto\mathrm{dir}(\rho(e_{4l}a)e_{4l}),$$

which lifts to an isomorphism of one spin structure onto the other, given by $[a]_+\mapsto[e_{4l}ae_{4l}]_-$.

(ii) The even-dimensional real projective spaces are nonorientable; they will be shown to admit pin structures. For any $k=1,2,...,$ the space $\mathbb{R}P_{2k}$ admits two inequivalent pin structures. Depending on whether k is even or odd, one has to consider the covering map $\mathrm{Pin}(2k)\to\mathrm{O}(2k)$ corresponding to a pin group associated with an Euclidean space $\mathbb{R}^{2k}$ with a quadratic form that is positive or negative, respectively (cf. Sec. II).

The pin structures on $\mathbb{R}P_{2k}$ are

$$\pi^\pm:\ \mathrm{Pin}(2k+1)/\mathbb{Z}_2^\pm\to\mathrm{Pin}(2k+1)/\mathbb{Z}_2^+\times\mathbb{Z}_2^-=\mathrm{SO}(2k+1),$$

where the projections $\pi^\pm$ are obvious and the action δ of $\mathrm{Pin}(2k)$ in $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^\pm$ comes from the natural embedding $\mathrm{Pin}(2k)\to\mathrm{Pin}(2k+1)$ by passing to the quotient, i.e.,

$$\delta_b([a]_\pm)=[ab]_\pm\,,$$

for any $a\in\mathrm{Pin}(2k+1)$ and $b\in\mathrm{Pin}(2k)$. The inequivalence of π^+ and π^- may be seen as follows. Consider a curve in $\mathrm{Pin}(2k+1)$ beginning at 1 and ending at $e_1\epsilon$. Its projection to $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^+\times\mathbb{Z}_2^-$ has the property that its end is obtained by applying $\rho(e_1)\in\mathrm{O}(2k)$ to its beginning. There are again exactly two lifts of this curve to each $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^+$ and $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^-$. The starting and end points of the curves in $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^\pm$ are related to each other by the action of $\pm e_1$, respectively. This difference in sign implies that there is no isomorphism of bundles h such that $\pi^-\circ h=\pi^+$.

The total spaces $\mathrm{Pin}(2k+1)/\mathbb{Z}_2^\pm$ are both diffeomorphic to $\mathrm{Spin}(2k+1)$. More precisely, let

$$\sigma:\mathrm{Spin}(2k+1)\to\mathbb{R}P_{2k} \tag{6a}$$

be the projection $a\mapsto\mathrm{dir}\,\rho(a)e_{2k+1}$ and

$$\delta^\pm:\ \mathrm{Spin}(2k+1)\times\mathrm{Pin}(2k)\to\mathrm{Spin}(2k+1) \tag{6b}$$

be right actions defined by

$$\delta_b^\pm(a)=\delta^\pm(a,b)=\begin{cases}ab\,, & \text{if } b \text{ is even}\,,\\ \pm\epsilon ab\,, & \text{if } b \text{ is odd}\,.\end{cases}$$

The two maps $h^\pm:\mathrm{Pin}(2k+1)/\mathbb{Z}_2^\pm\to\mathrm{Spin}(2k+1)$ given by

$$[a]_\pm\mapsto\begin{cases}a\,, & \text{if } a \text{ is even}\,,\\ \pm\epsilon a\,, & \text{if } a \text{ is odd}\,,\end{cases}$$

define, respectively, isomorphisms of the two principal bundles (6) with the bundles

$$\sigma^\pm: \mathrm{Pin}(2k+1)/\mathbf{Z}_2^\pm \to \mathbf{R}P_{2k}, \tag{7}$$

where $\sigma^\pm = \pi\circ\pi^\pm$ and

$$\pi: \mathrm{SO}(2k+1) \to \mathbf{R}P_{2k}.$$

We have indeed

$$\sigma\circ h^\pm = \sigma^\pm \quad \text{and} \quad h^\pm\circ\delta_b = \delta_b^\pm\circ h^\pm,$$

for any $b\in\mathrm{Pin}(2k)$.

Even though the two spin structures on $\mathbf{R}P_{2k}$ are inequivalent, the two bundles (6) are *isomorphic* to each other when considered as principal bundles over $\mathbf{R}P_{2k}$. Indeed, a based isomorphism

$$i: \mathrm{Pin}(2k+1)/\mathbf{Z}_2^+ \to \mathrm{Pin}(2k+1)/\mathbf{Z}_2^-$$

is given by

$$[a]_+ \mapsto [\alpha(a)e_{2k+1}]_-.$$

C. Complex projective spaces

It is well known that even-dimensional complex projective spaces have no spinor structure. In order to understand the difference between even and odd dimensions and to construct the spin structure in the latter case, it is convenient to consider the metaunitary group $\mathrm{MU}(n)$ (see Rf. 21) and find its center. This group may be defined as that subgroup of $\mathrm{Spin}(2n)$ that (doubly) covers the unitary group $\mathrm{U}(n)$ considered as a subgroup of $\mathrm{SO}(2n)$:

$$\begin{array}{ccc} \mathrm{MU}(n) & \xrightarrow{\text{inj}} & \mathrm{Spin}\ (2n) \\ \rho\downarrow & & \downarrow\rho\,. \\ \mathrm{U}(n) & \xrightarrow{\text{inj}} & \mathrm{SO}\ (2n) \end{array}$$

Let $(e_1,\ldots,e_{2n})$ be an orthonormal frame in $\mathbf{R}^{2n}$ embedded in the Clifford algebra $C^+(2n)$. Let $J\in\mathrm{SO}(2n)$, given by

$$J(e_\alpha) = \begin{cases} -e_{n+\alpha}, & \text{for } \alpha = 1,\ldots,n, \\ e_{\alpha-n}, & \text{for } \alpha = n+1,\ldots,2n, \end{cases}$$

define a complex structure in $\mathbf{R}^{2n}$ so that $\mathrm{U}(n) = \{a\in\mathrm{SO}(2n): J\circ a = a\circ J\}$. The center of $\mathrm{U}(n)$ is isomorphic to $\mathrm{U}(1)$ and consists of all elements of $\mathrm{SO}(2n)$ of the form $\cos 2t + J\sin 2t = \exp 2tJ$, $0\leqslant t<\pi$. Let

$$\iota = e_1e_{n+1} + \cdots + e_ne_{2n} \in \mathrm{spin}(2n),$$

then

$$\rho(\pm\exp t\iota) = \exp 2tJ.$$

Any element of $\mathrm{Spin}(2n)$ commuting with ι projects by ρ to an element of $\mathrm{SO}(2n)$ commuting with J. One can, therefore, define the metaunitary group as follows:

$$\mathrm{MU}(n) = \{s\in\mathrm{Spin}(2n): s\iota = \iota s\}.$$

Its Lie algebra is spanned by the set of n^2 elements

$$e_\alpha e_{n+\beta} + e_\beta e_{n+\alpha}, \quad 1\leqslant\alpha,\beta\leqslant n.$$

Any element of the center of $\mathrm{MU}(n)$ is of the form $\exp t\iota$ or $-\exp t\iota$ for some $t\in\mathbf{R}$. Since

$$\exp t\iota = (\cos t + e_1e_{n+1}\sin t)\cdots(\cos t + e_ne_{2n}\sin t),$$

one sees that $\exp\frac{1}{4}\pi\iota$ covers J and $\exp\frac{1}{2}\pi\iota = \epsilon$. Moreover,

$$\exp\pi\iota = \epsilon^2 = (-1)^n,$$

and the center of $\mathrm{MU}(n)$ is the set

$$\{\exp t\iota: 0\leqslant t<2\pi\} = \mathrm{U}(1), \quad \text{for } n \text{ odd},$$

and

$$\{\pm\exp t\iota: 0\leqslant t<\pi\} = \mathbf{Z}_2\times\mathrm{U}(1), \quad \text{for } n \text{ even}.$$

If n is odd, then the spinor structure on $\mathbf{C}P_n$ can be described as follows. Let $\mathrm{U}(1)$ be embedded in $\mathrm{MU}(n+1)$ so as to coincide with the connected component of the identity of its center,

$$\exp 2t\sqrt{-1} \mapsto \exp t\iota, \quad 0\leqslant t<\pi,$$

and put

$$S = \mathrm{MU}(n+1)/\mathrm{U}(1).$$

A right action of $\mathrm{MU}(n)$ in S is obtained by passing to the quotient with the action defined by the natural embedding $\mathrm{MU}(n)\to\mathrm{MU}(n+1)$. On quotienting, the double cover $\mathrm{MU}(n+1)\to\mathrm{U}(n+1)$ passes to a double cover of the unitary frame bundle E_n,

$$S\to\mathrm{U}(n+1)/\mathrm{U}(1) = \mathrm{SU}(n+1)/\mathbf{Z}_{n+1} = E_n,$$

and the action of $\mathrm{MU}(n)$ in S projects to the action of $\mathrm{U}(n)$ in $\mathrm{U}(n+1)/\mathrm{U}(1)$, as defined in Sec. III.

The nonexistence of a spinor structure in $\mathbf{C}P_{2k}$ results from $w_1 = 0$ and $w_2\neq 0$ for such a space. It also may be deduced directly from a comparison of the fundamental groups of the total space of the fibration $E_{2k}\to\mathbf{C}P_{2k}$ and of its fiber $\mathrm{U}(2k)$. We have indeed

$$\pi_1(\mathrm{U}(2k)) = \mathbf{Z}$$

and

$$\pi_1(E_{2k}) = \mathbf{Z}_{2k+1}.$$

The injection $\mathrm{U}(2k)\to E_{2k}$ defines a homomorphism $\mathbf{Z}\to\mathbf{Z}_{2k+1}$ such that $2k+1\mapsto 0 \bmod (2k+1)$. This contradicts the existence of a spinor structure.[22] It is known, however, that all complex projective spaces admit a natural spin^c-structure.[14,23] Recently, Robinson and Rawnsley[24] have shown that any symplectic manifold admits a complex metaplectic structure. The metaplectic structure on $\mathbf{C}P_{2k+1}$ gives rise to symplectic spinors.[25]

D. Quaternionic projective spaces

This is the simplest and easiest case: since $w_1 = 0$ and $w_2 = 0$ for $\mathbf{H}P_n$, $n = 1,2,\ldots$, any such space admits a unique spinor structure given by the sequence

$$S = \mathrm{Sp}(n+1)\to\mathrm{Sp}(n+1)/\mathbf{Z}_2 = F\to\mathbf{H}P_n.$$

The right action of $\Sigma = \mathrm{Sp}(n)\times\mathrm{Sp}(1)$ in S is obtained from the natural embedding. Incidentally, our considerations prove the existence of a natural monomorphism of groups

$$\mathrm{Sp}(n)\times\mathrm{Sp}(1)\to\mathrm{Spin}(4n), \tag{8}$$

which covers the injection $(\mathrm{Sp}(n)\times\mathrm{Sp}(1))/\mathbf{Z}_2\to\mathrm{SO}(4n)$.

V. CONCLUDING REMARKS

Most of the work on spinor structures is based on methods of algebraic topology and concentrates on problems of existence. Our approach is differential geometric and Lie-group theoretic. It yields an explicit construction of all spaces and maps occurring in the description of spinor structures on projective spaces. It can be extended to other homogeneous spaces, such as the Grassmannians, as well as to pseudo-Riemannian manifolds.

Besides the two coverings of the orthogonal group, which we have used in the case of real projective spaces, there are coverings not coming from the Clifford scheme. The analogous coverings—Clifford and not—can be defined also for the pseudo-orthogonal groups and related to the transformation properties of fermions under space-time reflections considered by physicists.[26] Our method can also be used to construct "extended spinor structures" such as the spin^c and complex metaplectic structures. It is clear from this work that, in the nonorientable case, the topological condition for the existence of a pin structure depends on which particular double cover of the orthogonal group is being considered. It would also be of some interest to study the spinor connections on projective and other homogeneous spaces. Stiefel bundles over Grassmannians, together with their canonical connections, are universal. Can one give a meaning to the idea of "universal spinor structures and connections"?

ACKNOWLEDGMENTS

We both have had the opportunity, on different occasions and not always together, to discuss the questions related to this paper with several mathematicians and physicists. We are particulary indebted for illuminating comments and helpful advice to M. Cahen, C. J. S. Clarke, A. Crumeyrolle, S. Gutt, R. Harvey, H. B. Lawson, Jr., R. Penrose, and J. H. Rawnsley.

One of us (A.T.) was supported by the National Science Foundation under Grant No. PHY-8306104.

[1] R. Penrose, Ann. Phys. (NY) **10**, 71 (1960); E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[2] A. Haefliger, C. R. Acad. Sci. Paris **243**, 558(1956); A Borel and F. Hirzebruch, Am. J. Math. **81**, 315(1959); J. Milnor, Enseign. Math. **9**, 198 (1963).
[3] J. Milnor, Enseign. Math. **8**, 16(1962); R. Stong, *Notes on Cobordism Theory* (Princeton U. P., Princeton, NJ, 1968).
[4] A. Lichnerowicz, C. R. Acad. Sci. Paris A-B **257**, 7(1963); N. Hitchin, Adv. Math. **14**, 1 (1974).
[5] M. F. Atiyah and I. M. Singer, Ann. Math. **87**, 546 (1968).
[6] R. Penrose, J. Math. Phys. **8**, 345 (1967); R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U. P., Cambridge, 1984 and 1986), Vols. I and II.
[7] C. J. Isham, Proc. R. Soc. London Ser. A **364**, 591 (1978); H. R. Petry, J. Math. Phys. **20**, 231 (1979).
[8] A. Trautman, "Differential geometry and spinors," unpublished lectures at SISSA, Trieste, 1984; P. Budinich and A. Trautman, "Remarks on pure spinors," SISSA, Trieste, preprint 87/85/EP, 1985.
[9] G. Landi, Lett. Math. Phys. **11**, 171 (1986).
[10] B. Grossman, T. K. Kephart, and J. D. Stasheff, Commun. Math. Phys. **96**, 431 (1984).
[11] J. Rawnsley, Warwick Univ. preprint 1985.
[12] S. S. Chern and J. Simons, Ann. Math. **99**, 48 (1974).
[13] G. S. Whiston, Gen. Relativ. Gravit. **6**, 463 (1975).
[14] M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3** (Suppl. 1), 3 (1964).
[15] M. F. Atiyah and R. Bott, Ann. Math. **88**, 451 (1968).
[16] M. Karoubi, Ann. Ec. Norm. **4**, 161 (1968).
[17] A. Trautman, *Differential Geometry for Physicists*, Stony Brook Lectures (Bibliopolis, Naples, 1984).
[18] G. E. Bredon, *Introduction to Compact Transformation Groups* (Academic, New York, 1972).
[19] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1951).
[20] C. J. S. Clarke, Gen. Relativ. Gravit. **2**, 43 (1971).
[21] R. J. Blattner, Am. Math. Soc. Proc. Symp. Pure Math. **26**, 147 (1973); V. Guillemin and S. Sternberg, *Geometric Asymptotics, Mathematical Surveys*, No. 14 (Am. Math. Soc., Providence, RI, 1977).
[22] R. Geroch, J. Math. Phys. **9**, 1739 (1968).
[23] S. W. Hawking and C. N. Pope, Phys. Lett. B **73**, 42 (1978).
[24] P. Robinson and J. Rawnsley, Warwick Univ. preprint, 1984.
[25] B. Kostant, Symp. Math. **14**, 139 (1974); A. Crumeyrolle, C. R. Acad. Sci. Paris A **282**, 583 (1976).
[26] S. Watanable, Rev. Mod. Phys. **27**, 40 (1955); Yu. M. Shirokov, Sov. Phys. JETP **11**, 101 (1960) [Zh. Eksp. Teor. Fiz. **38**, 140 (1960)].

Spontaneous compactification and Ricci-flat manifolds with torsion

Brett McInnes[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 12 December 1985; accepted for publication 26 March 1986)

The Freund–Rubin mechanism is based on the equation $R_{ik} = \lambda g_{ik}$ (where $\lambda > 0$), which, via Myers' theorem, implies "spontaneous" compactification. The difficulties connected with the cosmological constant in this approach can be resolved if torsion is introduced and λ is set equal to zero, but then compactification "by hand" is necessary since the equation $R_{ik} = 0$ can be satisfied both on compact and on noncompact manifolds. In this paper we discuss the global geometry of Ricci-flat manifolds with torsion, and suggest ways of restoring the "spontaneity" of the compactification.

I. WHAT IS "SPONTANEITY"?

The problem of understanding the relationship between gravitation and the other interactions is currently being addressed in a variety of ways—Kaluza–Klein, supergravity, superstrings, and so on.[1] The majority of such models either suggest or require that the universe be represented by a manifold of more than four dimensions, and so are apparently faced with an immediate conflict with the observational evidence. This problem is usually circumvented by assuming that (i) the ground state of the universe is a product of some simple space-time with a spacelike manifold M, and (ii) M is compact and has a very small diameter. (Noncompact models for M have also been proposed; for a general discussion and further references, see McInnes.[2,3] We shall not consider this possibility here.) This strategy is often loosely described as "spontaneous compactification," but, in fact, it is not at all evident that there is anything "spontaneous" in this procedure. Perhaps "compactification by hand" would be a more apt description.

In the particular case of 11-dimensional supergravity, however, the Freund–Rubin[4] mechanism does indeed inject a strong element of "spontaneity." If the generalized Maxwell field has nonzero vacuum expectation values $F_{\mu\nu\rho\sigma}$ proportional to the alternating tensor $e_{\mu\nu\rho\sigma}$ (Greek letters for space-time), then the Einstein equations imply that the internal space M satisfies $R_{ik} = \lambda g_{ik}$, with $\lambda > 0$. Now if we further assume that M is connected and complete (in the sense that every Cauchy sequence in M converges to a point in M), then Myers' theorem (see below) implies that M *must* be compact. Thus, the compactness of M need not be presumed, but rather can be deduced from certain natural assumptions regarding the nature of the ground state. This compactification is therefore genuinely "spontaneous."

Taking the Freund–Rubin mechanism as our prototype, we can now give a much more general characterization of "spontaneous compactification" in the true sense. The key ingredients may be listed as follows.

(*a*) *Assumption concerning the vacuum expectation values of the matter fields:* Intuitively, we expect the ground state matter configuration to be as simple as possible. For example, the fact that the Freund–Rubin energy-momentum tensor (in space-time) is covariantly constant seems reasonable, since this means that the "matter" in the ground state is uniformly distributed. Similar considerations motivate the assumption that the off-diagonal components $F_{\mu\nu\rho k}$, $F_{\mu\nu ik}$, $F_{\mu ijk}$ vanish. There is, admittedly, an element of vagueness in this procedure, but this is inevitable in the absence of a rigorous formulation of the concept of a "gravitational ground state."

(*b*) *Assumption concerning the geometry of the ground state:* It is essential to realize that the field equations alone *cannot* induce compactification, since they constrain the manifold only at a local level. Thus, for example, Freund and Rubin implicitly assume that the internal manifold M is a complete metric space, since otherwise Myers' theorem cannot be applied. Indeed, if M were not complete, it would necessarily be noncompact, independently of all other conditions. (This follows from the Hopf–Rinow theorem—see Cheeger and Ebin[5].) Therefore, any compactification scheme must assume completeness. This is quite reasonable, since otherwise the internal manifold will have "holes" or "rips," and we do not expect the ground state to display such pathologies. (This is particularly true in those theories in which the higher-dimensional geodesics have a direct physical interpretation. See McInnes[3] for a discussion.) We conclude, then, that in setting up a compactification mechanism we will have no option but to make *some* assumptions as to the global structure of the ground state. Obviously these should be as general as possible, and should be physically motivated; but, above all, they must not themselves automatically imply compactness, since this would be tantamount to compactification "by hand." In the Freund–Rubin case, the assumption that M is connected and complete does not, of course, imply that M is compact.

(*c*) *Gravitational field equations:* These allow us to translate the assumptions in (a) above into constraints on the geometry of M.

(*d*) *A "compactification theorem":* We must have a theorem that states that, under certain conditions compatible with (a)–(c) above, M is necessarily compact. This theorem should refer *only* to those aspects of the geometry that are controlled by the gravitational field equations. For example, the Einstein equations give full information only about the Ricci tensor, not the curvature tensor: fortunately,

[a)] Permanent address: Department of Mathematics, National University of Singapore, Kent Ridge 0511, Republic of Singapore.

however, the Myers "compactification theorem" only requires conditions on the Ricci tensor.

We have given this rather detailed and general formulation of spontaneous compactification in the hope that the technique can be extended beyond the specific model considered by Freund and Rubin. For, as is well known, that model encounters difficulties connected with the cosmological constant. Specifically, if the space-time cosmological constant is assumed to be very small, then the internal cosmological constant also becomes very small, and this is thought to be in conflict with the supposition that the internal manifold has a submicroscopic diameter. (Strictly speaking, this conclusion is unwarranted, because there does not seem to be any direct relationship between the "cosmological constant" of an Einstein manifold and its diameter, especially if it is multiply connected. Thus, for example, a sphere of given curvature has a totally different diameter to a real projective space of the same curvature. By means of a sufficiently large number of topological identifications, one might be able to reduce the diameter very substantially. But a topological structure of this level of complexity probably can be justified only in the context of quantum gravity.)

A solution of this problem, which is of course not peculiar to the Freund–Rubin mechanism, has been proposed by Orzalesi and collaborators (Destri *et al.*[6] and references therein), who propose to consider manifolds with torsion as models of internal space. Various generalizations have been proposed, notably the "seven-sphere with torsion" and the "squashed seven-sphere with torsion" (Duff *et al.*[7] and Wu[8]). In none of these cases is the compactification "spontaneous" in the sense in which we are using the term.

The principal purpose of this work is to consider the form that a genuine spontaneous compactification mechanism might take if the internal manifold is endowed with torsion. This will be done by relaxing some of the more restrictive assumptions made by Destri *et al.*[6] and then by proving a "compactification theorem" for the resulting class of manifolds. These manifolds have a vanishing (Riemann–Cartan) Ricci tensor. In the case of Riemannian manifolds, the vanishing of the Ricci tensor very severely restricts the isometry group, with serious consequences for Kaluza–Klein theories. We therefore explain in detail precisely why this does not, in general, occur if the torsion is nonzero. Finally, we consider the possibilities for generalizing or modifying our compactification scheme.

Notation: In general, we adhere strictly to the conventions of Kobayashi and Nomizu.[9] A Riemann–Cartan manifold or RC manifold is a manifold endowed with a positive-definite metric tensor g and a connection ∇ with $\nabla g = 0$. On such a manifold there are always (if the torsion $T \neq 0$) *two* fundamental connections, namely ∇ and the Levi-Cività connection $\mathring{\nabla}$ generated by g. Hence we need a dual notation (R is the curvature tensor for ∇, and $\mathring{R}$ that for $\mathring{\nabla}$) and also a dual nomenclature. We shall use the letters RC to indicate that we are referring to quantities generated by ∇, and the ordinary term for $\mathring{\nabla}$. Thus, for example, a tensor A will be called constant if $\mathring{\nabla} A = 0$, and "RC constant" if $\nabla A = 0$. Finally, note that we follow Kobayashi and Nomizu in using S to denote the Ricci tensor, but R_{ik} to denote its components. The symmetric part of S will be denoted (S).

II. COMPACTIFICATION WITH LIE GROUPS

A somewhat drastic solution of the cosmological constant problem in the Freund–Rubin framework would be to assume that *all* components of the vacuum expectation value of the tensor F vanish. The space-time cosmological constant is then precisely zero, and the internal space M satisfies $\mathring{R}_{ik} = 0$, which can be satisfied on certain compact manifolds.

There are two objections to this procedure, if we retain Riemannian geometry without torsion. In order to discuss the first, we need the following theorem (which unifies various results in Kobayashi and Nomizu[9]).

Theorem 1: Let M be a connected compact Ricci-flat Riemannian manifold. Then we have the following.

(i) The connected component of the identity of the isometry group $\mathrm{Isom}(M)$ is Abelian.

(ii) If $\dim(\mathrm{Isom}(M)) \geqslant \dim M$, then M is a flat manifold (and is therefore $\mathbb{R}^n/D$ for some discrete group D).

(iii) If the universal covering manifold of M is compact, then $\mathrm{Isom}(M)$ is a finite group.

The proof will be given later as a consequence of a more general result.

From the point of view of Kaluza–Klein theories, this result is disastrous, since it means that only an Abelian gauge group (at best) can be obtained. [This comment does not apply to superstring theories, but, even there, the absence of symmetries is a major technical impediment to explicit calculation with the metric. Note that the spaces considered by Candelas *et al.*[10] as vacuum configurations for superstrings are of the form (compact and simply connected)/(discrete) and therefore satisfy part (iii) of the above theorem.]

A second, and in our view, equally serious objection to internal manifolds with $\mathring{R}_{ik} = 0$ is that this equation can be satisfied both on compact *and* on complete noncompact manifolds. The compactification is therefore not spontaneous.

The introduction of torsion immediately resolves the first of these objections: for example, any semisimple compact Lie group can be endowed with a Riemann–Cartan connection which is RC Ricci-flat (in fact, RC flat), but obviously the symmetry group is not Abelian. Indeed, Destri *et al.*[6] propose just such a space as a model of M. In more detail, these authors assume (a) the Einstein–Cartan theory of gravitation,[11] where the field equations are (A, B, C, D being indices for the full multidimensional universe)

$$R_{AB} - \tfrac{1}{2} g_{AB} R = k E_{AB},$$
$$T^{A}{}_{BC} + \delta^{A}{}_{B} T^{D}{}_{CD} - \delta^{A}{}_{C} T^{D}{}_{BD} = k S^{A}{}_{BC},$$

where E_{AB} is the canonical (nonsymmetric) energy-momentum tensor, T^{A}_{BC} is the torsion, and S^{A}_{BC} is the spin tensor; (b) that the internal manifold M is a semisimple connected Lie group with a positive-definite Cartan–Killing metric; and (c) that the spin tensor S^{A}_{BC} has nonzero vacuum expectation values only in the internal space $S^{i}{}_{jr} = f^{i}{}_{jr}/k$ (where $f^{i}{}_{jk}$ are the structural constants in the appropriate basis), and the $E_{AB} = 0$.

The resulting ground state is a product of Minkowski space with a compact Lie group. However, it is clear that the compactification is not spontaneous. In particular, the formal condition (b)—which has no clear physical meaning—automatically entails the compactness of M: a well-known theorem of Weyl (see Kobayashi and Nomizu[12] for a geometric proof) states that any connected semisimple Lie group with a positive-definite Cartan–Killing is compact. Thus, the compactness of M is implicit in the geometric assumptions at the outset.

The difficulty with this mechanism, then, arises from the lack of a physical motivation for (b) and (c). The status of these assumptions can be greatly clarified with the aid of the following theorem, which simplifies and slightly modifies certain results of Hicks[13] and Wolf.[14]

Theorem 2: Let M be a complete, connected, simply connected Riemann–Cartan manifold such that (i) the curvature tensor $R=0$; and (ii) the "fully covariant" torsion tensor, defined by $T(X,Y,Z)=g(X,T(Y,Z))$, is totally antisymmetric.

Then M is a homogeneous (coset) space. If, in addition, we have (iii) T is constant, i.e., $\mathring{\nabla}T=0$, then M is a connected, simply connected Lie group.

Proof (outline): Using (i), let $\{e_j\}$ be a global basis of RC constant vector fields. From the definition of torsion we obtain, for each j, k, the equations $[e_j,e_k]=-T^r{}_{jk}e_r$. Now consider a vector field of the form a^ie_i, where the a^i are fixed numbers. Then $X=a^ie_i$ is a Killing vector field, since (if L denotes the Lie derivative)

$$(L_Xg)(e_j,e_k)=a^ie_ig_{jk}+g_{rk}T^r{}_{ij}a^i+g_{jr}T^r{}_{ik}a^i.$$

Here the second and third terms vanish by total antisymmetry, and the first term vanishes since $\nabla g=0$ implies $e_ig_{jk}=0$ in this basis. Furthermore, X is obviously of constant length, and its integral curves are therefore geodesics.

Now let x and y be arbitrary points in M. Then since M is complete, there exists (by the Hopf–Rinow theorem—see Cheeger and Ebin[5]) a minimizing geodesic joining x to y. (This curve is both a ∇ geodesic and $\mathring{\nabla}$ geodesic.) The tangent vector to this curve at x may be expressed as a linear combination $a^ie_i(x)$. Now define $X=a^ie_i$ as a vector field; then, as we have seen, the integral curves of X are geodesics. By the uniqueness of geodesics with given initial conditions, the geodesic joining x to y is an integral curve of X. But the isometry group G of M acts on M through motions along the integral curves of the Killing fields: hence G maps x into y. Since x and y are arbitrary, G acts transitively, and so M must be a homogeneous space G/H.

Now suppose that T is constant. Choosing the e_j to be orthonormal, one finds that the components of the Levi-Cività connection in this basis are $\mathring{\Gamma}^i{}_{jk}=-\frac{1}{2}T^i{}_{jk}$, and so $\mathring{\nabla}T=0$ yields

$$e_iT^j{}_{km}-\tfrac{1}{2}T^r{}_{km}T^j{}_{ir}+\tfrac{1}{2}T^j{}_{rm}T^r{}_{ik}+\tfrac{1}{2}T^j{}_{kr}T^r{}_{im}=0,$$

or

$$e_iT^j{}_{km}+\tfrac{1}{2}T^j{}_{ir}T^r{}_{mk}+\tfrac{1}{2}T^j{}_{mr}T^r{}_{ki}+\tfrac{1}{2}T^j{}_{kr}T^r{}_{im}=0.$$

Permuting twice on ikm and adding, we find

$$e_iT^j{}_{km}+e_kT^j{}_{mi}+e_mT^j{}_{ik}+\tfrac{3}{2}(T^j{}_{ir}T^r{}_{mk}+T^j{}_{mr}T^r{}_{ki}+T^j{}_{kr}T^r{}_{im})=0.$$

But the Jacobi identities corresponding to the basis $\{e_i\}$ are

$$e_iT^j{}_{km}+e_kT^j{}_{mi}+e_mT^j{}_{ik}+(T^j{}_{ir}T^r{}_{mk}+T^j{}_{mr}T^r{}_{ki}+T^j{}_{kr}T^r{}_{im})=0.$$

Substracting these two equations, we find

$$T^j{}_{ir}T^r{}_{mk}+T^j{}_{mr}T^r{}_{ki}+T^j{}_{kr}T^r{}_{im}=0$$

and so $e_iT^j{}_{km}=0$, that is, the $T^j{}_{km}$ are constants. In view of $[e_ie_j]=-T^r_{ij}e_r$, it is evident that the $\{e_i\}$ generate a Lie algebra. Let G be the corresponding simply connected, connected Lie group. Then, G acts on M through motions along the integral curves of the vector fields a^ie_i. Arguing as before, we find $M=G/H$. But now we have (since the Lie algebra of G is generated by the $\{e_i\}$) the relations $\dim G=\dim M=\dim G-\dim H$, whence H must be discrete. Since M is simply connected, H must be trivial, and so $M=G$. This concludes the proof.

Remark 1: If M is not assumed to be simply connected, then one finds (by applying the above argument to the universal covering space of M) that M has the form G/D, where G is a connected, simply connected Lie group, and D is a discrete subgroup.

Remark 2: It is easy to see that the given connection on M coincides with standard $(-)$ connection on G. But the given metric on M may not coincide with the Cartan–Killing metric on G—indeed G may not be semisimple. See the remarks after the proof of Proposition 1 in the next section.

The great virtue of this theorem from our present point of view is that we can now do away with the assumption that M is a Lie group. Instead we can make assumptions regarding the torsion and RC curvature of M, and it may be possible to motivate such assumptions in a physical way through the field equations. Thus, assumption (b) of the mechanism of Destri *et al.* can be reformulated as follows.

(b′) The internal manifold M is a complete, connected, simply connected Riemann–Cartan manifold, which satisfies (i) $R^i{}_{jkl}=0$; (ii) T_{ijk} is totally antisymmetric; (iii) T_{ijk} is constant (thus M is a Lie group); and (iv) the metric on M is the Cartan–Killing metric.

(Note that the fourth assumption is independent of the others—the Cartan–Killing metric is not the only possible metric on a Lie group.) Now let us consider the justification for these assumptions. Since the second Einstein–Cartan field equation relates the torsion to the spin tensor by an invertible algebraic equation, (iii) means that the vacuum expectation value of the spin tensor S_{ijk} is constant: in other words, the spin source is uniformly distributed throughout M. This appears to be quite reasonable for a ground state—indeed, as we have already mentioned, one reason for accepting anti-de Sitter space as a space-time ground state is that the "matter" is uniformly distributed throughout the space (that is, the cosmological term is constant). Similarly, (ii) means that the spin tensor is taken to be totally antisymmetric. Destri *et al.*[6] justify this on the grounds that the simplest source of spin—the Dirac field—does indeed generate a totally antisymmetric spin tensor. Thus a spin tensor of this type is naturally associated with spin-$\frac{1}{2}$ condensates. While this is plausible, in practice one might also be interested in Rarita–Schwinger condensates, for which the spin tensor need not be totally antisymmetric.[8] However, while (ii) is less well motivated than (iii), it does at least have a physical interpretation.

The same cannot, unfortunately, be said of (i), and still less of (iv). Assumption (i) can be partially justified on the grounds that it implies the vanishing of the internal energy-momentum tensor—again, this is a reasonable condition for the ground state to satisfy—but is obviously unnecessarily strong, since the vanishing of the RC Ricci tensor would suffice for this. Finally, (iv) appears to have no physical meaning whatever; again, since M is spacelike, this assumption amounts (via the theorem of Weyl mentioned earlier) to compactification by hand.

This analysis clearly indicates that some of these assumptions must be relaxed or replaced if we are to obtain a genuinely spontaneous compactification mechanism. Such modifications will lead us to spaces that are not necessarily Lie groups, but this is desirable: in particular, one would hope that (at least some) homogeneous (coset) manifolds can be included. A specific procedure will be proposed in the next section.

We conclude this section with a brief discussion of the "seven-sphere with torsion" solutions. As is well known, the round seven-sphere with torsion[15] satisfies the first two assumptions listed under (b′) above, but, since it is not a Lie group, it cannot satisfy the third. The squashed seven-sphere with torsion violates (i) and (iii) but satisfies (ii) (see Duff *et al.*[7]). Finally, Wu[8] proposes to retain (i) but neither (ii) nor, in general, (iii). We shall not consider these models further, since in each case the topology is assumed at the outset to be that of the seven-sphere.

III. A FRAMEWORK FOR SPONTANEOUS COMPACTIFICATION WITH TORSION

We now propose to extend the concept of spontaneous compactification to spaces with torsion by generalizing the mechanism of Destri *et al.*[6] In this paper we shall not propose a specific model, but rather a general framework to guide the construction of such models. We proceed according to the principles laid down in the first section, and list our assumptions in the same way.

A. Assumptions concerning the vacuum expectation values

(i) We shall retain the following assumptions of Destri *et al.* First we take the canonical energy-momentum tensor E_{AB} to be zero. (Thus R_{AB} is zero and the problem of the cosmological constant does not arise.) Second, we assume that the spin tensor has nonvanishing components S_{ijk} only in the internal direction. (Thus space-time may be taken as Minkowski space.) Third, we assume that the "spin condensate" is uniformly distributed throughout the internal ground state manifold M—that is, we take S_{ijk} to be a covariant constant tensor. We do *not* assume that the RC curvature tensor vanishes.

(ii) We shall not assume that the tensor S_{ijk} is totally antisymmetric; however, we will be forced to assume that (in a certain technical sense to be clarified later) the totally antisymmetric part gives the dominant contribution to S_{ijk}.

B. Assumption concerning the geometry of the ground state

Beyond the standard assumptions that the universe splits into a product of space-time with a connected, complete, spacelike manifold M, we wish to impose the following additional condition: *the universe splits only once*. That is, M itself must not split into a product. Let us be more precise about this. Let N be any Riemannian manifold, and x be an arbitrary point in N. Suppose that there exist two submanifolds N' and N'' of N and an open neighborhood U of x such that U is the Riemannian product of U' with U'', where U' is an open neighborhood of x in N', and U'' is an open neighborhood of x in N''. Then we shall say that N is "locally decomposable." More simply, a locally decomposable manifold is one in which, at each point, it is possible to find a coordinate system such that the metric has block-diagonal form throughout a neighborhood around that point—in short, a locally decomposable manifold is what is usually (but not quite correctly) described in the physics literature as a "product manifold." We propose to forbid such manifolds as candidates for M.

There are two major reasons for imposing this condition. The first is that, in general, a locally decomposable manifold has an isometry group that is not simple. In the Kaluza–Klein context, this would lead to a nonsimple gauge group with (possibly) different gauge couplings, and so unification would be lost even among the nongravitational interactions (unless the Candelas–Weinberg[16] method can be applied, which seems somewhat problematic). It is true that the presence of torsion complicates this picture; in principle, one could use the torsion to break completely all of the factors in the gauge group save one. But this is clearly very unnatural.

The second reason for requiring local indecomposability is more vague but also more deep. Throughout this paper, we have assumed as usual that the ground state of the universe is a (pseudo-) Riemannian product $W = L \times M$, where L is space-time. Why should W split in this way? The answer is of course unknown, but it is significant that W should split into a compact part and a part that *cannot* be compact. As is well known, compact space-times violate causality.[17] Thus we may speculate that space-time splits off from W in order that causality be preserved. (Freund and Rubin[4] have attempted to explain why precisely four dimensions split off.) The fact that the Myers compactification theorem is not valid for pseudo-Riemannian manifolds[18] reinforces this idea. The point is that if the existence of time is indeed responsible for the splitting of W, then M, being entirely spacelike, should not split.

To summarize, then, we assume that the internal manifold M is a spacelike, connected, complete, locally indecomposable manifold. These conditions in themselves are quite general and certainly do not imply compactness.

C. Gravitational field equations

We use the Einstein–Cartan theory.

D. A "compactification theorem"

Before stating the appropriate result, we must use the gravitational field equations to express the assumptions concerning the vacuum expectation values in geometrical form. First, $E_{AB} = 0$ implies that the RC Ricci tensor is zero. Second, the assumption that S_{ijk} is constant means that the torsion T is constant. The idea that the "spin-$\frac{1}{2}$-like" contribution to S_{ijk} is dominant can be expressed as follows. (As we shall see, our torsion has to be traceless, and so T_{ijk} and S_{ijk} essentially coincide; therefore we refer directly to T_{ijk}.) The tensor T_{ijk} corresponds to reducible representation of the appropriate rotation group, but can be split into two "irreducible" parts (not three, because T is traceless) according to

$$T_{ijk} = T_{[ijk]} + U_{ijk}, \tag{3.1}$$

where $T_{[ijk]}$ is the totally antisymmetric part and

$$U_{ijk} = \tfrac{2}{3}(T_{ijk} - T_{[jk]i}). \tag{3.2}$$

The deviation of T from total antisymmetry can be measured by the quantity

$$\theta = U_{ijk} U^{ijk} / T_{[ijk]} T^{[ijk]}. \tag{3.3}$$

Since T is constant, θ is a fixed number that is zero if and only if T is totally antisymmetric. The statement that the totally antisymmetric part of T is dominant simply means that θ is small.

We are now in a position to state and prove our main result. [We denote the symmetric part of the RC Ricci tensor by (S).]

Proposition 1: Let M be a complete, connected, locally indecomposable Riemann–Cartan manifold with constant nonzero torsion. If $(S) = 0$ and $\theta < \frac{1}{2}$, M must be compact.

Corollary: Let M be a complete, connected, locally indecomposable Riemann–Cartan manifold with constant, nonzero, totally antisymmetric torsion. If $(S) = 0$, then M must be compact.

This proposition has the following partial converse.

Proposition 2: Let M be a complete, connected, locally indecomposable Riemann–Cartan manifold with constant nonzero torsion, and with a continuous non-Abelian isometry group. If $(S) = 0$, then M is compact if and only if $\theta < \frac{1}{2}$.

These results will be proved with the aid of the following definitions and theorems. (See Kobayashi and Nomizu.[9,12]) Let N be any Riemannian manifold (assumed to be connected, but not necessarily complete). The holonomy group acts (via parallel transport around closed loops) as a group of linear transformations on the tangent space T_x at any point x. Thus T_x yields a representation of the holonomy group. If this representation is reducible, then N is said to be a reducible manifold. Similarly, if the restricted holonomy group (obtained by parallel transport around null-homotopic loops) acts reducibly, then we shall say that N is locally reducible. We now have the following theorem.

Theorem 3 (local de Rham decomposition theorem): Every connected locally reducible Riemannian manifold is locally decomposable.

Theorem 4 (global de Rham decomposition theorem): Every connected, simply connected, complete, reducible Riemannian manifold is globally decomposable, that is, globally isometric to a Riemannian product.

We shall also use the following results.

Theorem 5 (Myers): Let N be a connected, complete Riemannian manifold. If all eigenvalues of the Ricci tensor are bounded from below by a strictly positive constant, then N must be compact.

Theorem 6 (Bochner): Let N be a compact, connected Riemannian manifold, with a negative-definite Ricci tensor. Then the isometry group of N is finite.

Theorem 7 (Schur): Let G be a subgroup of the orthogonal group $O(n)$ that acts irreducibly on $\mathbb{R}^n$. Then every symmetric bilinear form on $\mathbb{R}^n$ that is invariant by G is a multiple of the standard Euclidean inner product.

We may now prove Propositions 1 and 2.

Proof of Proposition 1: Define a (1,2) tensor on M by $K(X,Y) = \nabla_X Y - \mathring{\nabla}_X Y$, where X and Y are vector fields. In components with respect to an arbitrary basis, K is

$$K^i{}_{jl} = -\tfrac{1}{2} g^{ir}(T_{jlr} + T_{ljr}) + \tfrac{1}{2} T^i{}_{jl}. \tag{3.4}$$

Note that $K_{ijl} = g_{ir} K^r{}_{jl}$ is antisymmetric in its first and third indices. A straightforward computation yields

$$R_{ij} = \mathring{R}_{ij} + K^n{}_{nm} K^m{}_{ji} - K^n{}_{jm} K^m{}_{ni} + \mathring{\nabla}_n K^n{}_{ji} - \mathring{\nabla}_j K^n{}_{ni} \tag{3.5}$$

as the general relation between R_{ij} and $\mathring{R}_{ij}$. In the present case, the last two terms on the right-hand side vanish, as also does the second—for the following reason. Since M is locally indecomposable, it must also (by Theorem 3) be locally irreducible. Now T and (therefore) T^m_{mi} are constant with respect to $\mathring{\nabla}$ and so T^m_{mi} is an invariant vector under the action of the restricted holonomy group. Hence T^m_{mi} is zero, that is, T is traceless. But $K^m_{mi} = T^m_{mi}$, and so only the first and third terms in (3.5) are nonzero. Thus, setting the symmetric part of R_{ij} equal to zero, we find

$$\mathring{R}_{ij} = \tfrac{1}{2} K^m{}_{ni} K^n{}_{jm} + \tfrac{1}{2} K^m{}_{nj} K^n{}_{im}. \tag{3.6}$$

Clearly $\mathring{\nabla}_p \mathring{R}_{ij} = 0$, and so $\mathring{R}_{ij}$ is a symmetric bilinear form invariant under the action of the restricted holonomy group. As we have seen, this action is irreducible, and so by Theorem 7 we have

$$\mathring{R}_{ij} = c g_{ij}. \tag{3.7}$$

Substituting this into (3.6), contracting and setting $n = \dim M$, we obtain

$$cn = K_{mni} K^{nim}.$$

Using the antisymmetry of K and the easily derived relations

$$K^{nim} = \tfrac{1}{2} T^{inm} + \tfrac{1}{2}(T^{nim} - T^{min}),$$

$$\tfrac{1}{2}(K_{mni} - K_{nmi}) = \tfrac{1}{2} T_{imn},$$

one obtains

$$cn = \tfrac{1}{4} T_{imn} T^{inm} + \tfrac{1}{2} T_{imn} T^{nim}.$$

Setting $T_{imn} = T_{[imn]} + U_{imn}$ and using the relations $T_{imn} U^{imn} = 0$ and $\frac{1}{2}(U_{imn} - U_{min}) = \frac{1}{2} U_{nmi}$, one finds

$$cn = \tfrac{1}{4} T_{[imn]} T^{[imn]} - \tfrac{1}{2} U_{imn} U^{imn}. \tag{3.8}$$

From the definition of θ, assuming $T_{[imn]}$ is not zero, we have finally

$$cn = \tfrac{1}{2}(\tfrac{1}{2} - \theta) T_{[imn]} T^{[imn]}. \tag{3.9}$$

Thus $c > 0$ if $\theta < \frac{1}{2}$. Since M is connected and complete, it now follows from (3.7) and Theorem 5 that M is compact. This completes the proof.

Remark 3: In our general discussion, we assume that the full Ricci tensor vanishes, but clearly this assumption can be weakened to $(S) = 0$. However, this has no clear physical meaning.

Remark 4: Combining Theorem 2 with Proposition 1, we can show that any complete, connected, simply connected; locally indecomposable RC manifold that satisfies the three conditions of Theorem 2 must be a simply connected compact Lie group. Furthermore, the given connection coincides with the $(-)$ connection, and the given metric coincides with the Cartan–Killing metric. [This follows from (3.6), which becomes $cg_{ij} = \frac{1}{4}T^{m}_{ni}T^{n}_{jm}$. Here T^{m}_{ni} are essentially the structural constants.] Thus, the group must also be semisimple in this case.

We now prove Proposition 2: The proof is precisely as above, except that we now must also show that M is noncompact if $\theta \geqslant \frac{1}{2}$. In this case $c \leqslant 0$ and so if M were compact we could use either Theorem 6 (if $\theta > \frac{1}{2}$) or Theorem 1 (if $\theta = \frac{1}{2}$) to produce a contradiction. This completes the proof.

Remark 5: Proposition 2 fails without the assumption that the isometry group is continuous and non-Abelian. In Kaluza–Klein theories we must in practice have such an isometry group, and so in this context the condition $\theta < \frac{1}{2}$ is not only sufficient but also necessary for compactification.

Remark 6: If $T_{[ijk]} = 0$, then θ is undefined, but it is clear from (3.8) that $c < 0$ in this case; therefore, we can include this case in the statements of Propositions 1 and 2 by formally allowing θ to be infinite.

With Proposition 1 we conclude our general description of spontaneous compactification for spaces with torsion. The problem of constructing nontrivial particular examples will be considered in the next section.

IV. SYMMETRIES OF MANIFOLDS WITH TORSION

As Theorem 1 clearly shows, the vanishing of the Ricci tensor of a compact Riemannian manifold strongly restricts the symmetry group. The example of Lie groups shows that the restrictions are less severe in the Riemann–Cartan case. But this tells us nothing about the more interesting case in which the RC Ricci tensor vanishes but the RC curvature does not. Part (ii) of Theorem 1 means that, in the Riemannian case, such a manifold is less symmetric than its flat counterparts. It is important to determine whether this is so for RC manifolds.

In this section we shall show that, unless it is supplemented by some quite unnatural technical conditions, RC Ricci-flatness imposes only a very weak condition on the symmetry group of a compact RC manifold. It is even conceivable that some Riemannian coset manifolds can be "Ricci-flattened" by torsion without losing any symmetry, though we have no examples as yet.

Let Aff(M) be the identity component of the group of affine automorphisms of a RC manifold M: that is, Aff(M) consists of mappings of M into itself, which preserve the affine connection. In the case of compact Riemannian manifolds, Aff(M) coincides with the identity component of the isometry group, denoted Isom(M), but for a general RC manifold this is not so. A vector field X on M that generates local isometries will be called a metric Killing vector, while a field that generates local affine automorphisms will be called an affine Killing vector. The corresponding algebras can, under certain natural conditions,[9] be identified with the Lie algebras of Isom(M) and Aff(M). The following simple result specifies the relationships between these various objects.

Proposition 3: A vector field X on a compact Riemann–Cartan manifold is an affine Killing vector if and only if it is a metric Killing vector and also the Lie derivative $L_x T$ is zero.

Proof: Let K be the tensor defined in the proof of Proposition 1. Clearly $L_x T = 0$ if and only if $L_x K = 0$. Now it may be shown that X is an affine Killing vector if and only if, for every vector field Y, $L_x \nabla_y - \nabla_y L_x = \nabla_{[xy]}$. Let X, Y, Z be arbitrary vector fields. Then using $\nabla_y Z = \mathring{\nabla}_y Z + K(Y,Z)$ one finds

$$
\begin{aligned}
&L_x \nabla_y Z - \nabla_y L_x Z - \nabla_{[xy]} Z \\
&\quad = L_x \mathring{\nabla}_y Z - \mathring{\nabla}_y L_x Z - \mathring{\nabla}_{[xy]} Z + (L_x K)(Y,Z). \quad (4.1)
\end{aligned}
$$

Thus it is clear that if X is a metric Killing field and $L_x T = 0$, then X is also an affine Killing field. Conversely, if X is an affine Killing field, the left-hand side vanishes. Exchanging Y and Z and subtracting, we find

$$
\begin{aligned}
&L_x (\mathring{\nabla}_y Z - \mathring{\nabla}_z Y) + (\mathring{\nabla}_z L_x Y - \mathring{\nabla}_{[xy]} Z) \\
&\quad + (\mathring{\nabla}_{[xz]} Y - \mathring{\nabla}_y L_x Z) + (L_x T)(Y,Z) = 0.
\end{aligned}
$$

Since $\mathring{\nabla}$ is torsionless, we have

$$
\begin{aligned}
&[X,[Y,Z]] + [Z,[X,Y]] \\
&\quad + [[X,Z],Y] + (L_x T)(Y,Z) = 0,
\end{aligned}
$$

and so, by the Jacobi identities, $L_x T = 0$. Substituting this into (4.1) we find that X is a metric Killing field. This completes the proof.

Corollary: Aff(M) is a subgroup of Isom(M).

In physical language, one would say that, unless the torsion is invariant by the isometry group, it breaks the symmetry from Isom(M) down to Aff(M). Thus, the "symmetry group" is Aff(M), not Isom(M).

In order to proceed, we introduce the following notation. For any vector field X, let $A_x = L_x - \nabla_x$. Since A_x annihilates any function, it may be treated as a $(1,1)$ tensor. If X is a metric Killing vector, then A_x, regarded as a $(0,2)$ tensor, is antisymmetric. For any vector fields X, Y, one has $A_x Y = -T_x Y - \nabla_y X$, where T_x is the $(1,1)$ tensor defined by $T_x(Z) = T(X,Z)$. It is also possible to show that if X is an affine Killing vector and Y is arbitrary, then $\nabla_y(A_x) = R(X,Y)$, where R is the RC curvature tensor.[9]

The following result is a direct generalization of Theorem 1.

Proposition 4: Let M be a compact connected RC manifold with traceless torsion and RC Ricci tensor equal to zero. Suppose either that every affine Killing vector satisfies

(i) Trace $A_x T_x \leqslant 0$ everywhere on M,

or that every affine Killing vector satisfies

(ii) $T_x = bA_x$ everywhere on M, where b is a constant.

Then (a) $[X,Y] = \pm T(X,Y)$ for every pair X, Y, of affine Killing fields, where the $(-)$ sign occurs only in case (ii) and only if $b = -1$; and (b) if dim Aff$(M) \geqslant$ dim M and T

is totally antisymmetric, then M has the form G/D, where G is a simply connected Lie group and D is a discrete subgroup.

Proof (outline): Let X be an affine Killing vector field, and let Y, Z be arbitrary vector fields. Then

$$(\nabla_y A_x)(Z) = \nabla_y(A_x(Z)) - A_x(-A_z Y - T_z Y)$$
$$= R(X,Y)Z = -R(Y,X)Z.$$

By definition, the Ricci tensor S is given as the trace $S(X,Z)$ of the map $Y \to R(Y,X)Z$, and so

$$-S(X,Z) = \operatorname{div}(A_x Z) + \operatorname{Trace} A_x A_z + \operatorname{Trace} A_x T_z. \tag{4.2}$$

Now Stokes' theorem for a compact orientable RC manifold takes the form

$$\int \operatorname{div} W\,dv = -\int \operatorname{Trace} T_W\,dv,$$

where W is any vector field and dv is the volume element. In our case T_W is traceless, and so, setting $S=0$ in (4.2) and integrating, we obtain

$$\int (\operatorname{Trace} A_x A_z + \operatorname{Trace} A_x T_z)\,dv = 0. \tag{4.3}$$

(If M is not orientable, we can take the appropriate twofold covering, without altering our conclusions.)

Now assume conditions (i). Set $Z = X$ in (4.3). Since A_x is antisymmetric, $\operatorname{Trace} A_x A_x \leqslant 0$ at each point, and since the same is true of $\operatorname{Trace} A_x T_x$, we must have $\operatorname{Trace} A_x A_x = 0$ and therefore $A_x = 0$ at each point. Thus if X and Y are affine Killing fields, we have $T_x Y + \nabla_y X = 0$. Exchanging X and Y and subtracting, one finds (from the definition of T) that $[X,Y] = T(X,Y)$.

Now if $\dim \operatorname{Aff}(M) \geqslant \dim M$, then at any point one can set up a basis of affine Killing vectors $\{X_i\}$. We have just seen that $A_x = 0$ for an affine Killing field, and so $R(X_i,X_j) = \nabla_{xj} A_{xi} = 0$. Thus M is RC flat. Now from Proposition 3 and the relation $[X_i,X_j] = T(X_i,X_j)$, one finds

$$[X_i, T(X_j,X_k)] = T(T^r{}_{ij} X_r, X_k) + T(X_j, T^r{}_{ik} X_r),$$

which, after simplification, becomes

$$X_i T^m{}_{jk} = T^m{}_{ir} T^r{}_{kj} + T^m{}_{kr} T^r{}_{ji} + T^m{}_{jr} T^r{}_{ik}. \tag{4.4}$$

Permuting twice on ijk and adding, one obtains an equation that can be compared with the Jacobi identities for the basis $\{X_i\}$. The result is that both sides of (4.4) vanish. Now using $L_x g = 0$ and the assumed total antisymmetry of T, one shows that $X_i g_{jk} = 0$ in this basis and so, since the commutator coefficients are $T^i{}_{jk}$, $\mathring{\Gamma}^i{}_{jk} = \frac{1}{2} T^i{}_{jk}$. Hence

$$\mathring{\nabla}_i T^j{}_{kl} = X_i T^j{}_{kl} + \tfrac{1}{2} T^r{}_{kl} T^j{}_{ir} - \tfrac{1}{2} T^j{}_{rl} T^r{}_{ik} - \tfrac{1}{2} T^j{}_{kr} T^r{}_{il},$$

which is zero since both sides of (4.4) vanish. Finally, M must be complete because it is compact. Thus, by Theorem 2, M is essentially a Lie group.

The proof is similar in case (ii). Putting $Z = X$ in (4.3), as well as $T_x = bA_x$, one has

$$(1+b)\int \operatorname{Trace} A_x A_x\,dv = 0,$$

whence $A_x = 0$ as above, unless $b = -1$. But in that case $\nabla_y X = -(A_x + T_x)Y = 0$ for every affine Killing vector X and arbitrary Y. Thus every affine Killing vector is RC constant, and so, from the definition of torsion, every pair X, Y of affine Killing fields satisfies $[X,Y] = -T(X,Y)$. Except for unimportant details, the proof that M is RC flat and is essentially a Lie group now proceeds as before. This concludes the proof.

Remark 7: Any compact connected Riemannian manifold obviously satisfies the conditions of this proposition if its Ricci tensor is zero, and so one obtains the first two parts of Theorem 1 by setting $T=0$ in Proposition 4. The third part of Theorem 1 is obtained as follows. It is clear from the above proof that for such a manifold, every Killing vector satisfies $A_x = 0$, and so $\nabla_y X = \mathring{\nabla}_y X = 0$. Now let $\widetilde{M}$ be the universal covering manifold of M, which inherits its local geometry from M. Then the algebra of Killing vectors is invariant under the action of the holonomy group, and so $\widetilde{M}$ is reducible and consequently splits, according to Theorem 4. The Killing vectors generate a flat simply connected manifold. But such a manifold is noncompact, which is impossible if $\widetilde{M}$ is compact. Hence, there can in fact be no Killing vectors, and so the Lie algebra of the isometry group is trivial and the isometry group is discrete. As M is compact, so also is its isometry group, which must therefore be not only discrete but actually finite.

Remark 8: Any compact semisimple Lie group satisfies the conditions of Proposition 4. The $(+)$ connection corresponds to case (i), and the $(-)$ connection to case (ii) (with $b = -1$).

Apart from its general interest, Proposition 4 is mainly of interest to us because it shows that, in general, the vanishing of the Ricci tensor imposes a remarkably weak condition on the symmetry group of a Riemann–Cartan manifold. The restrictions of Theorem 1 are so strong in the Riemannian case simply because these manifolds "accidentally" satisfy condition (i) of Proposition 4. But for a general Riemann–Cartan manifold with zero Ricci tensor, there is no reason whatever to expect that either (i) or (ii) will hold. In this case the only restriction is Eq. (4.3), which, being an integral equation, is a weak constraint. We conclude, therefore, that the Ricci-flatness condition is unlikely to restrict the symmetry of a Riemann–Cartan manifold to any significant extent.

The existence of symmetry is of great value in constructing explicit examples of manifolds. In the present context, homogeneous (coset) manifolds are of particular interest. The vast majority of such manifolds do not admit RC flat connections, but it is certainly possible that many may admit RC Ricci-flat connections. In view of our assumption (in Propositions 1 and 2) that $\mathring{\nabla} T = 0$, the following result suggests one approach. (Here we use the term "symmetric space" in the technical sense; see Kobayashi and Nomizu.[12])

Proposition 5: Let $M = G/H$ be a Riemannian symmetric space with a G-invariant metric g. Then if T is the torsion of any G-invariant Riemann–Cartan connection on M, we have $\mathring{\nabla} T = 0$, where $\mathring{\nabla}$ is the Levi-Cività connection for g.

Proof: Since the RC connection is G-invariant, it follows from Proposition 3 that T is G-invariant. But standard results on symmetric spaces state (i) that the Levi-Cività connection induced by a G-invariant metric coincides with the canonical connection, and (ii) that any G-invariant tensor is constant with respect to the canonical connection. Hence $\mathring{\nabla}T = 0$, which completes the proof.

Remark 9: This result could be regarded as further motivation for the assumption in Proposition 1 that T is constant.

Proposition 5 suggests that examples of compact manifolds compatible with Proposition 1 may possibly be found by substituting the metric of a symmetric space (in particular, a space of constant curvature) into the left-hand side of (3.6) and solving for $K^i{}_{jk}$ subject to the constraint $\theta < \frac{1}{2}$. One hopes that nontrivial solutions (with nonzero RC curvature) can be found in this way.

V. MODIFICATIONS OF THE COMPACTIFICATION THEOREM

As several of the known examples of internal manifolds with torsion do not satisfy all conditions of Proposition 1, it is of some interest to ask whether these conditions can be modified or dropped. Here we list briefly some relevant remarks.

First, note that it is not possible to remove either completeness or local indecomposability. Without the first, the manifold would necessarily be noncompact. Without the second, we would be including manifolds of the type $M \times \mathbb{R}^n$, where M satisfies all conditions of Proposition 1. This manifold satisfies all conditions of Proposition 1 except local indecomposability, and fails to be compact.

Second, note that both the round and the squashed seven-spheres with torsion have $\mathring{\nabla}T$ totally antisymmetric but nonzero. Although the physical motivation is not clear, one may ask whether the condition $\mathring{\nabla}T = 0$ in Proposition 1 can be relaxed to total antisymmetry (on all four indices) for $\mathring{\nabla}T$. Interestingly, the answer depends on the solution to a problem in pure mathematics which, to the author's knowledge, remains unresolved. Taking both T and $\mathring{\nabla}T$ to be totally antisymmetric in equation (3.5), and setting $R_{ij} = 0$, one obtains

$$\mathring{R}_{ij} = \tfrac{1}{4}T^m{}_{ni}T^n{}_{jm}.$$

Clearly $\mathring{R}_{ij}$ is non-negative, but this certainly does not imply compactification. (Myers' theorem requires that the eigenvalues be positive and *bounded away from zero.*) However, if one calculates the gradient of the scalar curvature, it is found (by judicious use of the antisymmetry properties) that $\mathring{\nabla}_i\mathring{R} = -2\mathring{\nabla}^j\mathring{R}_{ji}$. The Bianchi identities then imply that $\mathring{R}$ is a positive constant. It is apparently unknown at present whether it is possible for a noncompact complete manifold with non-negative Ricci tensor to have a constant positive scalar curvature, assuming local indecomposability of course. This is related to an extension of the well-known Yamabe conjecture.[19]

Finally, one may wish to consider replacing $\mathring{\nabla}T = 0$ by equation $\nabla T = 0$. The consequences of this can be explored as follows. Let M be a Riemann–Cartan manifold that is RC locally reducible—that is, the restricted holonomy group of ∇ (not $\mathring{\nabla}$) acts reducibly. Let $x \in M$ and let U be an open neighborhood of x. Let H'_x be an invariant subspace of the tangent space at x, and let H''_x be the orthogonal complement of H'_x. Then H''_x is also invariant, and in fact this splitting of the tangent spaces can be extended throughout U in a consistent and continuous way. We shall say that the torsion T *splits holonomically* if, for each x, $T(H'_x,H'_x) \subseteq H'_x$, $T(H''_x,H''_x) \subseteq H''_x$, and $T(H'_x,H''_x) = 0$. The following result now generalizes the local de Rham decomposition theorem to Riemann–Cartan manifolds, and should be compared with Theorem 3. (The proof is given in the Appendix.)

Proposition 6 (local de Rham for RC manifolds): Let M be a connected, RC locally reducible Riemann–Cartan manifold such that the torsion splits holonomically. Then M is locally decomposable, regarded as a Riemannian manifold. If in addition $\nabla T = 0$, then M is locally decomposable, regarded as an RC manifold.

Remark 10: If ∇T is not zero, then the torsion may not decompose, and so one has only a Riemannian decomposition.

It is now clear that $\mathring{\nabla}T = 0$ cannot be replaced by $\nabla T = 0$, because Proposition 6 means that local indecomposability does not imply RC local irreducibility, and this is what one needs in order for Schur's lemma to apply and for the proof to go through. In fact, all other parts of the proof of Proposition 1 can be suitably modified, and the result remains valid if $\nabla T = 0$, but only if T splits holonomically. But there is no physical motivation for this last assumption.

In conclusion, then, we see that Proposition 1 is very sensitive to modifications of the hypotheses.

VI. CONCLUSION

In this work we have examined the foundations of the Freund–Rubin spontaneous compactification technique, and have indicated the form that an extension of these ideas to Riemann–Cartan manifolds could take. Our purpose has been to provide a framework that not only guides the construction of particular models (in the sense that these should be compatible with the hypotheses of Proposition 1 or some similar result), but which also sheds some light on the whole question of compactification. Although our treatment has been primarily concerned with the Kaluza–Klein approach, many of the results apply also to the "field-theoretic limit" of superstring theories and possibly to other multidimensional theories. (This is why we have avoided, as far as possible, any assumptions as to the symmetry group of the internal manifold: in particular, Proposition 1 is independent of any such assumption.)

The specific technical problems that afflict Kaluza–Klein theories (chiral fermions, zero-mass modes—see Witten[20] and Muzinich[21] need not be rehearsed here. We shall conclude instead by pointing out some more general problems that deserve greater attention.

It is often claimed as a virtue of Kaluza–Klein theories that they reduce gauge theories to gravitation, and that they explain the origin of gauge symmetries. This is somewhat dubious, however, because these theories postulate *at the outset* that the internal space has a nontrivial isometry group. Most manifolds, of course, do not have this property. In general relativity theory, nontrivial isometry groups are imposed only as a useful approximation. The electromagnetic gauge symmetry, however, is practically exact. From this point of view, the Kaluza–Klein manifolds are thus highly "nongeneric." The virtue of the Kaluza–Klein formulation of gauge theory is not so much that it explains the symmetry as that it may provide a route to an explanation. One could imagine high degrees of symmetry arising, for example, from quantum gravitational effects of the type that tend to reduce anisotropies in cosmology.[22] Another line of approach is suggested by the work of Isenberg and Moncrief,[23] who show that, under certain conditions, a space-time must inevitably develop nontrivial isometries.

At the other extreme, the claim is often made that gravitation can be reduced to gauge theory or the field theory of spin-2 particles. The analyses of Trautman[24] and Penrose,[25] respectively, make it quite clear that these viewpoints are considerably, and perhaps grossly, oversimplified.

However, the most basic problem confronting all higher-dimensional theories is that of understanding the reason for the "factorization" of the universe into "internal" and "external" parts. Spontaneous compactification has the relatively modest aim of explaining the topological differences of the two factors in terms of their geometric differences—the point being that Myers' theorem (and related results such as Proposition 1) is not valid for pseudo-Riemannian manifolds. Thus, the internal space differs topologically from space-time because the latter has a time dimension and the former does not (and cannot, lest causality be violated). But in all this, the "factorization" is presumed to be given—it is certainly not explained.

We have already remarked, in Sec. III, that the existence of a time dimension may be partly responsible for the factorization of the universe, since it implies that the full multidimensional space cannot be compact. But this is obviously a very incomplete explanation. No theory that makes use of multidimensional spaces can be considered complete *unless* it gives a detailed account of the origin of the internal/external dichotomy. It is sometimes stated that this problem can be resolved by considering nontrivial fiber bundles instead of product spaces, but this is in fact not correct. Not every manifold can be regarded as the bundle space of a fiber bundle—the structure must be imposed by means of a postulate that is hardly less arbitrary than the assumption that the space factorizes. What is required is a physically motivated scheme which splits the universe "spontaneously." The de Rham decomposition theorems and their generalizations may be of value here.

ACKNOWLEDGMENTS

The author is grateful to Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for the hospitality at the International Centre for Theoretical Physics, Trieste, where this work was completed. He also wishes to thank Dr. J. A. Hëlayel-Neto and Dr. O. Foda for several very helpful discussions.

APPENDIX: PROOF OF PROPOSITION 6

Here we shall only indicate those parts of the proof that differ from the corresponding parts in the Riemannian case; the remainder will be asserted without proof.[9]

Let H' and H'' be the distributions defined throughout U in the way described in the text. It may be shown that these distributions are differentiable and that, if X is any vector field, $\nabla_x(H')\subseteq H'$ and $\nabla_x(H'')\subseteq H''$. Thus if X and Y are vector fields belonging to H', $[X,Y]$ also belongs to H', since $[X,Y]=\nabla_x Y-\nabla_x X-T(X,Y)$ and T splits holonomically. Thus H', and similarly H'', is involutive. Given the point x, one can therefore (by the Frobenius integrability theorem) find a submanifold M' generated by H' and containing x, and similarly for M'' generated by H''. It now can be shown that there exists a neighborhood V around x in U, which is of the form $V'XV''$, where V' is an open neighborhood of x, in M', and similarly for V'' in M''. Let $k=\dim H'$, and let $\{x^1,\ldots,x^k,x^{k+1},\ldots,x^n\}$ be a coordinate system adapted to the splitting, that is, $\partial_1\cdots\partial_k$ belong to H', and $\bar\partial_{k+1}\cdots\bar\partial_n$ belong to H''. We now show that the metric in V' is independent of the coordinates in V'', and vice versa.

The "off-diagonal" components of g, $g(\partial_i,\bar\partial_j)$ are obviously zero. To see that $\partial_i(g(\bar\partial_j,\bar\partial_m))$ is zero, we use $\nabla g=0$ to write

$$\partial_i(g(\bar\partial_i,\bar\partial_m))=g(\nabla_{\partial i}\bar\partial_j,\bar\partial_m)+g(\bar\partial_j,\nabla_{\partial i}\bar\partial_m). \tag{A1}$$

Now by definition of T we have

$$\nabla_{\partial i}\bar\partial_i-\nabla_{\bar\partial j}\partial_i-[\partial_i,\bar\partial_j]=T(\partial_i,\bar\partial_j).$$

The third term on the left vanishes since we are dealing with a coordinate basis and the right-hand side vanishes since T splits holonomically. Thus, the first two terms are equal. But the first belongs to H'', and the second to H': hence, both must be zero. Substituting into (A1), we find that the V'' part of the metric is independent of the V' coordinates. Similarly the V' part of the metric is independent of the V'' coordinates, and so V is the Riemannian product of V' and V''.

Notice that the proof depends only on $\nabla g=0$. Thus T will decompose in the same way if $\nabla T=0$, and so, in this case V is also the Riemann–Cartan product of V' and V''. This completes the proof.

[1] *Supersymmetry and Supergravity '84*, edited by B. de Wit, P. Fayet, and P. van Nieuwenhuizen (World Scientific, Singapore, 1984).
[2] B. T. McInnes, Phys. Lett. B **150**, 113 (1985).
[3] B. T. McInnes, Class Quantum Grav. **2**, 661 (1985).
[4] P. G. O. Freund and M. A. Rubin, Phys. Lett. B **97**, 233 (1980).
[5] J. Cheeger and D. G. Ebin, *Comparison Theorems in Riemannian Geometry* (North-Holland, Amsterdam, 1975).
[6] C. Destri, C. A. Orzalesi, and P. Rossi, Ann. Phys. (NY) **147**, 321 (1983).
[7] M. J. Duff, B. E. W. Nilsson, and C. N. Pope, Nucl. Phys. B **233**, 433 (1984).
[8] X. Wu, Phys. Lett. B **144**, 51 (1984).
[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[10] P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985).
[11] F. W. Hehl, P. von der Heyde, and G. D. Kerlick, Rev. Mod. Phys. **48**, 393 (1976).
[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II.
[13] N. J. Hicks, Ill. J. Math. **3**, 242 (1959).
[14] J. A. Wolf, J. Diff. Geom. **6**, 317 (1972).
[15] F. Englert, Phys. Lett. B **119**, 339 (1982).
[16] P. Candelas and S. Weinberg, Nucl. Phys. B **237**, 397 (1984).
[17] R. P. Geroch, J. Math. Phys. **8**, 782 (1967).
[18] J. K. Beem and P. E. Ehrlich, *Global Lorentzian Geometry* (Dekker, New York, 1981).
[19] J. P. Bourguignon, *Global Differential Geometry and Analysis, Lecture*

Notes in Mathematics, Vol. 838, edited by D. Ferus (Springer, Berlin, 1981), p. 42.
[20]E. Witten, Princeton preprint, October 1983.
[21]I. J. Muzinich, J. Math. Phys. **26**, 1942 (1985).
[22]L. Parker, *Quantum Gravity: Essays in Honour of the 60th Birthday of Bryce S. De Witt*, edited by S. M. Christensen (Adam Hilger, Bristol, 1984).
[23]J. Isenberg and V. Moncrief, J. Math. Phys. **26**, 1024 (1985).
[24]A. Trautman, *General Relativity and Gravitation, One Hundred Years After the Birth of Albert Einstein*, edited by A. Held (Plenum, New York, 1980).
[25]R. Penrose, *Essays in General Relativity, A Festschrift for Abraham Taub*, edited by F. J. Tipler (Academic, New York, 1980).

Transverse nearest neighbor degeneracies on a $2 \times N$ lattice

William Roger Fuller, CSC
Departments of Mathematics and Physics, University of Portland, Portland, Oregon 97203

(Received 2 January 1986; accepted for publication 26 March 1986)

An exact 15-term recursion relation and the associated generating function are derived for the number of arrangements of q particles on a $2 \times N$ lattice such that s occupied nearest neighbor pairs, v of which are transverse, and t unoccupied nearest neighbor pairs are formed.

I. INTRODUCTION

In the mathematical statistical mechanics of adsorption of molecules on double chains of polymers the following combinatoric problem arises: Obtain a recursion relation for the number $A(N,q,s,t,v)$ of arrangements of q identical, indistinguishable particles on a $2 \times N$ rectangular lattice, composed of two horizontal rows of N square cells, in such a way that there are a number s of occupied nearest neighbor pairs, v of which are vertical, and a number t of unoccupied nearest neighbor pairs. In Fig. 1, for example, ten particles are arranged on a 2×8 lattice. In this case there are a total of eight occupied nearest neighbor pairs and a total of three unoccupied nearest neighbor pairs. Three of the occupied nearest neighbor pairs are vertical.

The desired recursion relation may be obtained by first utilizing a method of McQuistan and Hock[1] to derive a set of coupled recursion relations for conditional arrangement numbers and then by employing a technique developed by Phares[2] for decoupling the recursion relations. These two steps are presented in Secs. II and III of this article. In Sec. IV a previously obtained result is derived by a reduction of the relation for $A(N,q,s,t,v)$.

II. RECURSION RELATIONS FOR CONDITIONAL ARRANGEMENT NUMBERS

The conditional arrangement numbers are the numbers of arrangements of the particles with the specified numbers and types of nearest neighbor pairs, given the configuration of occupied and unoccupied cells on the leftmost edge of the lattice. The four such arrangement numbers will be labeled $A_j(N,q,s,t,v)$, with $j = 1,2,3,4$. The given configurations are indicated in Fig. 2. For example, $A_1(N,q,s,t,v)$ denotes the number of arrangements of q particles on a $2 \times N$ lattice such that there are s occupied nearest neighbor pairs, v of which are vertical, and t unoccupied nearest neighbor pairs, provided that the upper leftmost cell is occupied and that the lower leftmost cell is unoccupied.

As a consequence of the relation

$$A(N,q,s,t,v) = \sum A_j(N,q,s,t,v), \tag{1}$$

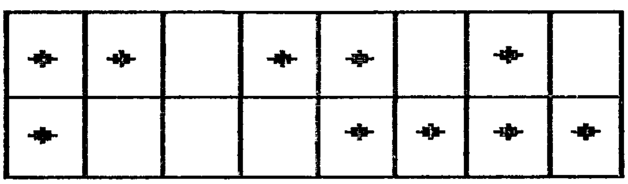


FIG. 1. On this 2×8 lattice ten particles have been placed so that $s = 8$, $t = 3$, $v = 3$, and $e = 3$.

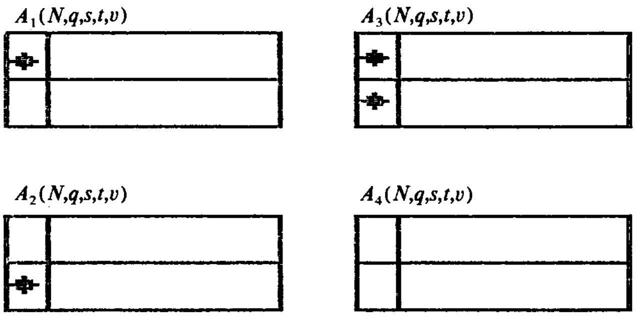


FIG. 2. Arrangements may be characterized by the state of the two leftmost cells of the lattice. The conditional arrangement numbers associated with each state are indicated.

the question of obtaining a recursion relation for $A(N,q,s,t,v)$ reduces to one of obtaining relations for the conditional arrangement numbers, and this shall be the focus of the following discussion.

The variables $N,q,s,t,$ and v have for their domains the non-negative integers, and the arrangement number $A(N,q,s,t,v)$ and the conditional arrangement numbers $A_j(N,q,s,t,v)$ have specific initial conditions. For the sake of simplicity, however, we shall assume that the variables range unrestrictedly over the entire set of integers, and that the arrangement number functions take nonzero values only for those values of their arguments that make sense and are self-consistent. We adopt the following convention concerning the values of the domain variables for which the arrangement number functions may be nonzero. Since the lattice must have some length, the arrangement numbers may be nonzero only for positive N. For nonzero values of these numbers the number q of particles distributed on the lattice must be non-negative and less than $2N$. The number s of occupied nearest neighbor pairs must also be non-negative and less than some maximum number s_m, which is never greater than $3(q/2) - 2$. It is clear that v can only range from zero to the lesser of the two numbers N and s_m. Finally, t must be a non-negative number equal to $3(N-q) - 2 + s + e$, where e is the number of particles found in the two leftmost and the two rightmost cells of the lattice.

The convention just defined permits a reduction of the question of listing the initial conditions of the conditional arrangement numbers to that of listing their first nonzero values when their domain quintuples are ordered lexicographically. These conditions are

$$A_1(1,1,0,0,0) = 1, \tag{2a}$$

$$A_2(1,1,0,0,0) = 1, \tag{2b}$$

$$A_3(1,2,1,0,1) = 1, \tag{2c}$$

 0022-2488/86/082039-04$02.50

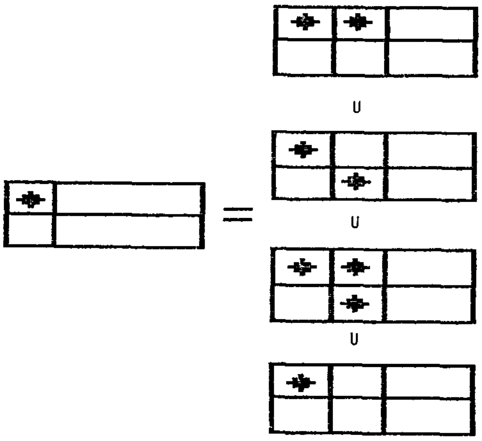


FIG. 3. The partition of the set of arrangements for $A_1(N,q,s,t,v)$ is indicated schematically. Such partitions give rise to the recursion relations (3).

$$A_4(1,0,0,1,0) = 1. \tag{2d}$$

The conditional arrangement numbers are not independent. It is clear from Fig. 2, for example, that since the $2 \times N$ lattice is symmetric under a reflection about the central line, $A_1(N,q,s,t,v)$ and $A_2(N,q,s,t,v)$ are identical. The task at hand is thus to obtain the recursion relations for three functions in five variables.

One way to obtain coupled recursion relations for the conditional arrangement numbers is to take each of the cases in Fig. 2 and partition the set of associated arrangements into four disjoint classes characterized by how the two cells second from the left in the lattice are occupied. The decomposition is shown schematically for $A_1(N,q,s,t,v)$ in Fig. 3. Consideration of similar decompositions for the other conditional arrangement numbers leads to four coupled recursion relations:

$$\begin{aligned} A_1&(N,q,s,t,v) \\ &= A_1(N-1,q-1,s-1,t-1,v) \\ &\quad + A_2(N-1,q-1,s,t,v) \\ &\quad + A_3(N-1,q-1,s-1,t,v) \\ &\quad + A_4(N-1,q-1,s,t-1,v), \end{aligned} \tag{3a}$$

$$\begin{aligned} A_2&(N,q,s,t,v) \\ &= A_1(N-1,q-1,s,t,v) \\ &\quad + A_2(N-1,q-1,s-1,t-1,v) \\ &\quad + A_3(N-1,q-1,s-1,t,v) \\ &\quad + A_4(N-1,q-1,s,t-1,v), \end{aligned} \tag{3b}$$

$$\begin{aligned} A_3&(N,q,s,t,v) \\ &= A_1(N-1,q-2,s-2,t,v-1) \\ &\quad + A_2(N-1,q-2,s-2,t,v-1) \\ &\quad + A_3(N-1,q-2,s-3,t,v-1) \\ &\quad + A_4(N-1,q-2,s-1,t,v-1), \end{aligned} \tag{3c}$$

$$\begin{aligned} A_4&(N,q,s,t,v) \\ &= A_1(N-1,q,s,t-2,v) \\ &\quad + A_2(N-1,q,s,t-2,v) \\ &\quad + A_3(N-1,q,s,t-1,v) \\ &\quad + A_4(N-1,q,s,t-3,v). \end{aligned} \tag{3d}$$

III. DECOUPLING THE RECURSION RELATIONS

The recursion relations for the conditional arrangement numbers may be decoupled by introduction of the generating function associated with each arrangement number. For each $j = 1,2,3,4$, we define a polynomial $G_j(X)$ in the quintuple of variables $X = (x,y,z,a,b)$:

$$G_j(X) = \sum A_j(N,q,s,t,v)\, x^N y^q z^s a^t b^v. \tag{4}$$

The ranges of the indices of summation may be considered to be the set of all integers; the convention adopted concerning the arrangement numbers ensures that for each value of N almost all terms are zero.

Substitution of the relations (3) and of the initial conditions (2) into the summations (4) yields a set of simultaneous equations for the generating functions:

$$(1 - xyza - xy)G_1(X) - xyzG_3(X) - xyaG_4(X) = xy, \tag{5a}$$

$$-2xy^2z^2bG_1(X) + (1 - xy^2z^3b)G_3(X) - xy^2zbG_4(X) = xy^2zb, \tag{5b}$$

$$-2xa^2G_1(X) - xaG_3(X) + (1 - xa^3)G_4(X) = xa. \tag{5c}$$

Solution of the system (5) yields the generating functions for $j = 1,2,3,4$:

$$G_j(X) = H_j(X)/D(X),$$

where

$$\begin{aligned} D(X) = {} & 1 - xy - xyza - xa^3 - x^2ya^3 - xy^2z^3b - x^2y^2zab \\ & + x^2yza^4 + x^3y^3zab - x^2y^3z^3b \\ & + x^2y^3z^4ab + x^2y^2z^3a^3b \\ & - 3x^3y^3z^2a^2b + 3x^3y^3z^3a^3b - x^3y^3z^4a^4b, \\ H_1(X) = {} & xy + x^2ya^2 - x^2ya^3 + x^2y^3z^2b \\ & - x^3y^3zab - x^2y^3z^3b + x^3y^3za^2b \\ & + x^3y^3z^2ab - x^3y^3z^3a^2b - x^3y^3z^2a^3b + x^3y^3z^3a^3b, \\ H_2(X) = {} & H_1(X), \\ H_3(X) = {} & xy^2zb - x^2y^3zb + x^2y^2zab + 2x^2y^3z^2b \\ & - x^2y^3z^2ab - x^2y^2za^3b - x^3y^3zab \\ & + 2x^3y^3za^2b + x^3y^3z^2a^2b \\ & - x^3y^3za^3b - 2x^3y^3z^2a^3b + x^3y^3z^2a^4b, \\ H_4(X) = {} & xa - x^2ya + 2x^2ya^2 - x^2yza^2 + x^2y^2zab \\ & - x^2y^2z^3ab - x^3y^3zab + 2x^3y^3z^2ab \\ & - x^3y^3z^3ab + x^3y^3z^2a^2b \\ & - 2x^3y^3z^3a^2b + x^3y^3z^4a^2b. \end{aligned}$$

Comparison of the coefficients of like monomials in the equations

$$D(X)G_j(X) = H_j(X) \tag{6}$$

yields recursion relations and associated sets of initial conditions for each of the conditional arrangement numbers.

TABLE I. Nonzero initial values for A_1.

(N,q,s,t,v)	$A_1(N,q,s,t,v)$
(1,1,0,0,0)	1
(2,1,0,2,0)	1
(2,2,0,0,0)	1
(2,2,1,1,0)	1
(2,3,2,0,1)	1
(3,1,0,5,0)	1
(3,2,0,2,0)	1
(3,2,0,3,0)	2
(3,2,1,3,0)	1
(3,3,0,0,0)	1
(3,3,1,1,0)	2
(3,3,1,2,1)	1
(3,3,2,1,1)	1
(3,3,2,2,0)	1

Since the highest power of x in any of the functions $H_j(X)$ is the third, all the $A_j(N,q,s,t,v)$, and consequently $A(N,q,s,t,v)$, obey the same recursion relation when $N > 3$:

$$\begin{aligned}0 = {} & A(N,q,s,t,v) - A(N-1,q-1,s,t,v) \\ & - A(N-1,q-1,s-1,t-1,v) \\ & - A(N-1,q,s,t-3,v) \\ & - A(N-2,q-1,s,t-3,v) \\ & - A(N-1,q-2,s-3,t,v-1) \\ & - A(N-2,q-2,s-1,t-1,v-1) \\ & + A(N-2,q-1,s-1,t-4,v) \\ & + A(N-3,q-3,s-1,t-1,v-1) \\ & - A(N-2,q-3,s-3,t,v-1) \\ & + A(N-2,q-3,s-4,t-1,v-1) \\ & + A(N-2,q-2,s-3,t-3,v-1) \\ & - 3A(N-3,q-3,s-2,t-2,v-1) \\ & + 3A(N-3,q-3,s-3,t-3,v-1) \\ & - A(N-3,q-3,s-4,t-4,v-1). \end{aligned} \tag{7}$$

The nonzero initial values for the conditional arrangement numbers are given in Tables I–III. The appropriate initial conditions for $A(N,q,s,t,v)$ may be obtained from these tables and Eq. (1). The generating function for $A(N,q,s,t,v)$ is the sum of the generating functions for the conditional arrangement numbers:

$$G(X) = G_1(X) + G_2(X) + G_3(X) + G_4(X).$$

TABLE II. Nonzero initial values of A_3.

(N,q,s,t,v)	$A_3(N,q,s,t,v)$
(1,2,1,0,1)	1
(2,2,1,1,1)	1
(2,3,2,0,1)	2
(2,4,4,0,2)	1
(3,2,1,4,1)	1
(3,3,1,2,1)	2
(3,3,2,2,1)	2

TABLE III. Nonzero initial values of A_4.

(N,q,s,t,v)	$A_4(N,q,s,t,v)$
(1,0,0,1,0)	1
(2,0,0,4,0)	1
(2,1,0,2,0)	2
(2,2,1,1,1)	1
(3,0,0,7,0)	1
(3,1,0,4,0)	2
(3,1,0,5,0)	2
(3,2,0,2,0)	2
(3,2,1,2,1)	1
(3,2,1,3,0)	2
(3,3,2,1,1)	2
(3,3,2,2,1)	2

IV. CONCLUSION

In this section we indicate how restrictions of the recursion relation (7) and of its generating function lead to further results. The summation

$$\sum_v A(N,q,s,t,v) \tag{8}$$

represents the number of ways of arranging q particles on a $2\times N$ lattice with s occupied nearest neighbor pairs and with t unoccupied nearest neighbor pairs. Let us denote this number by $A(N,q,s,t)$; then $G(x,y,z,a,1)$ is the generating function for $A(N,q,s,t)$. A reduced form of Eq. (6),

$$D(x,y,z,a,1)G(x,y,z,a,1) = H(x,y,z,a,1),$$

indicates that the recursion relation for $A(N,q,s,t)$ may be read off from $D(x,y,z,a,1)$:

$$\begin{aligned}0 = {} & A(N,q,s,t) - A(N-1,q-1,s,t) \\ & - A(N-1,q-1,s-1,t-1) - A(N-1,q,s,t-3) \\ & - A(N-2,q-1,s,t-3) - A(N-1,q-2,s-3,t) \\ & - A(N-2,q-2,s-1,t-1) \\ & + A(N-2,q-1,s-1,t-4) \\ & + A(N-3,q-3,s-1,t-1) \\ & - A(N-2,q-3,s-3,t) \\ & + A(N-2,q-3,s-4,t-1) \\ & + A(N-2,q-2,s-3,t-3) \\ & - 3A(N-3,q-3,s-2,t-2) \\ & + 3A(N-3,q-3,s-3,t-3) \\ & - A(N-3,q-3,s-4,t-4). \end{aligned} \tag{9}$$

The same convention regarding the ranges of the domain variables and regarding the nonzero values of $A(N,q,s,t)$ holds just as it does for $A(N,q,s,t,v)$. The initial values for $A(N,q,s,t)$ may be obtained by use of (8) and the initial values for the conditional arrangement numbers. Relation (9) was obtained by McQuistan and Hock[1] by a different method.

Similarly, $G(x,y,z,1,1)$ is the generating function for the numbers $A(N,q,s)$ of arrangements of q particles on a $2\times N$ lattice with exactly s occupied nearest neighbor pairs. As above, the recursion relation for $A(N,q,s)$ may be read off from $D(x,y,z,1,1)$:

$$
\begin{aligned}
0 = {} & A(N,q,s) - A(N-1,q,s) \\
& - A(N-1,q-1,s) - A(N-1,q-1,s-1) \\
& - A(N-2,q-1,s) + A(N-2,q-1,s-1) \\
& - A(N-2,q-2,s-1) - A(N-1,q-2,s-3) \\
& + A(N-3,q-3,s-1) + A(N-2,q-2,s-3) \\
& - A(N-2,q-3,s-3) - 3A(N-3,q-3,s-2) \\
& + A(N-2,q-3,s-4) + 3A(N-3,q-3,s-3) \\
& - A(N-3,q-3,s-4).
\end{aligned}
$$

This expression is a generalization of a result obtained for the $1\times N$ case by McQuistan.[3]

The recursion relation (7) may be used to gain information about the grand canonical partition function of a physical system of gas molecules being adsorbed onto double strands of polymers. Results concerning such a system will be published elsewhere.

[1]R. B. McQuistan and J. L. Hock, "An exact recursion for the composite nearest-neighbor degeneracy for a $2\times N$ lattice space," J. Math. Phys. **25**, 261 (1984).
[2]A. J. Phares, "Decoupling of a system of partial difference equations with constant coefficients and application," J. Math. Phys. **25**, 2169 (1984).
[3]R. B. McQuistan, "Exact nearest neighbor statistics for one-dimensional lattice spaces," J. Math. Phys. **13**, 1317 (1972).

Special solutions of the Sparling equation

S. H. Hsieh
Department of Physics, Soochow University, Shih-Lin, Wei Shauang Shi, Taipei, Taiwan, Republic of China
Steven L. Kent
Department of Mathematics and Computer Science, Youngstown State University, Youngstown, Ohio 44555
E. T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260 and Institute of Theoretical Physics, University of California, Santa Barbara, California 93106

(Received 12 March 1986; accepted for publication 9 April 1986)

The Sparling equation, a first-order, matrix-valued linear differential equation that is equivalent to the self-dual Yang–Mills equations for any group, has recently been solved by quadratures for the case of SL(2,C) or its subgroup. It is the purpose of this paper to show how for a series of special cases, rather than integrating the quadratures, the Sparling equation can be reduced to an algebraic equation and then solved, yielding the single- and multi-instanton fields parallel to isospace.

I. INTRODUCTION

The self-dual Yang–Mills equations have, for a variety of reasons, played an important role in both physics and applied mathematics. They have supplied the instanton solutions used in the quantum theory of the Yang–Mills field and recently they have been shown[1] (for different gauge groups and special symmetries) to reduce to a wide variety of differential equations as, for example, the Bogomolny equation, the stationary, axial symmetric Einstein equations, the sine–Gordon equation, the Euler equations for the spinning top, and others. It is thus clear that solution generating techniques are of considerable interest. In this paper one such technique will be discussed. In a future paper a more powerful method [which promises to reduce the (at least) SL(2,C) case to quadratures] will be presented.

It has been known[2,3] for some time that one can express the self-dual Yang–Mills field equations on Minkowski space in terms of a single, matrix, first-order differential equation for a GL(n,c) matrix-valued function $G(x^a,\zeta,\bar{\zeta})$. This equation (the so-called Sparling equation) has the form

$$(1+\zeta\bar{\zeta})\,\frac{\partial G}{\partial \zeta}\equiv\eth G=-GA, \tag{1.1}$$

where A is a given but arbitrary (Lie-algebra) matrix-valued function on a characteristic surface (usually taken as null infinity, I^+), representing the characteristic initial data for the Yang–Mills field. Thus A is a function of three variables that coordinatize $I^+(R\times S^2)$: the retarded time $u=t-r$ and the complex stereographic coordinates ζ and $\bar{\zeta}$. If an interior point x^a of Minkowski space is chosen, the light cone from x^a intersects I^+ on a two-surface that can be described by

$$u=l(x^a,\zeta,\bar{\zeta}), \tag{1.2}$$

with l a simple, explicit function of x^a and $(\zeta,\bar{\zeta})$. When Eq. (1.2) is substituted into $A(u,\zeta,\bar{\zeta})$, Eq. (1.1) becomes a differential equation in ζ for $G(x^a,\zeta,\bar{\zeta})$, with the x^a entering only as parameters in A.

The solutions of (1.1) with the demand of regularity on the $(\zeta,\bar{\zeta})$ sphere (i.e., existence of expansion in spherical harmonics) constitute the solution of the associated Yang–Mills equations in the sense that from the regular $G(x^a,\zeta,\bar{\zeta})$, one can, by differentiating, directly construct the "self-dual" Yang–Mills vector potential[3] γ_a,

$$\gamma_a=\nabla_a G\cdot G^{-1}+\eth h l_a-hm_a, \tag{1.3}$$

with $h=l^a\bar{\eth}(\nabla_a GG^{-1})$, and l^a and m^a explicit functions $(\zeta,\bar{\zeta})$. (See Sec. II.)

The freedom in the regular solutions

$$G\to G'(x^a,\zeta,\bar{\zeta})=g(x^a)G(x^a,\zeta,\bar{\zeta}) \tag{1.4}$$

is a manifestation of the usual gauge freedom in the choice of γ_a, i.e.,

$$\gamma'_a=g\gamma_a g^{-1}+\nabla_a g\cdot g^{-1}. \tag{1.5}$$

Though on a formal level, Eq. (1.1) can always be solved in terms of path-ordered integrals

$$G(x^a,\zeta,\bar{\zeta})=G_0(x^a,\bar{\zeta})O\exp\left(\int^{\zeta}A\,\frac{d\zeta}{1+\zeta\bar{\zeta}}\right), \tag{1.6}$$

the difficulty of choosing the $G_0(x^a,\zeta)$ so that G is a regular function has never been solved. Recently,[4] an alternative method of solution has been developed [in the SL(2,C) case] that allows the G to be expressed in terms of quadratures over A.

It is the purpose of this paper to show how, for a special class of A, the method of quadratures can be circumvented and the Sparling equation can be reduced to an algebraic equation. The solutions correspond to the single- or multi-instanton fields, which are parallel in isospace.

In Sec. II, we discuss the meaning and derivation of the Sparling equation, while in Sec. III, we choose a special A and obtain the single-instanton solution. In Sec. IV, this in generalized to the parallel multi-instanton fields.

II. THE SPARLING EQUATION

In this section we will discuss the meaning of the Sparling equation and its derivation. The details will be omitted as they have appeared elsewhere.[5,6]

We begin with a GL(n,C) bundle, with connection $\gamma_a(x^a)$, over Minkowski space M. Let x^a be an arbitrary point of M and C_x denote the future null cone of x^a. The null

 0022-2488/86/082043-04$02.50

generators of C_x will be labeled by $(\zeta,\bar{\zeta})$ obtained from their intersection with the generators of I^+, and will be denoted by $l_x(\zeta,\bar{\zeta})$. The $GL(n,C)$-valued matrix $G(x^a,\zeta,\bar{\zeta})$ is the parallel propagator of the fiber over x^a to I^+ along $l_x(\zeta,\bar{\zeta})$ via the connection γ, i.e.,

$$G(x^a,\zeta,\bar{\zeta}) = O \exp\left(\int_{l_x(\zeta,\bar{\zeta})} \gamma_a \, dx^a\right) = O \exp\left(\int \gamma_a l^a \, ds\right), \tag{2.1}$$

where O indicates the path-ordered integral and $l^a(\zeta,\bar{\zeta})$ is the (null) tangent vector to $l_x(\zeta,\bar{\zeta})$.

Note that from the definition of G one can, by differentiation, obtain

$$l^a \nabla_a G \cdot G^{-1} = \gamma_a l^a, \tag{2.2}$$

from which Eq. (1.3) can be derived.

If we now restrict the discussion to self-dual Yang–Mills fields, the Sparling equation can be derived in the following fashion.

Consider the two infinitesimally close generators $l(\zeta,\bar{\zeta})$ and $l(\zeta + d\zeta,\bar{\zeta})$ of C_x. (The second is slightly into the complexified Minkowski space.) They lie in an integrable two-blade, which is anti-self-dual. If we connect these two generators by a connecting vector on I^+, we have an infinitesimally narrow closed loop. Since we are dealing with a self-dual field F_{ab}, parallel transport around an anti-self-dual loop is integrable, i.e., parallel transport around an anti-self-dual loop always yields the identity. Expressed in terms of the parallel propagator G we have

$$G^{-1}(\zeta,\bar{\zeta})G(\zeta + d\zeta,\bar{\zeta})[I + A\, d\zeta/(1+\zeta\bar{\zeta})] = I, \tag{2.3}$$

where $I + A\, d\zeta/(1+\zeta\bar{\zeta})$ is the infinitesimal parallel propagator along the connecting vector on I^+, and $A(u,\zeta,\bar{\zeta})$ is an asymptotic component (along the connecting vector) of the connection on I^+. The point $(u,\zeta,\bar{\zeta})$ on I^+ is the intersection of $l_x(\zeta,\bar{\zeta})$ with I^+ and has the form

$$u = l(x^a,\zeta,\bar{\zeta}) = x^a l_a(\zeta,\bar{\zeta}). \tag{2.4}$$

Finally, by expanding Eq. (2.3) we obtain

$$\eth G = -GA, \tag{2.5}$$

the Sparling equation.

III. THE SINGLE-INSTANTON SOLUTION

Before studying the solutions of the Sparling equations, we introduce some notation. A useful representation of $l^a(\zeta,\bar{\zeta})$ is

$$l^a = \frac{1}{\sqrt{2}}\left(1, \frac{\zeta+\bar{\zeta}}{1+\zeta\bar{\zeta}}, \frac{i(\bar{\zeta}-\zeta)}{1+\zeta\bar{\zeta}}, \frac{-1+\zeta\bar{\zeta}}{1+\zeta\bar{\zeta}}\right), \tag{3.1}$$

and the related vectors[5]

$$m^a = \eth l^a, \tag{3.2}$$

$$\bar{m}^a = \bar{\eth} l^a, \tag{3.3}$$

$$n^a = l^a + \eth\bar{\eth} l^a. \tag{3.4}$$

The set l^a, m^a, $\bar{m}^a$, n^a are closed under $\eth$ and $\bar{\eth}$ since

$$\eth m^a = \bar{\eth}\bar{m}^a = 0, \quad \eth n^a = -m^a, \quad \bar{\eth} n^a = -\bar{m}^a. \tag{3.5}$$

For arbitrary but fixed $(\zeta,\bar{\zeta})$, l^a, m^a, $\bar{m}^a$, and n^a form a standard null tetrad field with all scalar products vanishing except

$$l^a n_a = 1 = -m^a \bar{m}_a. \tag{3.6}$$

One easily sees that the Minkowski metric can be written for any $(\zeta,\bar{\zeta})$

$$\eta_{ab} = 2l_{(a}n_{b)} - 2m_{(a}\bar{m}_{b)}. \tag{3.7}$$

By multiplying (3.7) with x^a we have

$$x^a = l^a n + n^a l - m^a \bar{m} - \bar{m}^a m, \tag{3.8}$$

where

$$l = x^a l_a, \quad n = x^a n_a, \quad m = x^a m_a, \quad \bar{m} = x^a \bar{m}_a. \tag{3.9}$$

We also introduce the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \tag{3.10}$$

and $\sigma_0 = I$. We then have

$$\sigma_i \sigma_j = -\delta_{ij} + \epsilon_{ijk}\sigma^k, \tag{3.11}$$

with $\epsilon_{123} = 1$.

We now consider solving the Sparling equation for the following characteristic data:

$$A(u,\zeta,\bar{\zeta}) = A_0(\zeta,\bar{\zeta})/u^2, \tag{3.12}$$

with $u = l \equiv x^a l_a(\zeta,\bar{\zeta})$ and A_0 a 2×2 trace-free matrix with the properties

$$A_0^2 = 0 \quad \text{and} \quad \eth A_0 = 0. \tag{3.13}$$

From the trace-free condition and (3.13), we have that

$$A = a_0 m^i \sigma_i / l^2, \tag{3.14}$$

with a_0 a constant, and m^i the spatial part of m^a.

Now using (3.14) in the Sparling equation with (3.13) it is easy to show that

$$\eth^2(lG) = 0, \tag{3.15}$$

which implies that G has the form

$$G = \beta_a l^a / l, \tag{3.16}$$

where β_a is a 2×2 matrix-valued vector function of x^a only, which can be written

$$\beta_a = \beta_a^\mu \sigma_\mu, \quad \mu = 0,1,2,3.$$

The idea is now to substitute (3.16) back into the Sparling equation, obtaining an algebraic equation for the determination of β_a^μ. On the assumption that β_a^μ is a nonsingular matrix, one can show that choosing $\beta_a^0 = x^a$ is a unique gauge-fixing device and hence we can choose G to have the form

$$G = I + \beta_a^i l^a \sigma_i / l, \quad i = 1,2,3. \tag{3.17}$$

Rather than directly substituting (3.17) into the Sparling equation, it turns out to be simpler to put (3.17) into a slightly different form and then use the Sparling equation itself to find the solution.

We write

$$\beta_a^i l^a \equiv G_0 l^i + G_- m^i + G_+ \bar{m}^i, \tag{3.18}$$

with l^i, m^i, $\bar{m}^i$ the spatial parts of l^a, m^a, and $\bar{m}^a$, respectively. Substituting (3.17), with (3.18), into the Sparling equation, and considering the coefficients of I, $\sigma_i l^i$, $\sigma_i m^i$, $\sigma_i \overline{m}^i$, We obtain four equations. From this list of these equations, one has immediately that

$$G_+ = 0. \tag{3.19}$$

This identically satisfies the last equation. The second and third equations become

$$\eth G_0 - (m/l) G_0 = 0, \tag{3.20}$$

$$\eth G_- + G_0 - \frac{m}{l} G_- = \frac{-a_0}{l} + \frac{i}{\sqrt{2}} \frac{a_0 G_0}{l^2}. \tag{3.21}$$

Equation (3.20) is easily integrated as

$$G_0 = \phi(x^a) x_a l^a = \phi l. \tag{3.22}$$

Now by applying $\eth$ to (3.21), we obtain

$$\eth^2 G_- = -2\,\eth G_0 = -2\phi m, \tag{3.23}$$

which integrates uniquely to

$$G_- = \phi x_a \overline{m}^a = \phi \overline{m}. \tag{3.24}$$

Finally, by substituting (3.22) and (3.24) into (3.21), we obtain an algebraic equation for the scalar function $\phi(x)$, yielding

$$\phi(x) = -2a_0/(x_a x^a - \sqrt{2} i a_0). \tag{3.25}$$

Thus, from (3.17)–(3.19), (3.22), (3.24), and (3.25), we have

$$G = I - \frac{2a_0}{(x^a x_a - \sqrt{2} i a_0)} \left(l^i + \frac{\overline{m}}{l} m^i \right) \sigma_i. \tag{3.26}$$

To conclude this section, we point out that the Yang–Mills field obtained from (3.26), using (1.3), is the analytic extension into Minkowski space of the R^4 instanton solution[7] centered on $x^a = 0$. This has been checked by taking the instanton solution (analytically extended) and examining its asymptotic behavior, showing that its characteristic data is the same as (3.12) and (3.13).

IV. THE MULTI-INSTANTON SOLUTIONS

In this section, we will first consider the problem of solving the Sparling equation for (what appears to be) the double-instanton solution, parallel in isospace. (Our lack of certainty is based on the fact that we have not checked our results with the literature but have simply assumed that the linear superposition of characteristic data for two single instantons centered at different points, x_1^a and x_2^a, is the data for the two-instanton field.) Following this, we will indicate how the parallel two-instanton solution can be generalized to the n-parallel instanton case. Though the $n = 2$ case is more difficult than the $n = 1$ case, the generalization to arbitrary n is readily apparent.

We generalize the characteristic data (3.12) and (3.13) to

$$A = (a_1/L_1^2 + a_2/L_2^2) m^i \sigma_i, \tag{4.1}$$

where

$$L_\alpha = (x^a - x_\alpha^a) l_a, \quad \alpha = 1,2,$$

and the a_α are constants. Using (4.1) in the Sparling equation, it is easy to show that

$$\eth^3 (L_1 L_2 G) = 0, \tag{4.2}$$

a generalization of (3.15). Its solution is

$$G(x^a, \zeta, \overline{\zeta}) = (\beta^{\mu}_{ab} l^a l^b / L_1 L_2) \sigma_\mu, \quad \mu = 0,1,2,3, \tag{4.3}$$

where the β^{μ}_{ab} are functions only of x^a. If we now impose the gauge condition that

$$\beta^{0}_{ab} = (x_a - x_{a1})(x_b - x_{b2}),$$

Eq. (4.3) can be written as

$$G = I + (G_0 l^i + G_- m^i + G_+ \overline{m}^i) \sigma_i / L_1 L_2. \tag{4.4}$$

(We could have taken this as the starting ansatz.) The G_0, G_-, and G_+ are to be determined from the Sparling equation with help from the derived equation (4.2), which now takes the form

$$\eth^3 (G_0 l^i + G_- m^i + G_+ \overline{m}^i) = 0. \tag{4.5}$$

Once again, from the Sparling equation we have immediately

$$G_+ = 0, \tag{4.6}$$

which, when used in (4.5), yields

$$\eth^3 G_0 = 0, \tag{4.7}$$

$$\eth^3 G_- = -3\,\eth^2 G_0. \tag{4.8}$$

Equation (4.7) integrates to

$$G_0 = \theta_{ab} l^a l^b, \tag{4.9}$$

with θ_{ab} symmetric and a function only of x^a. The coefficient of $l^i \sigma_i$ in the Sparling equation can be written as

$$\theta_{cd} y_{1(a} y_{2b)} l^a l^c l^{[b} m^{d]} = 0, \tag{4.10}$$

where $y_\alpha^a = x^a - x_\alpha^a$. The algebraic solution is

$$\theta^{ab} = (y_1^{(a} y_2^{b)} - \tfrac{1}{2} y_1^c y_{2c} \eta^{ab}) \psi(x^a), \tag{4.11}$$

where $\psi(x^a)$ is a scalar function to be determined.

Now using (4.9), with (4.11) in (4.8), we have

$$\eth^3 G_- = -6\psi y_1^a y_2^b m_a m_b, \tag{4.12}$$

with a general solution

$$G_- = y_1^a y_2^b \psi l_a \overline{m}_b + 2\alpha^{ab} l_{[a} \overline{m}_{b]}, \tag{4.13}$$

with $\alpha^{ab} = -\alpha^{ba}$ an arbitrary function of x^a. The first term is a particular solution of (4.12) while the second term is the general solution of the homogeneous equation.

We now make the ansatz for α^{ab} that

$$\alpha^{ab} = 2\alpha(x) \psi(x) y_1^{[a} y_{y_2}^{b]}, \tag{4.14}$$

and are left with the problem of determining $\alpha(x)$ and $\psi(x)$. Equation (4.4), using (4.6), (4.9), (4.11), (4.13), and (4.14), becomes

$$G = I + \psi \left[l^i + \left(\alpha \frac{\overline{M}_1}{L_1} - \alpha \frac{\overline{M}_2}{L_2} + \frac{\overline{M}_2}{L_2} \right) m^i \right] \sigma_i \tag{4.15}$$

(with $\overline{M}_\alpha = \eth L_\alpha$, $\alpha = 1,2$), which, when substituted into the Sparling equation, yields an algebraic equation, which, after a bit of manipulation, can be solved uniquely in the form

$$\psi = \frac{-2(a_2 y_1^2 + a_{12} y_2^2)}{y_1^2 y_2^2 - \sqrt{2} i (a_2 y_1^2 + a_1 y_2^2)}, \tag{4.16}$$

$$\alpha = \frac{a_1 y_2^2}{a_2 y_1^2 + a_1 y_2^2}. \tag{4.17}$$

We then finally have

$$G = I - \frac{2\lambda(x)}{1-\sqrt{2}i\lambda(x)}\left[l^i + \frac{1}{\lambda(x)}\left(\frac{a_1\overline{M}_1}{y_1^2 L_1} + \frac{a_2\overline{M}_2}{y_2^2 L_2}\right)m^i\right]\sigma_i, \tag{4.18}$$

with

$$\lambda(x) = \frac{a_1}{y_1^2} + \frac{a_2}{y_2^2}. \tag{4.19}$$

The procedure used here to obtain the two-(parallel-) instanton solution is easily generalized to the n-(parallel-) instanton solution. If

$$A = \left(\sum_\alpha^n \frac{\alpha_i}{L_\alpha^2}\right) m^i\sigma_i, \tag{4.20}$$

with

$$L_\alpha = (x^a - x_\alpha^a)l_a \equiv y_\alpha^a l_a, \tag{4.21}$$

one can show that the ansatz

$$G = I + \frac{(G_0 l^i + G_- m^i + G_+ \overline{m}^i)\sigma_i}{\Pi_\alpha^n L_\alpha}, \tag{4.22}$$

leads to

$$G = I - \frac{2\lambda}{1-\sqrt{2}i\lambda}\left[l^i + \frac{1}{\lambda}\left(\sum_\alpha^n \frac{a_\alpha}{y_\alpha^2}\frac{M_\alpha}{L_\alpha}\right)m^i\right]\sigma_i, \tag{4.23}$$

where $\lambda = \Sigma_\alpha a_\alpha / y_\alpha^2$.

The fields can now be directly calculated from (4.23) and (1.3).

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under Grant No. PHY82-17853, supplemented by funds from the National Aeronautics and Space Administration, at the University of California at Santa Barbara, and also by a University Research Council Grant No. 568 from Youngstown State University, Youngstown, Ohio.

[1]Richard Ward, "Integrable and Solvable Systems," preprint, Math. Dept. Univ. of Durham, Durham, England.

[2]G. Sparling, University of Pittsburgh report (unpublished).

[3]E. T. Newman, Phys. Rev. D **22**, 3023 (1980).

[4]S. Chakravarty and E. T. Newman, to be published.

[5]C. N. Kozameh and E. T. Newman, Found. Phys. **15**, 487 (1985).

[6]S. Kent and E. T. Newman, J. Math. Phys. **24**, 949 (1983).

[7]M. Atiyah, *The Geometry of Yang–Mills Fields*, Lezione Fermiane (Scuola Normale Superiore, Pisa, 1979).

Nonlinear classical scalar field theory in curved space-time

J. A. Okolowski
Department of Electrical Engineering, Widener University, Chester, Pennsylvania 19013

(Received 25 November 1985; accepted for publication 16 April 1986)

A generic Lagrangian based classical field theory is formulated for any space-time manifold in which certain postulated conditions remain valid. The choice of a specific field Lagrangian leads to a nonlinear model theory that admits a rigorous closed-form particlelike solution in isotropic homogeneous space-time of positive spatial curvature. This metastable solution, of finite positive energy, is discussed in relation to its counterpart in isotropic homogeneous space-time of negative spatial curvature.

I. INTRODUCTION

In isotropic homogeneous space-time of negative spatial curvature, a classical nonlinear model scalar field theory has been constructed for which a rigorous, closed-form, nonsingular, spatially localized solution exists.[1] With this theory is associated a field equation closely resembling the Heisenberg–Klein–Gordon equation.[2] However, due to coupling coefficient dependence on an arbitrary spatial scale factor,[3] covariance of the former field equation under general coordinate transformations is not evident.

The preceding theory is given a manifestly covariant generalization in Sec. II of the present work. There, we introduce a generic Lagrangian-based scalar field theory featuring a scalar coupling coefficient that is identified with the trace of a symmetric tensor field. This tensor field is constrained by postulated conditions that preclude any space-time manifold that does not admit a Ricci tensor having a timelike eigenvector associated with an eigenvalue of a certain type, namely, an eigenvalue specifically expressed in terms of the coupling coefficient.

In the third section of this paper, we confine our attention to the isotropic homogeneous space-time of positive spatial curvature, and we specialize the field Lagrangian to a form suggested by the Heisenberg–Klein–Gordon equation. For such conditions, the present theory admits an exact particlelike[4] solution resulting in a finite positive field energy (particle rest mass). A well-defined flat space-time limit, together with dynamical instability, are salient properties of the latter particlelike solution.

In the final section of this paper, the coupling coefficient is found to remain form-invariant under a group of conformal transformations that generate a restricted class of space-time geometries, including isotropic homogeneous space-time of negative spatial curvature. As evidenced in the remainder of this section, the aforementioned conformal invariance provides the principal connection between the field theory of Ref. 1 and its counterpart in the present report.

II. A MODEL FIELD THEORY

With prescribed metric tensor components $g_{\mu\nu}$, and an arbitrary real scalar ζ having no variational degrees of freedom, the generic field Lagrangian

$$\mathscr{L} = \tfrac{1}{2} g^{\mu\nu}\Psi_{,\mu}\Psi_{,\nu} + |\zeta|\, f(\Psi) \tag{2.1}$$

leads to the field equation

$$g^{\mu\nu}\Psi_{,\mu;\nu} = |\zeta|\, f'(\Psi)\,, \tag{2.2}$$

for the admissible real scalar field Ψ.

To further specialize our theory, we postulate the existence of a real scalar field τ such that its gradient components $\tau_{,\mu}$ form a timelike[5] vector field subject to the equations

$$\tau_{,\mu;\nu} = \tfrac{1}{3}(g_{\mu\nu} - \tau_{,\mu}\tau_{,\nu}) g^{\alpha\beta}\tau_{,\alpha;\beta} \tag{2.3}$$

and

$$g^{\mu\nu}\tau_{,\mu}\tau_{,\nu} = 1\,. \tag{2.4}$$

Next, we introduce the tensor field

$$\zeta_{\mu\nu} \equiv \tfrac{1}{3}(g_{\mu\nu} - \tau_{,\mu}\tau_{,\nu})\zeta\,, \tag{2.5}$$

and we postulate the conservation law

$$(\zeta^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu}\zeta)_{;\nu} = 0\,. \tag{2.6}$$

Observe that the coupling coefficient ζ is the trace of $\zeta_{\mu\nu}$,

$$\zeta = g^{\mu\nu}\zeta_{\mu\nu}\,. \tag{2.7}$$

In consequence of (2.4) and (2.5), Eqs. (2.6) are equivalent to the relations

$$\zeta = \zeta(\tau)\,, \tag{2.8}$$

and

$$g^{\mu\nu}\tau_{,\mu;\nu} = -\tfrac{3}{2}\,[\zeta'(\tau)/\zeta(\tau)]\,. \tag{2.9}$$

A noteworthy feature of both (2.9) and (2.3) is covariance with respect to the transformation

$$g_{\mu\nu} \to h^2 g_{\mu\nu}\,,\quad \tau_{,\mu} \to h\tau_{,\mu}\,,\quad \zeta \to h^{-2}\zeta\,, \tag{2.10}$$

for any differentiable, real-valued function $h = h(\tau) \neq 0$. Moreover, when (2.10) leaves Ψ unaffected, we have the necessary and sufficient condition for covariance of (2.2) under (2.10):

$$g^{\mu\nu}\tau_{,\mu}\Psi_{,\nu} = 0\,. \tag{2.11}$$

Because the validity of (2.11) can be maintained only if Ψ and τ are functionally independent, we do not regard (2.11) as a general restriction on solutions of (2.2).

For any prescribed space-time manifold that admits scalars τ and ζ according to (2.3)–(2.6), the Ricci tensor components $R^{\mu\nu}$ satisfy the equations

$$R^{\mu\nu}\tau_{,\nu} = \Lambda(\tau) g^{\mu\nu}\tau_{,\nu}\,, \tag{2.12}$$

$$\Lambda(\tau) \equiv R^{\mu\nu}\tau_{,\mu}\tau_{,\nu} = -3|\zeta|^{1/2}\frac{d^2}{d\tau^2}|\zeta|^{-1/2}\,. \tag{2.13}$$

 0022-2488/86/082047-04$02.50

These equations arise from the identity

$$(g^{\mu\sigma}\tau_{,\mu;\nu})_{;\sigma} - (g^{\mu\sigma}\tau_{,\mu;\sigma})_{,\nu} \equiv \tau_{,\mu} R^{\mu}{}_{\nu} . \quad (2.14)$$

Inspection of (2.12) shows that the matrix $\|R_{\mu}{}^{\nu}\|$ has an eigenvalue $\Lambda(\tau)$ belonging to an eigenvector with components $\tau_{,\nu}$.

Suppose that in addition to τ and ζ, a space-time manifold admits real scalars τ^* and ζ^*, which differ from τ and ζ, respectively, but otherwise satisfy (2.3), (2.4), (2.8), and (2.9). Then (2.12) and (2.13), together with the corresponding equations for τ^* and ζ^*, imply

$$\Lambda(\tau) = \Lambda^*(\tau^*) \equiv R^{\mu\nu}\tau^*_{,\mu}\tau^*_{,\nu} = -3|\zeta^*|^{1/2}\frac{d^2|\zeta^*|^{-1/2}}{d\tau^{*2}} . \quad (2.15)$$

In view of (2.12), (2.13), and (2.15), we construct the determinant

$$D \equiv \det\{R_{\mu\nu} - \Lambda(\tau)g_{\mu\nu} + Q\tau_{,\mu}\tau_{,\nu}\} , \quad (2.16)$$

where Q denotes any real nonzero scalar having dimensions[6] of $(\text{length})^{-2}$. For a given space-time manifold with $D \neq 0$ at each point, the scalar τ is unique to within a linear transformation

$$\tau \to \tau^* = \pm\tau + \text{const} ; \quad (2.17)$$

consequently, Eq. (2.9) ensures uniqueness of ζ to within an arbitrary multiplicative constant.

In flat space-time, where D vanishes identically, ζ can be unambiguously determined by choosing τ so that $\tau_{,\mu;\nu} = 0$. Thus, for flat space-time, the scalar ζ exists as an arbitrary constant quantity, and, relative to an arbitrary Lorentz frame, we have

$$\tau = \tau(x^0,x^1,x^2,x^3) = n_{\mu}x^{\mu} + \tau_0 ,$$
$$\tau_0 \equiv \tau(0,0,0,0) , \quad (2.18)$$

with any constant timelike vector n_{μ} of unit magnitude.

III. A SOLVABLE NONLINEAR FIELD THEORY

In isotropic homogeneous space-time characterized by the line element[7]

$$ds^2 = dt^2 - a^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)] ,$$
$$a = a(t) , \quad (3.1)$$

the scalar field τ is given by

$$\tau = t + \text{const} . \quad (3.2)$$

Hence, Eqs. (2.5) and (2.9) produce the tensor field

$$\zeta_0{}^0 = 0 , \quad \zeta_1{}^1 = \zeta_2{}^2 = \zeta_3{}^3 = \pm\tfrac{1}{3}(a_0/a)^2 ,$$
$$\zeta_{\mu}{}^{\nu} = 0 , \quad \mu \neq \nu , \quad (3.3)$$

where the arbitrary positive constant a_0 has dimensions of length. According to (2.7) and (3.3) we have the coupling coefficient

$$\zeta = \pm (a_0/a)^2 , \quad (3.4)$$

by reason of which the field equation (2.2), relative to the coordinates (t,χ,θ,ϕ), takes the form

$$a^{-3}(a^3\Psi_{,t})_{,t} - a^{-2}\csc^2\chi[(\Psi_{,\chi}\sin^2\chi)_{,\chi} + \csc\theta(\Psi_{,\theta}\sin\theta)_{,\theta} + \Psi_{,\phi\phi}\csc^2\theta]$$
$$= (a_0/a)^2 f'(\Psi) . \quad (3.5)$$

Notice that (3.4) does not preclude the existence of nontrivial, time-independent solutions to (3.5).

When Ψ is spherically symmetric, Eq. (3.5) reduces to

$$a^{-3}(a^3\Psi_{,t})_{,t} - a^{-2}\csc^2\chi(\Psi_{,\chi}\sin^2\chi)_{,\chi}$$
$$= (a_0/a)^2 f'(\Psi) . \quad (3.6)$$

This equation can be derived from the field Lagrangian

$$\mathscr{L} = \frac{1}{2}(\Psi_{,t})^2 - \frac{1}{2a^2}(\Psi_{,\chi})^2 + \left(\frac{a_0}{a}\right)^2 f(\Psi) ; \quad (3.7)$$

therefore, the field energy associated with a static solution to (3.6) is obtained by evaluating the functional

$$E[\Psi] = 4\pi a \int_0^{\pi} \left\{\frac{1}{2}(\Psi_{,\chi})^2 - a_0^2 f(\Psi)\right\} \sin^2\chi\, d\chi \quad (3.8)$$

at such a solution.

For a theory based on the field Lagrangian (2.1) and the field potential term

$$f(\Psi) \equiv \tfrac{1}{2}\lambda_0\Psi^2 + \tfrac{1}{4}\lambda_1\Psi^4 + \tfrac{1}{6}\lambda_2\Psi^6 , \quad (3.9)$$

in which the positive constants λ_0, λ_1, and λ_2 all have dimensions of $(\text{length})^{-2}$, Eq. (3.6) becomes

$$a^{-3}(a^3\Psi_{,t})_{,t} - a^{-2}\csc^2\chi(\Psi_{,\chi}\sin^2\chi)_{,\chi}$$
$$= (a_0/a)^2(\lambda_0\Psi + \lambda_1\Psi^3 + \lambda_2\Psi^5) . \quad (3.10)$$

The nonlinear field equation (3.10) admits the static, singularity-free, spatially localized solution

$$\Psi_0(\chi) = [(\lambda_1/2\lambda_2)(1-\sigma)]^{1/2}(\tan^2\chi + \sigma)^{-1/2} ,$$
$$0 \leqslant \chi < \pi/2 , \quad (3.11)$$
$$\Psi_0(\chi) = 0 , \quad \pi/2 \leqslant \chi \leqslant \pi ,$$

where

$$\sigma \equiv \tfrac{1}{12}(\lambda_1 a_0)^2/\lambda_2 = \tfrac{1}{3}\lambda_0 a_0^2 . \quad (3.12)$$

Since the scalar field (3.11) must be real valued, we immediately conclude that the dimensionless "size parameter" σ satisfies the inequality

$$0 < \sigma < 1 , \quad (3.13)$$

which restricts the values of λ_0, λ_1, and λ_2 according to $0 < \lambda_0 a_0^2 < 3$, and $0 < (\lambda_1 a_0)^2 < 12\lambda_2$.

Using (3.11)–(3.13) to evaluate the functional (3.8) results in a positive field energy or "particle rest mass"

$$E = \frac{\pi^2}{4}\left(\frac{a}{a_0}\right)\left(\frac{3}{\lambda_2}\right)^{1/2}(1 + 2\sigma^{1/2})(1 - \sigma^{1/2})^2 , \quad (3.14)$$

with

$$\frac{\partial E}{\partial a} = \frac{E}{a} < \frac{\pi^2}{4a_0}\left(\frac{3}{\lambda_2}\right)^{1/2} . \quad (3.15)$$

On account of (3.15), we impose the condition $(\lambda_2 a_0^2)^{1/2} \gg 1$, in order to secure the relation

$$\frac{\partial E}{\partial a} = \frac{E}{a} \ll 1 , \quad (3.16)$$

for all admissible values of σ. The requirement (3.16) is necessitated by our tacit assumption that the scalar field Ψ does not significantly alter the prescribed space-time geometry. In the limiting case of small particle sizes ($0<\sigma\ll 1$), the field energy

$$E\cong\frac{\pi^2}{4}\left(\frac{a}{a_0}\right)\left(\frac{3}{\lambda_2}\right)^{1/2} \tag{3.17}$$

exists as a quantity independent of λ_1.

We make the coordinate transformation

$$\chi=\sin^{-1}(r/a_0)\,,\quad 0\leqslant r<a_0\,, \tag{3.18}$$

whereby the line element (3.1) and the nonzero part of (3.11) are brought to the respective forms

$$ds^2=dt^2-\left(\frac{a}{a_0}\right)^2\left[\left(1-\frac{r^2}{a_0^2}\right)^{-1}dr^2 + r^2(d\theta^2+\sin^2\theta\, d\phi^2)\right], \tag{3.19}$$

and

$$\hat{\Psi}_0(r)=\left[\frac{\lambda_1 a_0^2(1-\sigma)}{2\lambda_2}\right]^{1/2} \times\left(1-\frac{r^2}{a_0^2}\right)^{1/2}[(1-\sigma)r^2+\sigma a_0^2]^{-1/2}. \tag{3.20}$$

By taking the limits $\lambda_1\to 0$, $a_0\to\infty$, and $a/a_0\to 1$ in such a way that $r/a_0\to 0$ for all $r\geqslant 0$, while the positive dimensionless constant $z\equiv\lambda_1 a_0^2$ remains finite but arbitrary, we obtain the flat space-time limits of (3.19), (3.20), and (3.14), respectively,

$$ds_0^2=dt^2-dr^2-r^2(d\theta^2+\sin^2\theta\, d\phi^2)\,, \tag{3.21}$$

$$\Phi_0(r)=\left(\frac{z}{2\lambda_2}\right)^{1/2}\left(r^2+\frac{z^2}{12\lambda_2}\right)^{-1/2}, \tag{3.22}$$

and

$$E_0=(\pi^2/4)(3/\lambda_2)^{1/2}. \tag{3.23}$$

The scalar field (3.22) is a time-independent, particlelike solution of the equation

$$\Phi_{,tt}-r^{-2}(r^2\Phi_{,r})_{,r}=\lambda_2\Phi^5\,, \tag{3.24}$$

in conformity with a previously investigated, Lorentz-covariant classical field theory.[8,9]

Let us now consider the dynamical stability of (3.11) when the perturbed solution is given by

$$\Psi(\chi,t)=\Psi_0(\chi)+\frac{\xi(\chi)}{\sin\chi}\eta(t)\,,$$
$$\xi(\chi)=0\,,\quad \pi/2\leqslant\chi\leqslant\pi\,, \tag{3.25}$$

where the function $\eta(t)$ satisfies the equation[10]

$$\ddot{\eta}+3(\dot{a}/a)\dot{\eta}+(k^2/a^2)\eta=0\,, \tag{3.26}$$

as well as the initial conditions

$$\eta(t_0)=1\,,\quad \dot{\eta}(t_0)=0\,. \tag{3.27}$$

The dimensionless constant k appearing in (3.26) may be either purely real or purely imaginary. For an initially small perturbation

$$|\xi(\chi)|/(\sin\chi)\ll|\Psi_0(\chi)|\,,\quad 0\leqslant\chi<\pi/2\,, \tag{3.28}$$

we substitute (3.25) into (3.10) and retain only terms linear in ξ to derive the eigenvalue equation

$$\xi''(\chi)+F(\chi)\xi(\chi)+(1+3\sigma+k^2)\xi(\chi)=0\,, \tag{3.29}$$

in which

$$F(\chi)=\frac{18\sigma(1-\sigma)}{(\tan^2\chi+\sigma)}+\frac{15\sigma(1-\sigma)^2}{(\tan^2\chi+\sigma)^2}. \tag{3.30}$$

Equation (3.29) must be supplemented with the appropriate boundary conditions for a singularity-free, localized perturbation

$$\xi(0)=\lim_{\chi\to\pi/2}\xi(\chi)=0\,. \tag{3.31}$$

As a direct consequence of (3.29)–(3.31), the negative quantity $11-15/\sigma$ is a lower bound of k^2. That the minimum value of k^2 has a negative upper bound $3(\sigma-1)$ follows from considerations based on the relation[11]

$$G(\chi)\equiv\frac{2(1-\sigma)(7\sigma-1)}{(\tan^2\chi+\sigma)}+\frac{8\sigma(1-\sigma)^2}{(\tan^2\chi+\sigma)^2}<F(\chi)\,, \tag{3.32}$$

in combination with the differential equation

$$u''(\chi)+G(\chi)u(\chi)+2(3\sigma-1)u(\chi)=0\,, \tag{3.33}$$

which admits the solution

$$u(\chi)=(\tan^2\chi+\sigma)^{-1}\tan\chi\,,\quad 0\leqslant\chi<\pi/2\,, \tag{3.34}$$

corresponding to the boundary conditions

$$u(0)=\lim_{\chi\to\pi/2}u(\chi)=0\,. \tag{3.35}$$

Because k is purely imaginary, the perturbation term in (3.25) increases with time in a dynamically unstable manner.[12]

To calculate the approximate minimum value of k^2, we employ a Rayleigh–Ritz procedure. For this purpose, it is convenient to introduce the new independent and dependent variables

$$\rho\equiv\tan^{-1}(\sigma^{-1/2}\tan\chi)\,,\quad 0\leqslant\rho\leqslant\pi/2\,,$$
$$\omega(\rho)\equiv\{\cos^2\rho+\sigma/(1-\sigma)\}^{1/2}\xi(\chi)\,. \tag{3.36}$$

By means of (3.36), we transform (3.29) to the equation

$$\omega''(\rho)+\frac{k^2\sigma\omega(\rho)}{[(1-\sigma)\cos^2\rho+\sigma]^2}-\frac{12\sigma\omega(\rho)}{[(1-\sigma)\cos^2\rho+\sigma]} + 16\omega(\rho)=0\,, \tag{3.37}$$

which leads to the variational principle

$$\delta\gamma^2=0\,,$$
$$\gamma^2\equiv\int_0^{\pi/2}\left\{\omega'(\rho)^2+\frac{12\sigma\omega(\rho)^2}{[(1-\sigma)\cos^2\rho+\sigma]} - 16\omega(\rho)^2\right\}d\rho\,, \tag{3.38}$$

with ω subject to the normalization condition

$$\int_0^{\pi/2}\frac{\omega(\rho)^2\,d\rho}{[(1-\sigma)\cos^2\rho+\sigma]^2}=1\,, \tag{3.39}$$

and the boundary conditions

$$\omega(0)=\omega(\pi/2)=0\,. \tag{3.40}$$

Next, we select a trial function of the form

$$\omega(\rho) = (4/\pi)^{1/2}[(1-\sigma)\cos^2\rho + \sigma] \times(\alpha \sin 2\rho + \beta \sin 4\rho), \quad (3.41)$$

where α and β are variational parameters constrained by (3.39) to satisfy $\alpha^2+\beta^2=1$. By minimizing the result of combining (3.41) with the definition part of (3.38), and using

$$\gamma^2 \equiv k^2\sigma, \quad (3.42)$$

we establish an approximate functional relation between σ and the minimum value of k^2:

$$k^2\sigma \cong -\tfrac{1}{2}\{[3(1+\sigma)^2+\tfrac{1}{2}(1-\sigma)^2]^2+9(1-\sigma)^4\}^{1/2} + \tfrac{3}{2}(1+\sigma)^2+\tfrac{1}{4}(1-\sigma)^2-3(1-\sigma^2). \quad (3.43)$$

From (3.43) we infer that $(-k^2\sigma)^{1/2}$ is a monotonically decreasing function of σ with

$$\lim_{\sigma\to 0}(-k^2\sigma)^{1/2} \cong 1.88, \quad (3.44)$$

and

$$\lim_{\sigma\to 1}(-k^2\sigma)^{1/2} = 0. \quad (3.45)$$

IV. CONFORMAL INVARIANCE OF THE COUPLING COEFFICIENT AND RELATED CONSIDERATIONS

By virtue of the coordinate transformation

$$\bar{\chi} = \ln(\tan\chi + \sec\chi), \quad 0 \leqslant \chi < \pi/2, \quad (4.1)$$

the nonzero part of (3.11) and the line element (3.1) are transformed, respectively, as follows:

$$\Psi_0(\chi) = \bar{\Psi}_0(\bar{\chi};\lambda_1 a_0^2,\lambda_2 a_0^2) \equiv \left[\frac{\lambda_1}{2\lambda_2}\left(1-\frac{\lambda_1^2 a_0^2}{12\lambda_2}\right)\right]^{1/2}\left(\sinh^2\bar{\chi}+\frac{\lambda_1^2 a_0^2}{12\lambda_2}\right)^{-1/2}, \quad (4.2)$$

and

$$ds^2 = dt^2 - \frac{a^2}{\cosh^2\bar{\chi}}[d\bar{\chi}^2+\sinh^2\bar{\chi}(d\theta^2+\sin^2\theta\, d\phi^2)]. \quad (4.3)$$

It is easily verified that (3.2)–(3.4) satisfy (2.3)–(2.6) for any space-time metric generated from (4.3) by means of the conformal transformation

$$dl^2 \equiv \frac{a^2}{\cos^2\bar{\chi}}[d\bar{\chi}^2+\sinh^2\bar{\chi}(d\theta^2+\sin^2\theta\, d\phi^2)] \to \kappa(\bar{\chi},\theta,\phi)\,dl^2, \quad (4.4)$$

where the function κ is positive valued and differentiable, but otherwise arbitrary. Therefore, the coupling coefficient (3.4) remains form invariant under (4.4). We choose

$$\kappa(\bar{\chi},\theta,\phi) = \cosh^2\bar{\chi} \quad (4.5)$$

to obtain the line element associated with isotropic homogeneous space-time of negative spatial curvature

$$ds^2 = dt^2 - a^2[d\bar{\chi}^2+\sinh^2\bar{\chi}(d\theta^2+\sin^2\theta\, d\phi^2)]. \quad (4.6)$$

Let (4.6) specify the space-time geometry for a field theory of the type described in Sec. II when the field potential term is given by (3.9). Then, with the aid of (3.2) and the ensuing conformal invariance of (3.4), the field equation (2.2) acquires the form featured in Ref. 1,

$$a^{-3}(a^3\Psi_{,t})_{,t} - a^{-2}\,\mathrm{csch}^2\bar{\chi}[(\Psi_{,\bar{\chi}}\sinh^2\bar{\chi})_{,\bar{\chi}} + \csc\theta(\Psi_{,\theta}\sin\theta)_{,\theta} + \Psi_{,\phi\phi}\csc^2\theta] = (a_0/a)^2(\lambda_0\Psi+\lambda_1\Psi^3+\lambda_2\Psi^5). \quad (4.7)$$

Equation (4.7) admits the nonsingular, time-independent, spherically symmetric solution

$$\tilde{\Psi}_0(\bar{\chi}) \equiv 3^{1/2}\bar{\Psi}_0\left(\bar{\chi};9\lambda_1 a_0^2\left[1+\frac{3\lambda_1^2 a_0^2}{4\lambda_2}\right]^{-1}, 9\lambda_2 a_0^2\left[1+\frac{3\lambda_1^2 a_0^2}{4\lambda_2}\right]^{-1}\right) = \left(\frac{2}{\lambda_1 a_0^2}\right)^{1/2}\left(1+\frac{4\lambda_2}{3\lambda_1^2 a_0^2}\right)^{-1/2} \times\left(\sinh^2\bar{\chi}+\left[1+\frac{4\lambda_2}{3\lambda_1^2 a_0^2}\right]^{-1}\right)^{-1/2}, \quad (4.8)$$

provided that $\lambda_0 a_0^2 = 1$. Here, λ_0 is independent of both λ_1 and λ_2 in contrast to the relation $\lambda_0 = \lambda_1^2/4\lambda_2$, which follows from (3.12).

In terms of the dimensionless "size parameter"

$$\tilde{\sigma} \equiv (1+4\lambda_2/3\lambda_1^2 a_0^2)^{-1}, \quad (4.9)$$

we write

$$\tilde{\Psi}_0(\bar{\chi}) = (2\tilde{\sigma}/\lambda_1 a_0^2)^{1/2}(\sinh^2\bar{\chi}+\tilde{\sigma})^{-1/2}. \quad (4.10)$$

The particlelike scalar field (4.10) has the flat space-time limit

$$\Theta_0(r) \equiv \left(\frac{3z}{2\lambda_2}\right)^{1/2}\left(r^2+\frac{3z^2}{4\lambda_2}\right)^{-1/2} = 3^{-1/2}\Phi_0\left(\frac{r}{3}\right). \quad (4.11)$$

Furthermore, (4.11) satisfies the field equation (3.24), which, under time-independent conditions, is invariant with respect to the scale transformation

$$\Phi(r) \to \epsilon^{1/2}\Phi(\epsilon r), \quad (4.12)$$

for all real $\epsilon > 0$. Finally, analysis reveals[1] that $\tilde{\Psi}_0(\bar{\chi})$ exists as a dynamically unstable field associated with non-negative energy only if $0 < \tilde{\sigma} \leqslant \tilde{\sigma}_0 \cong 0.142$.

[1] J. A. Okolowski, J. Math. Phys. **19**, 2253 (1978); **20**, 1834(E) (1979).
[2] E. W. Mielke, J. Math. Phys. **21**, 543 (1980).
[3] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 718.

[4] In the present work, the term "particlelike" is equivalent to the phrase "spatially localized and singularity-free."
[5] The metric tensor has signature $(+,-,-,-)$.
[6] Units are chosen so that the constant of gravitation and the speed of light in a vacuum are both equal to unity.
[7] The ranges of the coordinates (χ,θ,ϕ) are given by $0 \leqslant \chi \leqslant \pi$, $0 \leqslant \theta \leqslant \pi$, and $0 \leqslant \phi < 2\pi$.

[8] G. Rosen, J. Math. Phys. **6**, 1269 (1965).
[9] G. Rosen, J. Math. Phys. **7**, 2066 (1966).

[10] In what follows, we assume that $\dot{a}(t)$ is a continuous function of t for $t > t_0$.

[11] E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., London, 1962), Part I, 2nd ed., p. 113.
[12] G. Birkhoff and G. C. Rota, *Ordinary Differential Equations* (Blaisdell, Waltham, MA, 1969), 2nd ed., pp. 39–41.

Constraints, reduction, and quantization

Mark J. Gotay
Mathematics Department, United States Naval Academy, Annapolis, Maryland 21402

(Received 23 July 1985; accepted for publication 12 March 1986)

Theorems are proved that establish the unitary equivalence of the extended and reduced phase space quantizations of a constrained classical system with symmetry. Several examples are presented.

I. INTRODUCTION

Among classical dynamical systems, those which are "constrained" are often the most important and interesting. Typically, constraints arise when the equations of motion are overdetermined or when symmetries are present. In the first circumstance the constraints take the form of restrictions on the admissible initial data for the evolution equations of the system. The divergence constraints of electromagnetism and Yang–Mills theory and the super-Hamiltonian and supermomentum constraints of general relativity are standard examples. In the second case the constraints consist of *a posteriori* specifications of the constants of motion associated with the invariances of the system. The mass and charge constraints in the Kaluza–Klein formalism for a relativistic charged particle are of this type, as is, for example, fixing the angular momentum of a rotationally invariant system.

All constrained systems can be described naturally in terms of symplectic geometry.[1,2] Beyond this, however, many such systems have a rich group-theoretical structure. It is an amazing fact that the constraints are usually given by $J = \text{const}$, where J is a momentum mapping for an appropriately chosen group action.[3] These observations lead us to model a constrained dynamical system as follows.

Let (X,ω) be a symplectic manifold that represents the "extended" phase space of a system. Suppose that G is a Lie group which acts symplectically on (X,ω) and that $J: X \to \mathscr{g}^*$ is a momentum mapping for this action, where $\mathscr{g}$ is the Lie algebra of G. We interpret G as a "symmetry" or "gauge" group; J is the corresponding conserved quantity. A *constrained classical system with symmetry* is given by (X,ω,G,J) along with a fixed choice of $\mu \in \mathscr{g}^*$. The constraints are then $J = \mu$ and $J^{-1}(\mu) \subset X$ is the *constraint set.*

One may reduce the number of degrees of freedom of a constrained system by factoring out the symmetries of the constraint set. Subject to certain technical assumptions, Marsden and Weinstein[4] showed that the resulting orbit space $\overline{X}_\mu$ is a quotient manifold of $J^{-1}(\mu)$ and inherits a symplectic structure $\overline{\omega}_\mu$ from that on X. The symplectic manifold $(\overline{X}_\mu,\overline{\omega}_\mu)$ is the *reduced phase space* of invariant states of the system.

There are thus two symplectic manifolds associated to each constrained system: the extended and reduced phase spaces (X,ω) and $(\overline{X}_\mu,\overline{\omega}_\mu)$, respectively. *Classically,* there is no formal distinction between working on (X,ω) while carrying along the constraints versus solving the constraints, reducing the system and working on $(\overline{X}_\mu,\overline{\omega}_\mu)$. But these two approaches are not necessarily equivalent on the *quantum* level. This was recently emphasized by Ashtekar and Horowitz,[5] who showed that these two classical formalisms may engender real and significant physical differences in the quantum behavior of the system.

Recall that quantization associates to a phase space (X,ω) a Hilbert space $\mathscr{H}$ of quantum states and to some class of smooth functions f on X quantum operators $\mathscr{Q}f$ on $\mathscr{H}$. For a constrained classical system one may, as indicated above, quantize either the extended or the reduced phase space. The purpose of this paper is to determine under what conditions and in what sense these two quantizations will be equivalent.

We first consider the extended phase space quantization following Dirac.[6] The essential idea is that as the constraints have not been eliminated classically, they must be enforced quantum mechanically. This is possible if quantization provides a representation of $\mathscr{g}$ on $\mathscr{H}$. Since the constraints are given classically by $J = \mu$, it follows that the physically admissible quantum states are those which belong to the subspace $\mathscr{H}_\mu$ of $\mathscr{H}$ defined by

$$\mathscr{H}_\mu = \{\Psi \in \mathscr{H} \mid \mathscr{Q}J[\Psi] = \mu\Psi\}.$$

The situation is somewhat simpler for the reduced phase space $(\overline{X}_\mu,\overline{\omega}_\mu)$ as the constraints have already been solved and the symmetries divided out. There are no restrictions to be imposed on the quantum system and so, by construction, the associated Hilbert space $\overline{\mathscr{H}}_\mu$ consists of all the physically admissible states of the system.

These two quantizations each yield spaces of "physically admissible quantum states" which in general will not coincide. We may thus phase our question as follows: *When will $\mathscr{H}_\mu$ and $\overline{\mathscr{H}}_\mu$ be unitarily isomorphic?* There are three sets of obstructions to the existence of such an isomorphism, involving (i) the naturality of the extended phase space quantization, (ii) the compatibility of the extended and reduced phase space quantizations, and (iii) the unitary relatedness of the Hilbert space structures on $\mathscr{H}_\mu$ and $\overline{\mathscr{H}}_\mu$.

The first impediment is whether in fact the quantization of the extended phase space gives rise to a representation of the Lie algebra $\mathscr{g}$ of G on $\mathscr{H}$. A necessary condition is that $J^{-1}(\mu)$ be a coisotropic submanifold of (X,ω). This ensures the internal consistency of the quantization and effectively restricts the allowable values of $\mu \in \mathscr{g}^*$.

The next difficulty is to properly correlate the quantizations of the extended and reduced phase spaces. This can be accomplished by requiring that the auxiliary structures on (X,ω) necessary for quantization be G-invariant—provided this is possible—for they will then project to compatible quantization structures on $(\overline{X}_\mu,\overline{\omega}_\mu)$.

 0022-2488/86/082051-16$02.50

Physically, the above obstructions take the form of "quantization conditions" and/or "superselection rules" and place restrictions on the topology of G as well as the choice of quantization structures. Once they have been overcome, one obtains "smooth" quantizations of (X,ω) and $(\overline{X}_\mu,\overline{\omega}_\mu)$, i.e., linear spaces of C^∞ wave functions $\mathscr{H}$ and $\overline{\mathscr{H}}_\mu$, respectively. The final obstructions appear when one introduces the quantum inner products on these spaces. It may happen that $\overline{\mathscr{H}}_\mu$ does not inherit an inner product from $\mathscr{H}$ and, when it does, it must be checked that an equivalence of the underlying smooth quantizations extends to a unitary isomorphism of the corresponding Hilbert spaces.

Substantial progress towards answering this question has already been made by Gotay and Sniatycki,[7] Guillemin and Sternberg,[8] Puta,[9] Sniatycki,[10] Vaisman,[11] and Woodhouse.[12] Because of the intricacy of the problem and the vagaries of the quantization process, however, it is difficult to obtain results in a completely general setting that provide explicit information about concrete systems.

To rectify this, we concentrate in this paper on one specific class of constrained systems—those whose phase spaces are cotangent bundles and whose groups act by point transformations. There are numerous reasons for considering such systems.

(1) They are the most common and hence the most important physically. Indeed, all of the examples cited earlier—with the exception of the mass constraint in the Kaluza–Klein theory—fall into this class.

(2) Cotangent bundles, along with Kähler manifolds, are exceptional examples of symplectic manifolds as they have naturally defined polarizations (the vertical and antiholomorphic ones, respectively); this is a crucial advantage insofar as quantization is concerned. Guillemin and Sternberg[8] have studied the Kähler case and so the results we present here are, to some extent, complementary to theirs.

(3) Reduction keeps us within the cotangent bundle category: subject to certain assumptions (which are in any case necessary for quantization), the reduced phase space will also be a cotangent bundle. We may therefore quantize both the extended and reduced phase spaces using the corresponding vertical polarizations. This means, in physicists' terminology, that we always quantize in the "Schrödinger representation."

(4) We are able to obtain relatively "hard" results. Namely, we can explicitly identify and construct the momentum mapping, the reduced phase space, and all of the required quantization structures. The formalism we develop will also enable us to detail precisely the various obstructions discussed earlier as well as verify directly whether the assumptions we impose are satisfied in specific cases. Thus our general problem is reduced to a conceptually and computationally much simpler one.

The plan of attack is as follows. We consider systems of the form (T^*Q,ω,G,J,μ), where G acts on T^*Q by pullback and μ is Ad*-invariant. Applying the reduction technique of Kummer,[13] Satzer,[14] and Abraham and Marsden,[15] we show that the reduced phase space is symplectomorphic to $T^*(Q/G)$ (with a possibly noncanonical symplectic structure). These results are summarized in Sec. II.

Sections III and IV form the heart of the paper. After discussing some generalities on the quantization of constrained systems we quantize both the extended and reduced phase spaces. In particular, we show that quantization does indeed yield a representation of the symmetry algebra $\mathscr{g}$.

In the next section we construct a canonical unitary isomorphism between the two quantizations obtained in Sec. III. We also prove that it is possible to quantize invariant polarization-preserving functions in either formalism with equivalent results.

The following section presents several examples and we conclude with a discussion of possible generalizations of our results.

II. CONSTRAINED CLASSICAL SYSTEMS

We begin by reviewing some basic facts about group actions, momentum mappings, and reduction. The main references for what follows are Refs. 4, 15, and 16.

A. Hamiltonian G-spaces

Let G be a connected Lie group with Lie algebra $\mathscr{g}$ and let $\Phi: G\times Q\to Q$ be a smooth action of G on a manifold Q. For each $\zeta\in\mathscr{g}$, we denote by ζ_Q the corresponding infinitesimal generator on Q. The orbit of a point $q\in Q$ is written $G\cdot q$. Recall that when Φ is free and proper the orbit space $\overline{Q} = Q/G$ is a Hausdorff quotient manifold of Q and, furthermore, $\pi_Q: Q\to\overline{Q}$ is a left principal G-bundle.

Now suppose (X,ω) is a symplectic manifold on which G acts symplectically. A *momentum mapping* for this action is a map $J: X\to\mathscr{g}^*$ such that, for each $\zeta\in\mathscr{g}$, the associated function $J_\zeta(x) = \langle J(x),\zeta\rangle$ satisfies

$$\zeta_X \,\lrcorner\, \omega = -dJ_\zeta\ . \tag{2.1}$$

Then J is *Ad*-equivariant* provided

$$J(\Phi_g(x)) = \mathrm{Ad}^*_{g^{-1}}J(x)\ , \tag{2.2}$$

for all $g\in G$, where Ad* is the coadjoint action of G on $\mathscr{g}^*$. If an Ad*-equivariant momentum map J exists for the action Φ, we call (X,ω,G,J) a *Hamiltonian G-space*.

Let $\mu\in\mathscr{g}^*$ be a weakly regular value of J, so that the level set $J^{-1}(\mu)$ is a manifold with $TJ^{-1}(\mu) = \ker TJ$. The following result relates the geometry of $J^{-1}(\mu)$ with that of the orbits of G and G_μ, where G_μ is the isotropy group of μ under the coadjoint action.

Proposition (2.1): For $x\in J^{-1}(\mu)$,

(i) $T_x(G_\mu\cdot x) = T_x(G\cdot x)\cap T_xJ^{-1}(\mu)$,

and

(ii) $T_x(J^{-1}(\mu))^\perp = T_x(G\cdot x)$.

Here "$\perp$" denotes the ω-orthogonal complement.

By equivariance, $J^{-1}(\mu)$ is stable under the action of G_μ so that the orbit space $\overline{X}_\mu = J^{-1}(\mu)/G_\mu$ is well defined. Let $j_\mu: J^{-1}(\mu)\to X$ be the inclusion and $\pi_\mu: J^{-1}(\mu)\to\overline{X}_\mu$ the projection. The next result, due to Marsden and Weinstein,[4] is central to the theory.

Theorem (Marsden–Weinstein reduction): Let (X,ω,G,J) be a Hamiltonian G-space. If $\mu\in\mathscr{g}^*$ is a weakly regular value of J and the action of G_μ on $J^{-1}(\mu)$ is free and proper, then there exists a unique symplectic structure $\overline{\omega}_\mu$ on the manifold $\overline{X}_\mu$ such that $\pi_\mu^* \overline{\omega}_\mu = j_\mu^* \omega$.

Remark: If the Marsden–Weinstein reduction procedure fails [e.g., μ is not weakly regular or $(\overline{X}_\mu, \overline{\omega}_\mu)$ does not exist as a smooth symplectic manifold, as is the case in a number of important examples], one can still reduce $J^{-1}(\mu)$ on the level of Poisson algebras.[17]

B. The cotangent category

For the reasons cited in the Introduction, we restrict attention to constrained classical systems that belong to the "cotangent bundle category." Reduction in this category was first carried out by Satzer[14] for $\mu = 0$ and then extended to $\mu \neq 0$ by Abraham and Marsden[15] and Marsden.[16] Subsequently, Kummer[13] improved upon these results and put the theory into its present form. Our presentation is drawn from both Refs. 13 and 15.

Suppose that the system has configuration space Q and symmetry group G. We assume that G carries a bi-invariant metric and that Φ is a free and proper left action of G on Q. Let the extended phase space be $X = T^*Q$, where τ_Q is the cotangent bundle projection and ω and Θ denote the canonical two- and one-forms on T^*Q, respectively, with $\omega = d\Theta$. The induced symplectic action $T^*\Phi: G \times T^*Q \to T^*Q$, given by

$$T^*\Phi(g,\beta) = \Phi_{g^{-1}}{}^*\beta ,$$

is also free and proper. There is a natural Ad*-equivariant momentum map for $T^*\Phi$ defined by

$$\langle J(\beta),\xi \rangle = \Theta(\xi_{T^*Q}(\beta)) = \beta(\xi_Q) . \tag{2.3}$$

We refer to the Hamiltonian G-space (T^*Q,ω,G,J) so defined, along with a fixed choice of $\mu \in \mathscr{g}^*$, as a *constrained cotangent system*.

Regarding reduction, one of the main advantages of our formalism is the following proposition.

Proposition (2.2): Every $\mu \in \mathscr{g}^*$ is a regular value of J.

Proof: Suppose that $T_\beta J$ was not surjective for some $\beta \in T^*Q$, in which case there exists $\xi \in \mathscr{g}$ such that $\langle T_\beta J(v),\xi \rangle = 0$ for all $v \in T_\beta(T^*Q)$. Then (2.1) yields $\omega(\xi_{T^*Q}(\beta),v) = 0$ for all v and nondegeneracy implies that $\xi_{T^*Q}(\beta) = 0$, contradicting the fact that $T^*\Phi$ is free. ■

Each level set $J^{-1}(\mu)$ is therefore an imbedded submanifold of T^*Q. Furthermore, since $T^*\Phi$ is free and proper and G_μ is closed in G, the action of G_μ on $J^{-1}(\mu)$ is also free and proper. This observation, combined with Proposition (2.2) and the Marsden–Weinstein Theorem, give the following proposition.

Proposition (2.3): $J^{-1}(\mu)$ is reducible for every $\mu \in \mathscr{g}^*$.

Insofar as the quantization of these systems is concerned, however, it is not necessary to consider general $\mu \in \mathscr{g}^*$. We will see in Sec. III B that only those μ which are "invariant" are relevant.

Definition: $\mu \in \mathscr{g}^*$ is *invariant* if $\mathrm{Ad}_g^*(\mu) = \mu$ for all $g \in G$.

Equivalently, μ is invariant iff $G_\mu = G$. For such μ the reduction of $J^{-1}(\mu)$ is particularly simple and elegant. We first reduce $J^{-1}(0)$ and then transform the case $\mu \neq 0$ to this; note that $\mu = 0$ is always invariant.

Let $\overline{\omega} = d\overline{\Theta}$ be the canonical symplectic structure on $T^*\overline{Q}$, where $\overline{Q} = Q/G$.

Proposition (2.4): The reduced phase space $(\overline{T^*Q}_0, \overline{\omega}_0)$ is symplectomorphic to $(T^*\overline{Q},\overline{\omega})$.

Proof: First note that the pullback bundle

$$\pi_Q^*(T^*\overline{Q}) = J^{-1}(0) , \tag{2.4}$$

where $\pi_Q: Q \to \overline{Q}$ is the canonical submersion. Indeed, since $T\pi_Q(\xi_Q) = 0$, a 1-form β on Q belongs to $\pi_Q^*(T^*\overline{Q})$ iff $\beta(\xi_Q) = 0$ iff $\beta \in J^{-1}(0)$ by (2.3). Quotienting by G in (2.4) then gives $T^*\overline{Q} \approx \overline{T^*Q}_0$.

It remains to show that the reduced symplectic form $\overline{\omega}_0$ on $\overline{T^*Q}_0$ can be identified with $\overline{\omega}$ on $T^*\overline{Q}$. Using (2.4) and the induced commutative diagram

$$\begin{tikzcd} J^{-1}(0) \arrow[r, "j_0"] \arrow[d, "\pi_0"'] & T^*Q \arrow[d, "\pi_Q \circ \tau_Q"] \\ T^*\overline{Q} \arrow[r, "\tau_{\overline{Q}}"] & \overline{Q} \end{tikzcd} ,$$

a straightforward computation establishes $\pi_0^*\overline{\Theta} = j_0^*\Theta$ and consequently $\pi_0^*\overline{\omega} = j_0^*\omega$. The result now follows from the uniqueness of the reduced symplectic structure in the Marsden–Weinstein Theorem. ■

When μ is nonzero but invariant, we first choose a left connection α on the left principal G-bundle $\pi_Q: Q \to \overline{Q}$. Set $\alpha_\mu = \mu \circ \alpha$. Then α_μ is G-invariant and, viewed as a one-form on Q, takes values in $J^{-1}(\mu)$. Construct the invariant symplectic form $\Omega_\mu = \omega + \tau_Q^* \, d\alpha_\mu$ on T^*Q. The key step in the transition from $\mu = 0$ to $\mu \neq 0$ is the following proposition.

Proposition (2.5): There exists a G-equivariant presymplectomorphism of $(J^{-1}(0), j_0^*\Omega_\mu)$ with $(J^{-1}(\mu), j_\mu^*\omega)$.

Proof: Define a diffeomorphism δ_μ of T^*Q by

$$\delta_\mu(\beta) = \beta + \alpha_\mu(\tau_Q(\beta)) . \tag{2.5}$$

Since α_μ is invariant δ_μ is equivariant and, as $J(\alpha_\mu) = \mu$, δ_μ induces a diffeomorphism $J^{-1}(0) \to J^{-1}(\mu)$, which we also denote by δ_μ.

Now δ_μ is just translation along the fibers, so

$$\delta_\mu^*\Theta = \Theta + \tau_Q^*\alpha_\mu$$

and hence $\delta_\mu^*\omega = \Omega_\mu$. But this and the relation $j_\mu \circ \delta_\mu = \delta_\mu \circ j_0$ imply that δ_μ is a presymplectomorphism. ■

Propositions (2.4) and (2.5) enable us to identify the reduced manifolds $\overline{T^*Q}_\mu$ for μ invariant with $T^*\overline{Q}$. To complete the reduction we have only to compute the reduced symplectic forms $\overline{\omega}_\mu$.

Lemma (2.6): There exists a closed two-form F_μ on $\overline{Q}$ such that $\pi_Q^* F_\mu = d\alpha_\mu$.

Proof: We first claim that $d\alpha_\mu = \mu \circ D\alpha$, where $D\alpha$ is the curvature of the connection α. To prove this, take the μ component of the Cartan structure equation

$$d\alpha(u,v) = [\alpha(u),\alpha(v)] + D\alpha(u,v)$$

and observe that

$$\begin{aligned} \mu \circ [\alpha(u),\alpha(v)] &= \mu(\mathrm{ad}_{\alpha(u)}\alpha(v)) \\ &= (\mathrm{ad}^*_{\alpha(u)}\mu)\,(\alpha(v)) \end{aligned}$$

vanishes as μ is invariant. Thus $d\alpha_\mu$ is horizontal. Since in addition $d\alpha_\mu$ is invariant and π_Q is a submersion, $d\alpha_\mu$ projects to a two-form F_μ on $\overline{Q}$ with the required properties. ■

The reduction of $(J^{-1}(0), j_0^*\Omega_\mu)$ is clearly $(T^*\overline{Q}, \overline{\Omega}_\mu)$, where

$$\overline{\Omega}_\mu = \overline{\omega} + \tau_Q^* F_\mu . \tag{2.6}$$

Taking Proposition (2.5) into account and noting that, by equivariance, δ_μ passes to the quotient, we have proven the following theorem.

Theorem (Kummer–Marsden–Satzer reduction): Consider the constrained cotangent system (T^*Q,ω,G,J,μ). If $\mu\in\mathscr{g}^*$ is invariant then each choice of connection on Q defines a symplectomorphism between the reduced phase space $(\overline{T^*Q}_\mu,\overline{\omega}_\mu)$ and $(T^*\overline{Q},\overline{\Omega}_\mu)$, where $\overline{Q}=Q/G$ and $\overline{\Omega}_\mu$ is defined by (2.6).

In essence, the reduction of a cotangent bundle is again a cotangent bundle. This fact will be of paramount importance in the sequel.

Remarks: (1) It is crucial here that μ be invariant. When $G_\mu \subset G$, $\overline{T^*Q}_\mu$ can only be identified with a symplectic *subbundle* of $T^*(Q/G_\mu)$ (cf. Refs. 13 and 15). In other words, the invariance of μ is *necessary as well as sufficient* for the reduced manifold to be a cotangent bundle.

(2) In general the symplectic structure $\overline{\Omega}_\mu$ on $T^*\overline{Q}$ will *not* be canonical due to the presence of the curvature term F_μ. This "extra" term has been used as a means of introducing Yang–Mills-type interactions (see, e.g., Refs. 18 and 19 and the references contained therein for the details of this; we shall encounter an instance of this phenomenon in our study of the Kaluza–Klein theory in Sec. V C). However, when Q carries a flat connection we may take $\overline{\Omega}_\mu$ to be exact.

(3) Although the proof of the Kummer-Marsden–Satzer Theorem required choosing a connection, this choice is irrelevant. Other such choices simply lead to different, but nonetheless symplectomorphic, realizations of $(\overline{T^*Q}_\mu, \overline{\omega}_\mu)$. It is also possible to derive this theorem using a Riemannian metric to obtain the required invariant $J^{-1}(\mu)$-valued one-form α_μ (cf. Refs. 14 and 15). This approach seems more cumbersome and less "physical" than the one employed here, which is due to Kummer.[13]

(4) Montgomery[20] has recently shown, subject to certain additional assumptions, that the Kummer–Marsden–Satzer reduction procedure may be extended to the case when Φ is not free.

We close this section by noting that we may also reduce observables: if $f\in C^\infty(T^*Q)$ is invariant, then it projects to $\overline{T^*Q}_\mu$. To describe this function on $T^*\overline{Q}$, set $f_\mu = f\circ\delta_\mu$ and define $\overline{f}_\mu\in C^\infty(T^*\overline{Q})$ by $\overline{f}_\mu\circ\pi_0 = f_\mu$. Since $f_\mu\circ j_0 = f\circ j_\mu$, it follows that $\overline{f}_\mu$ represents the reduced observable. In particular, if h is a Hamiltonian on (T^*Q,ω) then $\overline{h}_\mu$ is the "amended" Hamiltonian on $(T^*\overline{Q},\overline{\Omega}_\mu)$ (cf. Refs. 14–16).

III. QUANTIZATION

To properly address the subtleties and complexities of the transition from the classical to the quantal domain it is essential to use a well-defined quantization technique. We choose the geometric quantization framework of Kostant and Souriau because it is formulated in terms of symplectic geometry. In Sec. III A we briefly outline those elements of this theory that are needed here, referring the reader to Sniatycki[21] and Woodhouse[12] for comprehensive expositions.

A. Quantization structures

Let (X,ω) be a $2n$-dimensional symplectic manifold. The supplementary structures needed for the geometric quantization of (X,ω) are a polarization,a prequantization line bundle, and a metalinear frame bundle.

A (real) *polarization* of (X,ω) is an involutive n-dimensional distribution P on X such that $P^\perp = P$.

A *prequantization* of (X,ω) consists of a complex line bundle $l: L\to X$ with a connection ∇ such that

$$\text{curvature } \nabla = -(1/h)\, l^*\omega , \tag{3.1}$$

where h is Planck's constant. A prequantization of (X,ω) exists iff the de Rham class of ω/h in $H^2(X,\mathbb{R})$ is integral and, if nonempty, the set of all prequantizations is parametrized up to equivalence by a principal homogeneous space for the character group of $\pi_1(X)$.

Remark: We make no distinction between L and its associated principal $\mathbb{C}^*$-bundle.

Fix a polarization P of (X,ω) and let FP be the linear frame bundle of P. It is a right principal $GL(n,\mathbb{R})$-bundle over X. Let $ML(n,\mathbb{R})$ be the $n\times n$ metalinear group, that is, the set of all matrices of the form

$$\widetilde{M} = \begin{pmatrix} M & 0 \\ 0 & z \end{pmatrix},$$

where $M\in GL(n,\mathbb{R})$ and $z^2=\det M$. A *metalinear frame bundle* for P is a right principal $ML(n,\mathbb{R})$-bundle $\widetilde{F}P$ over X along with a 2:1 projection $\rho: \widetilde{F}P\to FP$ such that the diagram

$$\begin{array}{ccc} \widetilde{F}P\times ML(n,\mathbb{R}) & \longrightarrow & \widetilde{F}P \\ \big\downarrow \rho\times\sigma & & \big\downarrow \rho \\ FP\times GL(n,\mathbb{R}) & \longrightarrow & FP \end{array}$$

commutes, where the horizontal arrows are the group actions and $\sigma: ML(n,\mathbb{R})\to GL(n,\mathbb{R})$ is the twofold projection $\widetilde{M}\mapsto M$.

The existence of a metalinear frame bundle is equivalent to the vanishing of a class in $H^2(X,\mathbb{Z}_2)$ characteristic of FP and, if nonempty, the set of all such is parametrized up to equivalence by $H^1(X,\mathbb{Z}_2)$.

Remark: We need not consider the more general metaplectic structures here since we will not be moving polarizations.

Let $\Delta: ML(n,\mathbb{R})\to\mathbb{C}$ be the unique holomorphic square root of the determinant function on $GL(n,\mathbb{R})$ such that $\Delta(\widetilde{I})=1$, where $\widetilde{I}$ is the identity. The bundle $\sqrt{\wedge^n P}$ of *half-forms* relative to P is the bundle associated to $\widetilde{F}P$ with typical fiber $\mathbb{C}$ on which $ML(n,\mathbb{R})$ acts by multiplication by Δ. This bundle has a canonically defined partial flat connection covering P. Denote by $\Gamma(\sqrt{\wedge^n P})$ the space of all smooth sections of $\sqrt{\wedge^n P}$. Each $\nu\in\Gamma(\sqrt{\wedge^n P})$ can be identified with a function $\nu^\#: \widetilde{F}P\to\mathbb{C}$ satisfying

$$\nu^\#(\widetilde{f}\widetilde{M}) = \Delta(\widetilde{M})^{-1}\cdot\nu^\#(\widetilde{f}) \tag{3.2}$$

for all metaframes $\widetilde{f}\in\widetilde{F}P$ and $\widetilde{M}\in ML(n,\mathbb{R})$.

Consider the bundle $L\otimes\sqrt{\wedge^n P}$. It carries a partial flat connection covering P induced from those on L and $\sqrt{\wedge^n P}$. A section $\Psi\in\Gamma(L\otimes\sqrt{\wedge^n P})$ is said to be *polarized* if it is covariantly constant along P. Let $\mathscr{H}$ be the subspace of $\Gamma(L\otimes\sqrt{\wedge^n P})$ consisting of polarized sections. Elements of

$\mathscr{H}$ are interpreted as smooth quantum wave functions, i.e., $\mathscr{H}$ is the *smooth* quantum state space associated to (X,ω) by the geometric quantization procedure in the representation defined by the polarization P.

We now turn to the quantization of classical observables $f \in C^\infty(X)$. Suppose f preserves P in the sense that $T\phi^t(P) = P$ for all $t \in \mathbf{R}$, where ϕ^t is the flow of (the Hamiltonian vector field of) f, which we assume is complete. Then f is quantizable as a first-order linear differential operator $\mathscr{Q}f$ on $\mathscr{H}$. The mechanics of this are as follows. The flow ϕ^t has a natural lift to L consisting of connection-preserving automorphisms. On the other hand, ϕ^t operates on FP by push forward of frames—this is well defined since f is polarization preserving—and this flow automatically lifts to $\widetilde{F}P$ because ρ is a 2:1 submersion. Assembling these, we obtain a one-parameter group of automorphisms of $L \otimes \surd \wedge^n P$ that in turn induces a one-parameter group of linear isomorphisms of $\mathscr{H}$, which we also denote by ϕ^t. Setting $\hbar = h/2\pi$, the quantum observable $\mathscr{Q}f$ is then defined by

$$\mathscr{Q}f[\Psi] = i\hbar \frac{d}{dt}(\phi^t \Psi)|_{t=0}, \tag{3.3}$$

for all $\Psi \in \mathscr{H}$.

Remark: This technique is not applicable if the observables to be quantized do not preserve the polarization. One must use the Blattner–Kostant–Sternberg kernels to quantize such functions and the corresponding quantum operators—if they exist—will generally be more complicated.

Of principal interest is when X is a cotangent bundle T^*Q with the canonical symplectic structure $\omega = d\Theta$. We study this case in detail and present several formulas which will be useful later.

Let (q^i, p_i), $i = 1,...,n$, be a canonical bundle chart on $U \subset T^*Q$. Then

$$\Theta|U = \sum_{i=1}^n p_i \, dq^i \tag{3.4}$$

and

$$\omega|U = \sum_{i=1}^n dp_i \wedge dq^i . \tag{3.5}$$

A cotangent bundle carries a naturally defined polarization: the *vertical polarization* $V = \ker T\tau_Q$. Locally,

$$V = \operatorname{span}\left\{\frac{\partial}{\partial p_1},...,\frac{\partial}{\partial p_n}\right\}.$$

Since ω is exact, T^*Q is always prequantizable. Relative to a locally trivializing section λ: $U \to L$, the covariant differential is given by

$$\nabla\lambda = (1/i\hbar)\, \Theta \otimes \lambda . \tag{3.6}$$

A *metalinear structure* on a cotangent bundle will always mean a metalinear frame bundle for the vertical polarization. In this case the existence criterion is quite simple: a metalinear structure exists on T^*Q iff $w_1(Q)^2 = 0$, where $w_1(Q)$ is the first Stiefel–Whitney class of Q (see Ref. 22).

Let $\underline{f} = (\partial/\partial p_1,...,\partial/\partial p_n)$ be a local frame field for V and $\tilde{\underline{f}}$ a lift of $\underline{f}$ to $\widetilde{F}V$. Define $\nu_{\tilde{f}} \in \Gamma(\surd \wedge^n V)$ according to

$$\nu_{\tilde{f}}^{\#} \circ \tilde{\underline{f}} = 1 . \tag{3.7}$$

Both λ and $\nu_{\tilde{f}}$ are covariantly constant along V and every section Ψ of $L \otimes \surd \wedge^n V$ may be written

$$\Psi|U = \psi(q,p)\lambda \otimes \nu_{\tilde{f}}, \tag{3.8}$$

for some smooth function ψ on U. Such a Ψ is polarized iff $\psi = \psi(q)$ only. In particular, when L and $\surd \wedge^n V$ are trivial with global sections λ and ν, respectively, the association $\psi(q,p)\lambda \otimes \nu \mapsto \psi(q,p)$ defines an isomorphism $\Gamma(L \otimes \surd \wedge^n V) \approx C^\infty(T^*Q,\mathbf{C})$. The space $\mathscr{H}$ of polarized sections may similarly be identified with $C^\infty(Q,\mathbf{C})$.

Now suppose g is an observable which preserves V. Then for Ψ given locally by (3.8), (3.3) reduces to

$$\mathscr{Q}g[\Psi]|U = [\{-i\hbar \nabla_{X_g} + g - \tfrac{1}{2} i\hbar \operatorname{tr} A_{\underline{f}}(X_g)\}\psi\lambda] \otimes \nu_{\tilde{f}}, \tag{3.9}$$

where X_g is the Hamiltonian vector field of g and the components a_j^i of the matrix $A_{\underline{f}}(X_g)$ are found from

$$\left[X_g, \frac{\partial}{\partial p_i}\right] = \sum_{j=1}^n a_j^i \frac{\partial}{\partial p_j}.$$

B. The quantization of constrained systems

We wish to study the equivalence of the geometric quantizations of the extended and reduced phase spaces of a constrained classical system with symmetry (X,ω,G,J,μ). In this section we outline our strategy and delineate general criteria which must be met before we can proceed with the more technical aspects of the theory (which occupy the remainder of Secs. III and IV).

Our main concerns here are obtaining a natural quantization of the extended phase space and constructing a compatible quantization of the reduced phase space.

First of all, the extended phase space quantization must be "natural" in the sense that the classical symmetry algebra $\mathscr{g}$ also appears as a symmetry algebra on the quantum level. Hence the functions J_ξ must all be quantizable and the association $J_\xi \mapsto \mathscr{Q}J_\xi$ must be a Lie algebra homomorphism:

$$[\mathscr{Q}J_\xi, \mathscr{Q}J_\eta] = i\hbar \mathscr{Q}J_{[\xi,\eta]}, \tag{3.10}$$

for all $\xi,\eta \in \mathscr{g}$. This enables us to express the constraints $J = \mu$ as conditions

$$\mathscr{Q}J_\xi[\Psi] = \langle \mu,\xi \rangle \Psi \tag{3.11}$$

on the quantum wave functions $\Psi \in \mathscr{H}$.

Unfortunately, such a quantization will usually be inconsistent: the constraint operators $\mathscr{Q}J_\xi$ will have no nonzero eigenstates corresponding to the eigenvalues $\langle \mu,\xi \rangle$. For suppose Ψ satisfied (3.11) so that

$$[\mathscr{Q}J_\xi, \mathscr{Q}J_\eta]\Psi = (\langle \mu,\eta \rangle \langle \mu,\xi \rangle - \langle \mu,\xi \rangle \langle \mu,\eta \rangle)\Psi$$

vanishes. But then (3.10) yields

$$\mathscr{Q}J_{[\xi,\eta]}[\Psi] = \langle \mu,[\xi,\eta] \rangle \Psi \not\equiv 0,$$

which forces $\Psi = 0$. Thus the space $\mathscr{H}_\mu$ of physically admissible wave functions will be trivial.

To obtain meaningful results the offending term

$$\langle \mu,[\xi,\eta] \rangle = \langle \mathrm{ad}_\xi^* \mu,\eta \rangle$$

in the last equation must vanish for all ξ and η, and this happens iff μ is invariant. *One can therefore consistently quantize (X,ω,G,J,μ) only if μ is invariant.*

This invariance condition can be expressed geometrically. From Proposition (2.1) we have that μ is invariant iff

$G_\mu = G$ iff $J^{-1}(\mu)$ is a coisotropic submanifold of (X,ω), i.e.,

$$TJ^{-1}(\mu)^\perp \subseteq TJ^{-1}(\mu) .$$

The invariance of μ thus plays two key roles in our formalism: it is the primary obstruction to obtaining a consistent natural quantization of the system, and it guarantees that the reduction of a cotangent bundle is again a cotangent bundle. *Henceforth we assume that μ is invariant.*

We now return to the naturality question, viz., under what conditions will quantization produce a representation of $\mathscr{g}$ on $\mathscr{H}$? Once the invariance of μ ensures that no outright inconsistencies will occur, this reduces to a problem of making suitable choices of the geometric quantization structures discussed in the previous section. We must choose these so that the J_ξ are all quantizable and moreover that the quantum operators $\mathscr{Q}J_\xi$ thus obtained satisfy (3.10). In general, this will be possible iff the polarization P is G-invariant for then the J_ξ are all polarization-preserving functions (see Ref. 21, §6.2). However, if P is not invariant the $\mathscr{Q}J_\xi$ need not exist and, even if they are defined, (3.10) will not necessarily follow.

Having obtained a natural quantization of the extended phase space we now turn to the quantization of the reduced phase space. Our task is to correlate these two quantizations. We first observe that, by construction, the quantization of $(\overline{X}_\mu,\overline{\omega}_\mu)$ is completely determined by the structure of the constraint set. Consequently, if there is to be any hope for an equivalence of the extended and reduced phase space quantizations, we must ensure that the extended phase space quantization has this same property. This translates into the requirement that the extended wave functions be uniquely determined by their restrictions to $J^{-1}(\mu)$, and effectively places a further restriction on the choice of polarization.[7]

It remains only to construct quantization structures on $(\overline{X}_\mu,\overline{\omega}_\mu)$ that are compatible with those we already have on (X,ω). The basic idea is to project the quantization structures on the latter down to the former. An invariant polarization P on (X,ω) will project to a polarization on $(\overline{X}_\mu,\overline{\omega}_\mu)$ if, for example, P is transverse to $TJ^{-1}(\mu)^\perp$. For the prequantization and metalinear structures, we accomplish this by first lifting the action of G on X to $L\,|J^{-1}(\mu)$ and $\widetilde{F}P\,|J^{-1}(\mu)$ in a suitable manner. This is always possible infinitesimally, and the obstruction to extending from $\mathscr{g}$ to G is purely topological. In particular, there is no problem if G is simply connected; otherwise, one must choose L and $\widetilde{F}P$ appropriately—provided, of course, such "invariant" structures exist. We then quotient by these G-actions, producing bundles which are the required quantization structures on $(\overline{X}_\mu,\overline{\omega}_\mu)$.

We have now laid the foundation for comparable quantizations of the extended and reduced phase spaces. The next step is to check the above criteria and to explicitly construct the appropriate quantization structures for constrained cotangent systems (T^*Q,ω,G,J,μ), where μ is invariant, $\dim Q = n$, $\dim G = r$, and $\dim \overline{Q} = \bar{n} = n - r$. In the remainder of this section we work out the details for the three geometric quantization structures. Then, in Sec. IV, we show that the conditions we have set forth here are sufficient to guarantee the equivalence of the extended and reduced phase space quantizations.

C. Polarization

Let $V = \ker T\tau_Q$ and $\overline{V} = \ker T\tau_{\overline{Q}}$ be the vertical polarizations on T^*Q and $T^*\overline{Q}$, respectively.

Lemma (3.1): $V \cap TJ^{-1}(\mu)^\perp = \{0\}$.

Proof: Let $v \in V_\beta \cap T_\beta J^{-1}(\mu)^\perp$. Since $G_\mu = G$, it follows from Proposition (2.1) that $v \in T_\beta(G\cdot\beta)$, i.e., $v = \xi_{T^*Q}(\beta)$ for some $\xi \in \mathscr{g}$. But then $T\tau_Q(v) = \xi_Q(\tau_Q(\beta)) = 0$, which implies that $v = 0$ because Φ is free. ■

Taking the symplectic orthogonal complement of Lemma (3.1) gives

$$V + TJ^{-1}(\mu) = TT^*Q$$

over $J^{-1}(\mu)$. Counting fiber dimensions, we have

$$\begin{aligned} 2n &= \dim(V + TJ^{-1}(\mu)) \\ &= \dim V + \dim TJ^{-1}(\mu) - \dim(V \cap TJ^{-1}(\mu)) . \end{aligned}$$

By Proposition (2.2) μ is a regular value of J, so $\dim J^{-1}(\mu) = 2n - r$. It follows that $V \cap TJ^{-1}(\mu)$ is an involutive $\bar{n}$-dimensional distribution on $J^{-1}(\mu)$. Furthermore, since $\tau_{\overline{Q}} \circ \pi_\mu = \pi_Q \circ \tau_Q$,

$$\begin{aligned} &T\tau_{\overline{Q}}(T\pi_\mu(V \cap TJ^{-1}(\mu))) \\ &\quad = T\pi_Q(T\tau_Q(V \cap TJ^{-1}(\mu))) = \{0\} , \end{aligned}$$

so that $T\pi_\mu(V \cap TJ^{-1}(\mu))$ is vertical on $T^*\overline{Q}$. We have therefore proven the following proposition.

Proposition (3.2): $\overline{V} = T\pi_\mu(V \cap TJ^{-1}(\mu))$.

Clearly the action $T^*\Phi$ leaves V invariant. It follows automatically—regardless of the choices of the prequantization and metalinear structures—that the quantization of (T^*Q,ω,G,J,μ) in the Schrödinger representation will be natural.

Also note that every leaf of V intersects $J^{-1}(\mu)$. Indeed, since $J^{-1}(0)$ contains the zero section of T^*Q this is certainly true for $J^{-1}(0)$. But $J^{-1}(\mu)$ is obtained from $J^{-1}(0)$ by translation along the leaves of V (cf. Sec. II B), so this holds for $J^{-1}(\mu)$ as well. As V-wave functions are covariantly constant along V, they will be uniquely determined by their restrictions to $J^{-1}(\mu)$.

These results, coupled with the fact that the vertical polarization on T^*Q projects to the vertical polarization on $T^*\overline{Q}$, imply that we may consistently and compatibly quantize both the extended and reduced phase spaces in the Schrödinger representation (which is the physicists' "canonical" quantization). Moreover, since geometric quantization is very sensitive to the choice of polarization, it is a definite advantage to have at our disposal concrete examples of polarizations that satisfy all the criteria of Sec. III B.

D. Prequantization

Let L be a prequantization line bundle for (T^*Q,ω) with connection form γ. We must construct a compatible prequantization line bundle for $(T^*\overline{Q},\overline{\Omega}_\mu)$. Preliminary results along these lines have been obtained by Puta.[9]

The first step is to lift the action of $\mathscr{g}$ on T^*Q to $L_\mu = L\,|J^{-1}(\mu)$. For each $\xi \in \mathscr{g}$ let $\xi_L = \xi^*_{T^*Q}$ be the horizontal lift of ξ_{T^*Q} to L_μ.

Proposition (3.3): $\zeta \mapsto \zeta_L$ is a Lie algebra antihomomorphism.

Proof: We have to verify that $[\zeta,\eta]_L = -[\zeta_L,\eta_L]$ and for this it suffices to prove that $[\zeta_L,\eta_L]$ is horizontal. The prequantization condition (3.1) gives

$$\gamma([\zeta_L,\eta_L]) = -d\gamma(\zeta_L,\eta_L) = (1/h)\, l^*[\omega(\zeta_{T^*Q},\eta_{T^*Q})]\,,$$

which vanishes by virtue of Proposition (2.1) and the fact that $J^{-1}(\mu)$ is coisotropic, since both ζ_{T^*Q} and η_{T^*Q} belong to $TJ^{-1}(\mu)^\perp \subseteq TJ^{-1}(\mu)$. ∎

Remarks: (1) The association $\zeta \mapsto \zeta_L$ is an *antihomomorphism* as the G-action is on the *left*.

(2) Proposition (3.3) will fail if μ is not invariant, so that we cannot lift the $\mathscr{g}$-action to L_μ for arbitrary μ. In particular, the action will generally not be defined on all of L (unless, e.g., G is Abelian), but, insofar as reduction is concerned, we need only obtain an action on L_μ.

To extend this $\mathscr{g}$-action to a G-action is more difficult. There are two possible obstructions: the incompleteness of some of the vector fields ζ_L and the nonsimple connectivity of G. The first of these presents no problem: since L is a *line* bundle and the ζ_{T^*Q} are complete the ζ_L will be also. When $\pi_1(G) \neq 0$, however, some L_μ may admit G-actions while others will not.

Fix an orbit $G\cdot\beta \subseteq J^{-1}(\mu)$. Since orbits are isotropic in T^*Q [cf. Proposition (2.1)], the prequantization condition implies that $L\,|\,(G\cdot\beta)$ is flat. As the ζ_L are horizontal the $\mathscr{g}$-action on $L\,|\,(G\cdot\beta)$ will integrate to a G-action iff the holonomy of $L\,|\,(G\cdot\beta)$ is trivial. The crucial observation is that the holonomy of $L\,|\,(G\cdot\beta)$ is the *same* for all orbits in a *given* level set $J^{-1}(\mu)$.

To show this let $c(t)$, $0 \leqslant t \leqslant T$, be a loop in G based at the identity and let $c_\beta(t) = T^*\Phi_{c(t)}(\beta)$ be the corresponding loop in $G\cdot\beta$ based at β. From Ref. 12, §5.5.2, we find that the element in the holonomy group of $l^{-1}(\beta)$ determined by c_β is

$$\exp\left(\frac{i}{\hbar}\int_{c_\beta}\Theta\right). \tag{3.12}$$

Let ζ_t be the curve in $\mathscr{g}$ defined by

$$\zeta_t = TL^{-1}_{c(t)}\,(c_*(t))\,,$$

where $L_{c(t)}$ is left translation by $c(t)$. Then a short calculation using (2.3) yields

$$\int_{c_\beta}\Theta = \int_0^T \langle J(c_\beta(t)),\ \zeta_t\rangle dt = \int_0^T \langle \mu,\zeta_t\rangle dt\,, \tag{3.13}$$

which depends only upon μ and the homotopy class of c and not the particular orbit $G\cdot\beta \subseteq J^{-1}(\mu)$. It therefore makes sense to speak of "the holonomy" of L_μ.

Proposition (3.4): The action of $\mathscr{g}$ on L_μ can be extended to a G-action iff L_μ has trivial holonomy.

This proposition is essentially a "quantization condition"; we will see it in operation in Sec. V.

Assuming that L_μ has trivial holonomy, we are now able to construct the reduced prequantization line bundle. Since the G-action on L_μ is necessarily free and proper we may form $\overline{L}_\mu = L_\mu/G$, which is clearly a complex line bundle over $T^*\overline{Q}$. Denote the projections $L_\mu \to \overline{L}_\mu$ and $\overline{L}_\mu \to T^*\overline{Q}$ by π_μ and $\bar{l}_\mu$, respectively.

Set $\gamma_\mu = j_\mu^*\gamma$, where j_μ: $L_\mu \to L$ is the inclusion. By (3.1) and (2.1)

$$\mathscr{L}_{\zeta_L}\gamma = -(1/h)l^*(\zeta_{T^*Q} \lrcorner\, \omega) = (1/h)l^*\, dJ_\zeta$$

and so

$$\mathscr{L}_{\zeta_L}\gamma_\mu = (1/h)l^*d\,\langle \mu,\zeta\rangle = 0\,.$$

Consequently γ_μ projects to a complex-valued one-form $\bar{\gamma}_\mu$ on $\overline{L}_\mu$ such that

$$\gamma_\mu = \pi_\mu^*\,\bar{\gamma}_\mu\,. \tag{3.14}$$

Theorem (3.5): $(\overline{L}_\mu,\bar{\gamma}_\mu)$ is a prequantization line bundle for $(T^*\overline{Q},\overline{\Omega}_\mu)$.

Proof: It is straightforward to check that $\bar{\gamma}_\mu$ is indeed a connection form on $\overline{L}_\mu$.

To prove the Theorem we must verify the prequantization condition

$$d\bar{\gamma}_\mu = -(1/h)\,\bar{l}_\mu^*\overline{\Omega}_\mu\,.$$

Now $d\bar{\gamma}_\mu = \bar{l}_\mu^*\rho$ for some two-form ρ on $T^*\overline{Q}$. By (3.14), (3.1), and the Kummer–Marsden–Satzer Theorem,

$$\begin{aligned}(\bar{l}_\mu\circ\pi_\mu)^*\rho &= d\gamma_\mu\\ &= -(1/h)j_\mu^*l^*\omega\\ &= -(1/h)l^*j_\mu^*\omega\\ &= -(1/h)l^*\pi_\mu^*\overline{\Omega}_\mu\\ &= -(1/h)(\bar{l}_\mu\circ\pi_\mu)^*\overline{\Omega}_\mu\,.\end{aligned}$$

Since $\bar{l}_\mu\circ\pi_\mu$ is a submersion, $\rho = -(1/h)\overline{\Omega}_\mu$. ∎

Thus if L_μ has trivial holonomy, $(T^*\overline{Q},\overline{\Omega}_\mu)$ is prequantizable. In particular, the de Rham class $(1/h)[\overline{\Omega}_\mu]_{T^*\overline{Q}} = (1/h)[F_\mu]_{\overline{Q}}$ must be integral. We will have a nice physical interpretation of this result in Sec. V C.

E. Metalinear structures

Let $\widetilde{F}V$ be a metalinear frame bundle for the vertical polarization V. Since V is invariant, G acts on FV by push forward of frames. Following the general technique of Ref. 10, we will relate FV on T^*Q to $F\overline{V}$ on $T^*\overline{Q}$ and use this relation, along with a lift of the G-action on FV to $\widetilde{F}V$, to induce a metalinear structure on $T^*\overline{Q}$ from that on T^*Q.

We can substantially simplify matters by working on configuration spaces rather than cotangent bundles. To this end, let FQ and F^*Q be the linear frame and coframe bundles of Q, respectively. There exist natural G-actions $F\Phi$ on FQ and $F^*\Phi$ on F^*Q again given by push forward of frames and coframes. Let Z: $Q \to T^*Q$ be the zero section.

Proposition (3.6): There exist canonical G-equivariant isomorphisms $FQ \approx Z^*(FV)$ and $FV \approx \tau_Q^*(FQ)$.

Proof: There is a canonical equivariant identification of T^*Q with Z^*V and hence of V with $\tau_Q^*(T^*Q)$. These induce similar identifications of F^*Q with $Z^*(FV)$ and FV with $\tau_Q^*(FQ)$. Moreover, associating to each basis its dual basis gives rise to a canonical isomorphism $FQ \approx F^*Q$ which is equivariant with respect to the actions $F\Phi$ and $F^*\Phi$. Combining these isomorphisms establishes the Proposition. ∎

This result allows us to transform back and forth from T^*Q to Q. Similarly, there are canonical isomorphisms $F\overline{Q}\approx\overline{Z}^*(F\overline{V})$ and $F\overline{V}=\tau_{\overline{Q}}^*(F\overline{Q})$.

Now consider the subbundle B of FQ consisting of frames of the form $\underline{b}=(\underline{v},\underline{\xi}_Q)$, where $\underline{\xi}$ is a positively oriented orthonormal frame for $\mathscr{g}$ with respect to the given bi-invariant metric on G. The space B is a right principal H-bundle over Q, where

$$H=\left\{\begin{pmatrix} N & O\\ P & R\end{pmatrix}\middle|\ N\in \mathrm{GL}(\bar{n},\mathbb{R}),\ R\in \mathrm{SO}(r)\right\}$$

is the subgroup of $\mathrm{GL}(n,\mathbb{R})$ that stabilizes B.

Let K be the subgroup of H consisting of those matrices which leave invariant the projection of $\underline{v}$ to $F\overline{Q}$; explicitly,

$$K=\left\{\begin{pmatrix} I & O\\ P & R\end{pmatrix}\middle|\ R\in \mathrm{SO}(r)\right\}.$$

It is a normal subgroup of H and $H/K\approx \mathrm{GL}(\bar{n},\mathbb{R})$. We may therefore identify

$$B/K\approx\pi_{\overline{Q}}^*(F\overline{Q}). \tag{3.15}$$

Since $F\Phi_g(\underline{\xi}_Q)=-(\mathrm{Ad}_{g^{-1}}\underline{\xi})_Q$ and the metric on G is bi-invariant, $\mathrm{Ad}_{g^{-1}}\underline{\xi}$ is also a positively oriented orthonormal frame for $\mathscr{g}$. It follows that the left action $F\Phi$ on FQ induces an action on B that commutes with the right action of H. This allows us to quotient by G in (3.15), thereby obtaining a natural isomorphism

$$G\setminus(B/K)\approx F\overline{Q}. \tag{3.16}$$

We next lift these constructions to the metabundles. Let ρ: $\widetilde{FQ}\to FQ$ be a metalinear frame bundle for Q and set $\widetilde{H}=\sigma^{-1}(H)$, where σ: $\mathrm{ML}(n,\mathbb{R})\to \mathrm{GL}(n,\mathbb{R})$ is the 2:1 projection. Then $\widetilde{B}=\rho^{-1}(B)$ is a right principal $\widetilde{H}$-bundle over Q. Since the determinant of any matrix in K is unity, $\sigma^{-1}(K)$ has two connected components. We identify K with the component of the identity in $\sigma^{-1}(K)$. Then K is a normal subgroup of $\widetilde{H}$ and $\widetilde{H}/K\approx \mathrm{ML}(\bar{n},\mathbb{R})$. The bundle $\widetilde{B}/K$ is therefore a right principal $\mathrm{ML}(\bar{n},\mathbb{R})$-bundle on Q such that the diagram

$$\begin{array}{ccc} \widetilde{B}/K\times \mathrm{ML}(\bar{n},\mathbb{R}) & \longrightarrow & \widetilde{B}/K\\ \downarrow & & \downarrow\\ B/K\times \mathrm{GL}(\bar{n},\mathbb{R}) & \longrightarrow & B/K \end{array} \tag{3.17}$$

commutes, where the horizontal arrows are the right group actions and the vertical arrows are twofold projections.

Suppose for the moment that the G-action $F\Phi$ on FQ lifts to an action $\widetilde{F\Phi}$ on $\widetilde{FQ}$. Then $\widetilde{F\Phi}$ will preserve $\widetilde{B}$ and commute with the right $\widetilde{H}$-action on $\widetilde{B}$, and thus give rise to a left G-action on $\widetilde{B}/K$ that commutes with the right action of $\mathrm{ML}(\bar{n},\mathbb{R})$. The space $G\setminus(\widetilde{B}/K)$ of G-orbits in $\widetilde{B}/K$ will therefore inherit the structure of a right principal $\mathrm{ML}(\bar{n},\mathbb{R})$-bundle over $\overline{Q}$ such that the projection $G\setminus(\widetilde{B}/K)\to G\setminus(B/K)$ is 2:1. Applying (3.17) and (3.16) it follows that $G\setminus(\widetilde{B}/K)$ will define a metalinear frame bundle $\widetilde{F\overline{Q}}$ for $\overline{Q}$.

In summary, if Q is metalinear and the action of G on FQ lifts to $\widetilde{FQ}$, then $\overline{Q}$ is metalinear. To complete the analysis we must determine whether in fact the group action lifts.

Since ρ is a 2:1 submersion, the action of $\mathscr{g}$ on FQ obtained by differentiating $F\Phi$ lifts to an action on $\widetilde{FQ}$ by complete vector fields. Thus the only possible obstruction to lifting $F\Phi$ is the nonsimple connectivity of G. Now this $\mathscr{g}$-action defines an involutive distribution on $\widetilde{FQ}$, and we may extend to a G-action iff this distribution has trivial holonomy when restricted to each orbit in FQ.

To measure the holonomy, introduce the characteristic homomorphism $\pi_1(FQ,\underline{f})\to\mathbb{Z}_2$ of $\widetilde{FQ}$ (see Ref. 23, §13) and consider the map $\pi_1(G)\to\pi_1(FQ,\underline{f})$ defined by $[c(t)]\mapsto[F\Phi_{c(t)}(\underline{f})]$. The above distribution has no holonomy over $G\cdot\underline{f}$ iff the composite homomorphism χ_G: $\pi_1(G)\to\mathbb{Z}_2$ is trivial. Furthermore, since $\mathbb{Z}_2$ has no nontrivial automorphisms, χ_G is independent of the choice of $\underline{f}\in FQ$. Thus if the $\mathscr{g}$-action extends over just one orbit, it extends over all of them.

Proposition (3.7): The action of G on FQ lifts to $\widetilde{FQ}$ iff the natural homomorphism $\chi_G:\pi_1(G)\to\mathbb{Z}_2$ is trivial.

There is another version of this result which is often useful. Note that the restriction of FQ to any orbit $G\cdot q$ in Q is trivial: each $\underline{f}\in F_qQ$ defines a global section $\Phi_g(q)\mapsto F\Phi_g(\underline{f})$ of $FQ\,|\,(G\cdot q)$. Proposition (3.7) then implies that we have lifting iff the restriction of $\widetilde{FQ}$ to any (and hence every) orbit in Q is trivial.

It remains to pull our results back to T^*Q and $T^*\overline{Q}$. We first observe that Proposition (3.6) holds on the metalinear level, i.e., if $\widetilde{FV}$ is a metalinear frame bundle for V then $\widetilde{FQ}=Z^*(\widetilde{FV})$ is one for Q and, conversely, every metalinear frame bundle $\widetilde{FV}$ is $\tau_Q^*(\widetilde{FQ})$ for some metalinear structure on Q. Similar resutls are true for $\widetilde{F\overline{V}}$ and $\widetilde{F\overline{Q}}$. Now, since the mechanics are the same in both cases, it is clear that the G-action on FQ lifts to $\widetilde{FQ}=Z^*(\widetilde{FV})$ iff that on FV lifts to $\widetilde{FV}=\tau_Q^*(\widetilde{FQ})$. It follows that these metalinear identifications are G-equivariant. Denote by the same letter $\widetilde{B}$ the pullback bundle $\tau_Q^*(\widetilde{B})\subset\widetilde{FV}$, and set $\widetilde{B}_\mu=\widetilde{B}\,|J^{-1}(\mu)$. We have proven the following theorem.

Theorem (3.8): If the action of G on FV lifts to $\widetilde{FV}$, then there exists a compatible metalinear structure

$$\widetilde{F\overline{V}}=G\setminus(\widetilde{B}_\mu/K)$$

on $T^*\overline{Q}$.

Remark: Similar results hold for metaplectic structures.[10]

IV. EQUIVALENCE OF COMPATIBLE QUANTIZATIONS

The stage is now set to prove the equivalence of the quantizations of the extended and reduced phase spaces. We have shown that quantization data on (T^*Q,ω,G,J,μ) consisting of the vertical polarization V, a prequantization line bundle (L,γ) and a metalinear frame bundle $\widetilde{FV}$ induce quantization data $\overline{V}$, $(\overline{L}_\mu,\overline{\gamma}_\mu)$ and $\widetilde{F\overline{V}}$ on $(T^*\overline{Q},\overline{\Omega}_\mu)$, provided both L_μ and $\widetilde{FV}$ have trivial holonomy. The corresponding quantizations of (T^*Q,ω) and $(T^*\overline{Q},\overline{\Omega}_\mu)$ are said to be *compatible*. Our main result is that *compatible quantizations are equivalent*. We will make this precise after disposing of some preliminaries.

A. Preliminaries

Let $\sqrt{\wedge^n V}$ and $\sqrt{\wedge^{\bar{n}}\overline{V}}$ be the bundles of half-forms relative to V and $\overline{V}$, respectively. Set $\sqrt{\wedge^n V_\mu}=\sqrt{\wedge^n V}|J^{-1}(\mu)$.

Proposition (4.1): There exists a canonical isomorphism $\surd\wedge^n V_\mu \approx \pi_\mu^* \surd\wedge^{\bar n}\bar V$.

Proof: Fix $\beta\in J^{-1}(\mu)$ and consider $\nu_\beta \in \surd\wedge^n V_\beta$. As $\Delta(\tilde M)=1$ for all $\tilde M\in K$, (3.2) implies that $\nu_\beta^\#(\underline{\tilde b}\tilde M) = \nu_\beta^\#(\underline{\tilde b})$. Since by Theorem (3.8) $\pi_\mu^* \bar F\bar V = \tilde B_\mu/K$, it follows that for $\underline{\tilde b}\in\tilde B_\beta$ the equation

$$\bar\nu_\beta^\#([\underline{\tilde b}]) = \nu_\beta^\#(\underline{\tilde b}) \tag{4.1}$$

defines an element $\bar\nu_\beta$ of $(\pi_\mu^*\surd\wedge^{\bar n}\bar V)_\beta$, where the brackets denote K-equivalence classes. Conversely, given $\bar\nu_\beta \in(\pi_\mu^*\surd\wedge^{\bar n}\bar V)_\beta$, (4.1) defines an element $\nu_\beta\in\surd\wedge^n V_\beta$ since, according to (3.2), any half-form is completely determined by its restriction to $\tilde B_\beta$. The association (4.1) is thus the desired complex line bundle isomorphism. ■

The action of G on $\tilde F V$ gives rise to a left action of G on $\surd\wedge^n V$ which commutes with the right action of $\mathrm{ML}(n,\mathbb{R})$ by multiplication by Δ. Since all our constructions are G-equivariant, Proposition (4.1) yields $G\backslash\surd\wedge^n V_\mu \approx \surd\wedge^{\bar n}\bar V$. Combining this with the results of Sec. III D, we have the following corollary.

Corollary (4.2): $G\backslash(L\otimes\surd\wedge^n V)_\mu \approx \bar L_\mu\otimes\surd\wedge^{\bar n}\bar V$.

Our next task is to relate the *two* induced actions of $\mathscr{g}$ on $\Gamma(L\otimes\surd\wedge^n V)_\mu$, which we must be careful to distinguish. The first is that provided by the naturality of the extended phase space quantization (cf. Sec. III B) and is used to quantize the momentum map. The second is needed to construct compatible quantization structures on the reduced phase space (cf. Secs. III D and III E and the preceding discussion). These two actions agree on $\Gamma(\surd\wedge^n V_\mu)$ but differ on $\Gamma(L_\mu)$. We derive expressions for the generators of both actions.

The first action on $\Gamma(L\otimes\surd\wedge^n V)_\mu$ is generated by the quantum constraint operators $\mathscr{Q}J_\xi$. According to (3.9) these are given by

$$\mathscr{Q}J_\xi[\Psi] = \{[-i\hbar\nabla_{\xi_{T^*Q}} + J_\xi - \tfrac12 i\hbar\,\mathrm{tr}\,A_{\underline f}(\xi_{T^*Q})]\psi\lambda\}\otimes\nu_{\underline f}, \tag{4.2}$$

for each local section Ψ of the form (3.8). Tracing through the derivation of this formula (cf. Ref. 21, §6.1), we find that the last term on the right-hand side of (4.2) arises from the action of $\mathscr{g}$ on $\surd\wedge^n V_\mu$ while the first two terms are due to the action of $\mathscr{g}$ on L_μ by *connection-preserving* vector fields

$$\xi^*_{T^*Q} - \zeta_{J_\xi/h}, \tag{4.3}$$

where $\zeta_{J_\xi/h}$ is the fundamental vector field on L defined by the function J_ξ/h. On the other hand, the second action of $\mathscr{g}$ on L_μ is generated by just the *horizontal* vector fields $\xi^*_{T^*Q}$. It is easy to see that removing the last term from (4.3) eliminates the second term from (4.2). Since both actions agree on $\surd\wedge^n V_\mu$, it follows from (4.2) that the generators $\mathscr{G}_\xi$ of the second action are related to the $\mathscr{Q}J_\xi$ by

$$\mathscr{Q}J_\xi[\Psi] = \mathscr{G}_\xi[\Psi] + J_\xi\Psi, \tag{4.4}$$

for all $\Psi\in\Gamma(L\otimes\surd\wedge^n V)_\mu$.

B. Smooth equivalence

We are finally ready to compare the extended and reduced phase space quantizations. They are correlated by the following theorem.

Smooth Equivalence Theorem: If the quantizations of the extended phase space (T^*Q,ω,G,J,μ) and the reduced phase space $(T^*\bar Q,\bar\Omega_\mu)$ are compatible, then there exists a canonical isomorphism $\mathscr{H}_\mu \approx \bar{\mathscr{H}}_\mu$.

Proof: Let $\Psi\in\mathscr{H}_\mu$. Equation (4.4) implies that $\mathscr{G}_\xi[\Psi]|J^{-1}(\mu)=0$, so $\Psi|J^{-1}(\mu)$ is G-invariant. By Corollary (4.2), Ψ projects to a smooth section $\bar\Psi$ of $\bar L_\mu\otimes\surd\wedge^{\bar n}\bar V$. Since Ψ is polarized, Proposition (3.2) shows that $\bar\Psi$ is also. Thus $\bar\Psi\in\bar{\mathscr{H}}_\mu$.

For the converse, suppose $\bar\Psi\in\bar{\mathscr{H}}_\mu$. Corollary (4.2), Proposition (3.2), and Eq. (4.4) imply that $\bar\Psi$ pulls back to a unique G-invariant section Ψ_μ of $(L\otimes\surd\wedge^n V)_\mu$ which is covariantly constant along $V\cap TJ^{-1}(\mu)$ and satisfies

$$\mathscr{Q}J_\xi[\Psi_\mu] = \langle\mu,\xi\rangle\Psi_\mu. \tag{4.5}$$

Since every leaf of V is simply connected and intersects $J^{-1}(\mu)$, parallel transport along V produces a globally defined polarized section Ψ of $L\otimes\surd\wedge^n V$ which agrees with Ψ_μ on $J^{-1}(\mu)$. Now consider the polarized sections

$$\Psi_\xi = \mathscr{Q}J_\xi[\Psi] - \langle\mu,\xi\rangle\Psi,$$

for each $\xi\in\mathscr{g}$. Every Ψ_ξ is uniquely determined by its restriction to $J^{-1}(\mu)$. But $\Psi_\xi|J^{-1}(\mu)=0$ by virtue of (4.5), so $\Psi_\xi\equiv0$ and hence $\Psi\in\mathscr{H}_\mu$.

This establishes the existence of the required isomorphism. ■

Remarks: (1) We emphasize that this isomorphism is entirely canonical since our constructions of the reduced quantization data are.

(2) When $\mu=0$ both of the $\mathscr{g}$-actions on $\mathscr{H}_0$ coincide. We may then restate the conclusion of this Theorem as follows: There exists a canonical isomorphism between the space of gauge-invariant smooth polarized sections of $L\otimes\surd\wedge^n V$ and the space of smooth polarized sections of $\bar L_0\otimes\surd\wedge^{\bar n}\bar V$. This special case is due to Sniatycki.[10]

We now derive a local expression for the isomorphism $\mathscr{H}_\mu\approx\bar{\mathscr{H}}_\mu$ which will be useful later. Let

$$(q^1,\ldots,q^n) = (\bar q^1,\ldots,\bar q^{\bar n},g^1,\ldots,g^r)$$

be a chart on $\pi_Q^{-1}(\bar U)\subset Q$ induced by a local trivialization $\pi_Q^{-1}(\bar U)\approx\bar U\times G$, and let (q^i,p_i), $i=1,\ldots,n$, and $(\bar q^i,\bar p_i)$, $i=1,\ldots,\bar n$, be the corresponding canonical charts on T^*Q and $T^*\bar Q$, respectively. Set $\underline f=(\partial/\partial p_1,\ldots,\partial/\partial p_n)$ and define $\nu_{\underline f}\in\Gamma(\surd\wedge^n V)$ by (3.7). It is convenient to construct another half-form $\nu_{\underline{\tilde b}}$ on T^*Q as follows (cf. Sec. III E). Using the given bi-invariant metric g on G, fix a positively oriented orthonormal frame $\underline\xi=(\xi_1,\ldots,\xi_r)$ for $\mathscr{g}$ and set

$$\underline b = \left(\frac{\partial}{\partial\bar q^1},\ldots,\frac{\partial}{\partial\bar q^{\bar n}},\xi_1,\ldots,\xi_r\right).$$

Then

$$\underline b = \left(\frac{\partial}{\partial\bar q^1},\ldots,\frac{\partial}{\partial\bar q^{\bar n}},\frac{\partial}{\partial g^1},\ldots,\frac{\partial}{\partial g^r}\right)\begin{pmatrix} I & O\\ O & C\end{pmatrix},$$

for some matrix C with $\det C=(\det g)^{-1/2}$. Applying Proposition (3.6) we find that under the isomorphism $FV\approx\tau_Q^*(FQ)$ the frame $(\partial/\partial\bar q^1,\ldots,\partial/\partial\bar q^{\bar n},\partial/\partial g^1,\ldots,\partial/\partial g^r)$ maps onto $\underline f$ while $\underline b$ maps onto a frame in B which we also denote by $\underline b$. Define $\nu_{\underline{\tilde b}}$ by $\nu_{\underline{\tilde b}}^\#\circ\underline{\tilde b}=1$; from the above and (3.2) it follows that

$$\nu_{\underline{b}} = (\det g)^{1/4}\,\nu_{\underline{f}}\,. \tag{4.6}$$

Then each $\Psi\in\mathscr{H}$ may be locally written as either

$$\Psi = \check{\psi}(q)\lambda\otimes\nu_{\underline{b}} \tag{4.7}$$

or

$$\Psi = \check{\psi}(q)\ (\det g)^{1/4}\lambda\otimes\nu_{\underline{f}}\,. \tag{4.8}$$

Now suppose $\Psi\in\mathscr{H}_\mu$ so that Ψ satisfies (3.11). Then using (4.7), (4.6), (4.2), (3.6), (2.3), and the fact that the components of ξ_Q are constant in this chart, we compute

$$\Psi = k(\bar{q})\exp\left(\frac{i}{\hbar}\sum_{a=1}^{r}\mu_a g^a\right)\lambda\otimes\nu_{\underline{b}}\,, \tag{4.9}$$

where k is arbitrary.

On the other hand, both $\underline{b}$ and $\underline{f}$ project to $\bar{\underline{f}} = (\partial/\partial\bar{p}_1,\ldots,\partial/\partial\bar{p}_n)\in F\bar{V}$. Defining $\bar{\nu}_{\underline{f}}\in\Gamma(\sqrt{\ }\wedge^{\bar{n}}\bar{V})$ by (3.7), it follows from (4.1) that $\nu_{\underline{b}}$ projects to $\bar{\nu}_{\underline{f}}$. Similarly, from Sec. III D we find that λ projects to the section $\bar{\lambda}_\mu$ of $\bar{L}_\mu$ defined by $\bar{\lambda}_\mu\circ\pi_\mu = \bar{l}_\mu\circ\lambda$. Since locally every $\bar{\Psi}\in\bar{\mathscr{H}}_\mu$ takes the form

$$\bar{\Psi} = \bar{\psi}(\bar{q})\bar{\lambda}_\mu\otimes\bar{\nu}_{\underline{f}}\,, \tag{4.10}$$

we have upon comparing (4.9) and (4.10) that the isomorphism $\mathscr{H}_\mu\to\bar{\mathscr{H}}_\mu$ is given by

$$k(\bar{q})\exp\left(\frac{i}{\hbar}\sum_{a=1}^{r}\mu_a g^a\right)\mapsto k(\bar{q})\,. \tag{4.11}$$

Compatible quantizations thus have canonically isomorphic spaces of physically admissible wave functions. But compatibility should ensure more than this: it should also intertwine the quantizations of G-invariant observables. More precisely, let $f\in C^\infty(T^*Q)$ be G-invariant in which case it reduces to $\bar{f}_\mu\in C^\infty(T^*\bar{Q})$ as indicated in Sec. II B. Then the quantum operators $\mathscr{Q}f$ on $\mathscr{H}$ and $\bar{\mathscr{Q}}\bar{f}_\mu$ on $\bar{\mathscr{H}}_\mu$ should be such that

$$\begin{array}{ccc} \mathscr{H}_\mu & \xrightarrow{\mathscr{Q}f} & \mathscr{H}_\mu \\ \downarrow & \bar{\mathscr{Q}}\bar{f}_\mu & \downarrow \\ \bar{\mathscr{H}}_\mu & \longrightarrow & \bar{\mathscr{H}}_\mu \end{array} \tag{4.12}$$

commutes, where the vertical arrows are the isomorphisms provided by the Smooth Equivalence Theorem. This is actually so, at least if f is polarization preserving.

Theorem (4.3): Let f be a G-invariant polarization-preserving observable. Then diagram (4.12) commutes.

Proof: First note that because f preserves V, $\bar{f}_\mu$ preserves $\bar{V}$ by Proposition (3.2). Consequently $\bar{f}_\mu$ is quantizable.

Let ϕ^t and $\bar{\phi}^t_\mu$ be the flows of f and $\bar{f}_\mu$ on T^*Q and $T^*\bar{Q}$, respectively, with

$$\pi_\mu\circ\phi^t = \bar{\phi}^t_\mu\circ\pi_\mu\,.$$

Since both f and $\bar{f}_\mu$ are polarization preserving, these flows induce one-parameter groups of bundle automorphisms of $L\otimes\sqrt{\ }\wedge^n V$ and $\bar{L}_\mu\otimes\sqrt{\ }\wedge^{\bar{n}}\bar{V}$, which we denote by the same symbols (cf. Sec. III A). Since f is invariant, a straightforward calculation using the techniques of Sec. III E along with the fact that $F\phi^t(\xi_Q(q)) = \xi_Q(\phi^t(q))$ shows that ϕ^t preserves $\tilde{B}_\mu$. Thus ϕ^t is equivariant and it follows from Corollary (4.2) that

$$\begin{array}{ccc} (L\otimes\sqrt{\ }\wedge^n V)_\mu & \xrightarrow{\phi^t} & (L\otimes\sqrt{\ }\wedge^n V)_\mu \\ \downarrow & \bar{\phi}^t_\mu & \downarrow \\ \bar{L}_\mu\otimes\sqrt{\ }\wedge^{\bar{n}}\bar{V} & \longrightarrow & \bar{L}_\mu\otimes\sqrt{\ }\wedge^{\bar{n}}\bar{V} \end{array} \tag{4.13}$$

commutes, where the vertical arrows are projections.

Now consider the corresponding one-parameter groups of linear isomorphisms ϕ^t of $\mathscr{H}$ and $\bar{\phi}^t_\mu$ of $\bar{\mathscr{H}}_\mu$. From the definition of $\mathscr{Q}J_\xi$ applied to $\phi^t\Psi$ we have $\mathscr{Q}J_\xi[\phi^t\Psi] = \phi^t\,\mathscr{Q}J_\xi[\Psi]$ as ϕ^t is equivariant. In particular, if $\Psi\in\mathscr{H}_\mu$, then so is $\phi^t\Psi$. Thus (4.13) and the Smooth Equivalence Theorem imply that the induced diagram

$$\begin{array}{ccc} \mathscr{H}_\mu & \xrightarrow{\phi^t} & \mathscr{H}_\mu \\ \downarrow & \bar{\phi}^t_\mu & \downarrow \\ \bar{\mathscr{H}}_\mu & \longrightarrow & \bar{\mathscr{H}}_\mu \end{array} \tag{4.14}$$

commutes.

The quantum operators $\mathscr{Q}f$ and $\bar{\mathscr{Q}}\bar{f}_\mu$ are defined by

$$\mathscr{Q}f[\Psi] = i\hbar\frac{d}{dt}(\phi^t\Psi)|_{t=0}$$

and

$$\bar{\mathscr{Q}}\bar{f}_\mu[\bar{\Psi}] = i\hbar\frac{d}{dt}(\bar{\phi}^t_\mu\bar{\Psi})|_{t=0}$$

[cf. (3.3)]. If $\Psi\in\mathscr{H}_\mu$ then $\phi^t\Psi\in\mathscr{H}_\mu$ and consequently $\mathscr{Q}f[\Psi]\in\mathscr{H}_\mu$. Thus diagram (4.12) is well defined and its commutativity now follows immediately from the above definitions and (4.14). ■

Roughly, this result states that one may quantize invariant observables in either formalism with equivalent results. However, the Theorem does not apply when f is not polarization preserving. In such cases diagram (4.12) may not exist and, when it does, it will generally not commute.

C. Unitary equivalence

We now discuss the one facet of the equivalence problem that we have overlooked thus far—the Hilbert space structure. The Theorems of Sec. IV B pertain only to *smooth* quantizations, i.e., the linear spaces $\mathscr{H}_\mu$ and $\bar{\mathscr{H}}_\mu$ of C^∞ wave functions. Do our results still apply when the quantum inner products are introduced? More precisely, does the linear isomorphism $\mathscr{H}_\mu\approx\bar{\mathscr{H}}_\mu$ of the Smooth Equivalence Theorem extend to a unitary isomorphism of the corresponding quantum Hilbert spaces?

For constrained cotangent systems, the spaces $\mathscr{H}$ and $\bar{\mathscr{H}}_\mu$ of polarized states carry canonically defined inner products.[12,21] Using the setup of the previous section, we may describe these as follows. The inner product of two wave functions $\Psi,\Phi\in\mathscr{H}$ of the form (4.7) with supports in $\pi_Q^{-1}(\bar{U})$ is

$$(\Psi,\Phi)_Q = \int_{\pi_Q^{-1}(\bar{U})}\psi(q)\phi^*(q)\sqrt{\det g}\,d^n q\,, \tag{4.15}$$

where the star denotes complex conjugation. Similarly, for $\bar{\Psi},\bar{\Phi}\in\bar{\mathscr{H}}_\mu$ of the form (4.10) with supports in $\bar{U}$,

$$(\bar{\Psi},\bar{\Phi})_{\bar{Q}} = \int_{\bar{U}}\bar{\psi}(\bar{q})\bar{\phi}^*(\bar{q})d^{\bar{n}}\bar{q}\,. \tag{4.16}$$

We complete these spaces with respect to these inner products thereby obtaining the quantum Hilbert spaces $\mathscr{h}$ and $\bar{\mathscr{h}}_\mu$, respectively.

Although this procedure is in itself straightforward, a difficulty arises when considering the space

$$\mathscr{h}_\mu = \{\Psi\in\mathscr{h} \,|\, \mathscr{Q}J_\xi[\Psi] = \langle\mu,\xi\rangle\Psi\}$$

of physically admissible states. It may happen that $\mathscr{h}_\mu$ will consist only of distributional wave functions. For instance, if G is noncompact some of the eigenvalues $\langle\mu,\xi\rangle$ will lie in the continuous spectra of the corresponding constraint operators $\mathscr{Q}J_\xi$. In such cases the inner product on $\mathscr{h}$ will not induce one on $\mathscr{h}_\mu$ so that, in general, $\mathscr{h}_\mu$ and $\bar{\mathscr{h}}_\mu$ can only be compared as linear spaces. However, one may use the Smooth Equivalence Theorem and the inner product on $\bar{\mathscr{h}}_\mu$ to induce one on $\mathscr{h}_\mu$ in such a way that $\mathscr{h}_\mu$ and $\bar{\mathscr{h}}_\mu$ will then be unitarily related. We will see an example of this phenomenon in Sec. V A.

This problem cannot occur if G is compact, in which case $\mathscr{h}_\mu$ is a genuine subspace of $\mathscr{h}$.

Unitary Equivalence Theorem: If G is compact then $\mathscr{h}_\mu$ and $\bar{\mathscr{h}}_\mu$ are unitarily equivalent.

Proof: Let $\Psi,\Phi\in\mathscr{h}_\mu$. Substituting (4.9) into (4.15) we obtain the induced inner product

$$(\Psi,\Phi)_\mu = \int_{\pi_Q^{-1}(\bar{U})} k(\bar{q})h^*(\bar{q})\sqrt{\det g}\, d^n q\,,$$

where h is to Φ as k is to Ψ. Writing $d^n q = d^r g\, d^{\bar{n}}\bar{q}$, this reduces to

$$(\Psi,\Phi)_\mu = \mathrm{vol}(G)\int_{\bar{U}} k(\bar{q})h^*(\bar{q})d^{\bar{n}}\bar{q}\,, \tag{4.17}$$

where

$$\mathrm{vol}(G) = \int_G \sqrt{\det g}\, d^r g$$

is finite since G is compact.

The isomorphism $\mathscr{H}_\mu \to \bar{\mathscr{H}}_\mu$ of the Smooth Equivalence Theorem clearly extends to $\mathscr{h}_\mu$ thereby enabling us to project Ψ and Φ on T^*Q to wave functions $\bar{\Psi}$ and $\bar{\Phi}$ on $T^*\bar{Q}$. Using the explicit form (4.11) of this projection in (4.16) yields

$$(\bar{\Psi},\bar{\Phi})_{\bar{Q}} = \int_{\bar{U}} k(\bar{q})h^*(\bar{q})d^{\bar{n}}\bar{q}\,. \tag{4.18}$$

It follows from (4.17) that $\bar{\Psi},\bar{\Phi}\in\bar{\mathscr{h}}_\mu$. The mapping $\mathscr{U}$: $\mathscr{h}_\mu\to\bar{\mathscr{h}}_\mu$ defined locally by

$$k(\bar{q})\exp\left(\frac{i}{\hbar}\sum_{a=1}^{r}\mu_a g^a\right)\lambda\otimes\nu_{\underline{b}} \mapsto \sqrt{\mathrm{vol}(G)}\, k(\bar{q})\bar{\lambda}_\mu\otimes\bar{\nu}_{\underline{\bar{b}}} \tag{4.19}$$

is therefore a vector space isomorphism, and a comparison of (4.17) with (4.18) shows that it is unitary. ■

Similarly, Theorem (4.3) carries over to the Hilbert space case when G is compact. Namely, if f is a G-invariant polarization-preserving observable, then the unitary isomorphism $\mathscr{U}$ intertwines the quantum operators corresponding to f and its reduction $\bar{f}_\mu$:

$$\bar{\mathscr{Q}}\bar{f}_\mu = \mathscr{U}^{-1}(\mathscr{Q}f)\mathscr{U}\,.$$

To summarize, if G is compact, then a smooth equivalence of the extended and reduced phase space quantizations naturally extends to a unitary equivalence which intertwines the quantizations of invariant observables. When G is noncompact no such natural unitary equivalence exists *a priori*, but the Smooth Equivalence Theorem may be used to induce one.

V. EXAMPLES

We present several examples which illustrate our techniques and theorems. In most cases we will explicitly verify our results by direct computation.

A. Center of mass reduction in the *N*-body problem

Our presentation follows that in §10.4 of Abraham and Marsden[15]; see also Robinson.[24]

Consider N masses $m_1,\dots,m_N$ moving in $\mathbb{R}^k$. Upon removing collisions we have

$$Q = (\mathbb{R}^k)^N - \Delta^N\,,$$

where

$$\Delta^N = \bigcup_{1<i,j<N} \Delta^N_{ij}$$

and

$$\Delta^N_{ij} = \{\mathbf{q} = (\underline{q}^1,\dots,\underline{q}^N)\in(\mathbb{R}^k)^N \,|\, \underline{q}^i = \underline{q}^j\}\,.$$

The translation group $\mathbb{R}^k$ acts freely and properly on Q according to

$$\Phi(\underline{g},\mathbf{q}) = (\underline{q}^1+\underline{g},\dots,\underline{q}^N+\underline{g})\,.$$

To construct the orbit space introduce the diffeomorphism

$$C\colon Q\to\{(\dot{\mathbb{R}}^k)^{N-1} - \Delta^{N-1}\}\times\mathbb{R}^k$$

given by

$$(\underline{q}^1,\dots,\underline{q}^N) \mapsto\left(m_1(\underline{q}^1-\underline{q}^N),\dots,m_{N-1}(\underline{q}^{N-1}-\underline{q}^N),\ \sum_{i=1}^N m_i\underline{q}^i\right), \tag{5.1}$$

where $\dot{\mathbb{R}}^k = \mathbb{R}^k - \{\underline{0}\}$. The corresponding $\mathbb{R}^k$-action $C\circ\Phi_{\underline{g}}\circ C^{-1}$ is just translation in the last factor by $M\underline{g}$, where $M = \Sigma_{i=1}^N m_i$ is the total mass of the system, so that

$$\bar{Q}\approx(\dot{\mathbb{R}}^k)^{N-1} - \Delta^{N-1}\,.$$

Thus $Q = \bar{Q}\times\mathbb{R}^k$ is trivial as a principal $\mathbb{R}^k$-bundle.

This result is useful for understanding the structure of $\bar{Q}$ which, in general, is quite complicated. More meaningful physically, however, is the representation of $\bar{Q}$ as the $N(k-1)$-dimensional submanifold $C^{-1}(\bar{Q}\times\{\underline{0}\})$ of Q obtained by fixing the center of mass of the system at the origin. Thus we view

$$\bar{Q} = \left\{\mathbf{q}\in Q \,\middle|\, \sum_{i=1}^N m_i\underline{q}^i = \underline{0}\right\}. \tag{5.2}$$

The extended phase space is $T^*Q = Q\times(\mathbb{R}^k)^N$ with symplectic form $\omega = d\Theta$, where $\Theta = \Sigma_{i=1}^N \underline{p}_i\cdot d\underline{q}^i$. The cotangent action is $T^*\Phi(\underline{g},(\mathbf{q},\mathbf{p})) = (\Phi_{\underline{g}}(\mathbf{q}),\mathbf{p})$ with momentum map

$$J(\mathbf{q},\mathbf{p}) = \sum_{i=1}^N \underline{p}_i\,. \tag{5.3}$$

Now $\mathbb{R}^k$ is Abelian so every $\mu\in\mathscr{g}^*\approx\mathbb{R}^k$ is invariant; for simplicity we consider only $\mu = 0$. Taking (5.2) and (2.4) into account, we can identify $T^*\bar{Q}$ with $\tau_Q^{-1}(\bar{Q})\cap J^{-1}(\underline{0})$, i.e.,

$$T^*\bar{Q} = \left\{(\mathbf{q},\mathbf{p})\in T^*Q \,\middle|\, \sum_{i=1}^N m_i \underline{q}^i = \underline{0},\quad \sum_{i=1}^N \underline{p}_i = \underline{0}\right\}. \tag{5.4}$$

Although there may be various prequantizations of (T^*Q,ω) depending upon the topology of Q, we can always take $L = T^*Q\times\mathbb{C}$ with trivializing section $\lambda(\mathbf{q},\mathbf{p}) = (\mathbf{q},\mathbf{p},1)$. Since $\mathbb{R}^k$ is simply connected, the action $T^*\Phi$ lifts horizontally to all of L; $\bar{L}_0$ therefore exists. Using (5.4) to explicitly identify $\bar{L}_0$ with $L\,|T^*\bar{Q}$, it follows that $\bar{L}_0 = T^*\bar{Q}\times\mathbb{C}$.

Since T^*Q is parallelizable one possibility for $\widetilde{F}V$ is simply $T^*Q\times\mathrm{ML}(Nk,\mathbb{R})$. The induced action of $\mathbb{R}^k$ on FV is trivial on the fibers and consequently lifts to $\widetilde{F}V$. Thus the corresponding metalinear frame bundle $\widetilde{F\bar{V}}$ for

$$\bar{V} = \mathrm{span}\left\{\underline{v}_1\cdot\frac{\partial}{\partial\underline{p}_1}+\dots+\underline{v}_N\cdot\frac{\partial}{\partial\underline{p}_N}\,\middle|\,\sum_{i=1}^N \underline{v}_i = \underline{0}\right\}$$

is also a product.

We now quantize the extended phase space. Setting $\underline{f} = (\partial/\partial\underline{p}_1,\dots,\partial/\partial\underline{p}_N)$, we have from Sec. III A that every polarized $\Psi\in\Gamma(L\otimes\sqrt{\wedge^{Nk}V})$ can be written

$$\Psi = \psi(\mathbf{q})\lambda\otimes\nu_{\underline{f}}. \tag{5.5}$$

From (4.2), (5.3), and (5.5) the quantum constraint operators are

$$\mathscr{Q}J[\Psi] = -i\hbar\{\nabla_1+\dots+\nabla_N)\psi(\mathbf{q})\}\lambda\otimes\nu_{\underline{f}},$$

where ∇_i is the ordinary gradient with respect to $\underline{q}^i$. Thus the physically admissible quantum states are those that satisfy

$$(\nabla_1+\dots+\nabla_N)\psi(\mathbf{q}) = 0.$$

It follows that $\mathscr{H}_0$ can be identified with the set of all $\psi\in C^\infty(Q,\mathbb{C})$ of the form, say,

$$\psi = \psi(\underline{q}^1-\underline{q}^N,\dots,\underline{q}^{N-1}-\underline{q}^N). \tag{5.6}$$

Similarly, quantizing the reduced phase space gives $\bar{\mathscr{H}}_0\approx C^\infty(\bar{Q},\mathbb{C})$. We have from (5.1) that every $\bar{\psi}\in C^\infty(\bar{Q},\mathbb{C})$ is of the form

$$\bar{\psi} = \bar{\psi}(m_1(\underline{q}^1-\underline{q}^N),\dots,m_{N-1}(\underline{q}^{N-1}-\underline{q}^N)). \tag{5.7}$$

A comparison of (5.6) with (5.7) yields the isomorphism $\mathscr{H}_0\to\bar{\mathscr{H}}_0$ predicted by the Smooth Equivalence Theorem.

From (4.15) and (4.16) it follows that the Hilbert spaces $\mathscr{h}$ and $\bar{\mathscr{h}}_0$ are $L^2((\mathbb{R}^k)^N-\Delta^N)$ and $L^2((\mathbb{R}^k)^{N-1}-\Delta^{N-1})$, respectively. Now $\underline{0}$ is in the continuous spectrum of $\mathscr{Q}J$ and from (5.6) it is clear that none of the translationally invariant wave functions are square integrable. Thus $\mathscr{h}_0$ and $\bar{\mathscr{h}}_0$ can only be compared as linear spaces.

Remark: For nonzero $\mu\in\mathbb{R}^k$, reduction fixes the center of mass as moving with velocity μ/M.

B. Angular momentum

We study the system consisting of a single particle moving in $\dot{\mathbb{R}}^n = \mathbb{R}^n - \{\underline{0}\}$ with constant angular momentum. This example is interesting for two reasons. First, when $n>2$, the action

$$\Phi(A,\underline{q}) = A\underline{q} \tag{5.8}$$

of $\mathrm{SO}(n)$ on $\dot{\mathbb{R}}^n$ is not free so that the fundamental assumption underlying all our results is violated. Nonetheless, as we shall see, the conclusions of our theorems still hold. Second, the case $n=2$ illustrates how the obstructions to lifting the group action to the various quantization structures give rise to "quantization conditions."

Since the case $n=2$ is exceptional, we first consider only $n>2$.

The action (5.8) is always proper and effective. The orbits are concentric spheres and thus

$$\dot{\mathbb{R}}^n/\mathrm{SO}(n)\approx\mathbb{R}^+. \tag{5.9}$$

Viewing $\underline{q}$ and $\underline{p}$ as column vectors, the cotangent action on $T^*\dot{\mathbb{R}}^n = \dot{\mathbb{R}}^n\times\mathbb{R}^n$ becomes

$$T^*\Phi(A,(\underline{q},\underline{p})) = (A\underline{q},A\underline{p}) \tag{5.10}$$

and, upon identifying $\mathrm{so}(n)^*$ and $\mathrm{so}(n)$ with $\wedge^2(\mathbb{R}^n)$, the angular momentum map can be written

$$J(\underline{q},\underline{p}) = \underline{q}\wedge\underline{p}. \tag{5.11}$$

The coadjoint action of $\mathrm{SO}(n)$ on $\wedge^2(\mathbb{R}^n)$ is

$$\mathrm{Ad}_A^*(\underline{q}\wedge\underline{p}) = A^{-1}\underline{q}\wedge A^{-1}\underline{p},$$

from which it follows that

$$\mathrm{SO}(n)_\mu\approx\mathrm{SO}(2)\times\mathrm{SO}(n-2),$$

for $\mu\neq0$. Consequently 0 is the *only* invariant element of $\mathrm{so}(n)^*$ for $n>2$.

Remark: With reference to the discussion in Sec. III B, it is not surprising that this system cannot be consistently quantized when $\mu\neq0$. Indeed, (3.11) would correspond to simultaneously specifying *all* the components of the angular momentum—a well-known quantum mechanical impossibility.

Now 0 is not a regular value of J, but it is weakly regular. Actually J has rank $n-1$ on $J^{-1}(0)$, so that $J^{-1}(0)$ is an $(n+1)$-dimensional submanifold of $\dot{\mathbb{R}}^n\times\mathbb{R}^n$. From (5.11) we have

$$J^{-1}(0) = \{(\underline{q},s\underline{q})\,|\,\underline{q}\neq0,\ s\in\mathbb{R}\}, \tag{5.12}$$

which shows that $J^{-1}(0)$ is a real line bundle over $\dot{\mathbb{R}}^n$. This bundle has a global nonvanishing section $\underline{q}\mapsto(\underline{q},\underline{q})$ so that in fact

$$J^{-1}(0) = \dot{\mathbb{R}}^n\times\mathbb{R}. \tag{5.13}$$

To reduce $J^{-1}(0)$, first note that the action of $\mathrm{SO}(n)$ on $\dot{\mathbb{R}}^n\times\mathbb{R}$ induced by (5.13), (5.12), and (5.10) is just $(\underline{q},s)\mapsto(A\underline{q},s)$. Then (5.13) and (5.9) imply that

$$J^{-1}(0)/\mathrm{SO}(n)\approx\mathbb{R}^+\times\mathbb{R},$$

with projection $\pi_0(\underline{q},s\underline{q}) = (\|\underline{q}\|,s\|\underline{q}\|)$. Now fix $\check{\underline{q}}\in\dot{\mathbb{R}}^n$ with $\|\check{\underline{q}}\| = 1$. The map i_0: $\mathbb{R}^+\times\mathbb{R}\to J^{-1}(0)\subset\dot{\mathbb{R}}^n\times\mathbb{R}^n$, defined by

$$i_0(r,s) = (r\check{\underline{q}},s\check{\underline{q}}), \tag{5.14}$$

is a section of π_0. Then from (3.5)

$$i_0^*\omega = i_0^*\left(\sum_{i=1}^n dp_i\wedge dq^i\right)$$

$$= \sum_{i=1}^n d(r\check{q}^i)\wedge d(s\check{q}^i)$$

$$= (\check{q}\cdot\check{q})\, dr\wedge ds$$
$$= dr\wedge ds\,,$$

which is the canonical symplectic structure on $T^*\mathbb{R}^+ = \mathbb{R}^+\times\mathbb{R}$. We have therefore shown that the reduced phase space $(J^{-1}(0)/\mathrm{SO}(n),\bar{\omega}_0)$ is just $(T^*(\dot{\mathbb{R}}^n/\mathrm{SO}(n)),\bar{\omega})$, i.e., the conclusions of the Kummer–Marsden–Satzer Theorem hold despite the fact that Φ is not free (see also Ref. 20).

Turn now to prequantization. Since $\dot{\mathbb{R}}^n$ is simply connected for $n>2$, the prequantization line bundle is unique and trivial. But $\pi_1(\mathrm{SO}(n)) = \mathbb{Z}_2$ for $n>2$, so we must check the holonomy of L_0. Let $(q,p)\in J^{-1}(0)$ and consider the orbit $\mathrm{SO}(n)\cdot(q,p)$. As $\mathrm{SO}(n)\cdot(q,p)\approx S^{n-1}$ is simply connected for $n>2$, it follows from Proposition (3.4) that $T^*\Phi$ lifts horizontally to L_0. Thus $\bar{L}_0$ exists and, since the reduced phase space is contractible, $\bar{L}_0$ is also trivial.

For the metalinear structure, the facts that $\dot{\mathbb{R}}^n$ is orientable and simply connected for $n>2$ imply that $\widetilde{FV}$ exists and is unique. Since FV is trivial so is $\widetilde{FV}$. By the remarks following Proposition (3.7) and Theorem (3.8), $T^*\bar{Q}$ is metalinear. Again, since $\mathbb{R}^+\times\mathbb{R}$ is contractible, $\widetilde{\bar{F}V}$ is trivial.

Quantizing this system, we have from Sec. III A that the extended wave functions are

$$\Psi = \psi(q)\lambda\otimes\nu_{\tilde{I}} \tag{5.15}$$

and from Sec. IV C that $\mathscr{h} = L^2(\dot{\mathbb{R}}^n)$. Using (4.2) and (5.11) we compute

$$\mathscr{Q}J_\xi[\Psi] = -i\hbar\left(\sum_{i,j=1}^{n}\xi_{ij}\left(q^i\frac{\partial}{\partial q^j} - q^j\frac{\partial}{\partial q^i}\right)\psi(q)\right)\lambda\otimes\nu_{\tilde{I}}, \tag{5.16}$$

for $\xi\in\wedge^2(\mathbb{R}^n)$. Thus the rotationally invariant states look like

$$\Psi = \psi(\|q\|)\lambda\otimes\nu_{\tilde{I}} \tag{5.17}$$

and, in hyperspherical coordinates, the induced inner product on $\mathscr{h}_0$ is

$$(\Psi,\Phi)_0 = \mathrm{vol}(S^{n-1})\int_0^\infty \psi(r)\phi^*(r)r^{n-1}\,dr \tag{5.18}$$

[compare (4.17)]. Hence $\mathscr{h}_0 = L^2(\mathbb{R}^+,r^{n-1})$.

Similarly, the space $\bar{\mathscr{h}}_0 = L^2(\mathbb{R}^+)$ consists of states

$$\bar{\Psi} = \bar{\psi}(r)\bar{\lambda}_0\otimes\bar{\nu}_{\partial/\partial s}\,. \tag{5.19}$$

Since the group action is not free, we have no set technique as in Sec. IV for constructing an isomorphism $\mathscr{h}_0\to\bar{\mathscr{h}}_0$. Nonetheless, it is clear from (5.17)–(5.19) that

$$\psi(r)\lambda\otimes\nu_{\tilde{I}}\mapsto\sqrt{\mathrm{vol}(S^{n-1})}\;r^{(n-1)/2}\;\psi(r)\bar{\lambda}_0\otimes\bar{\nu}_{\partial/\partial s} \tag{5.20}$$

defines a unitary isomorphism $\mathscr{U}$ of $L^2(\mathbb{R}^+,r^{n-1})$ with $L^2(\mathbb{R}^+)$. Thus we have unitary equivalence.

Now consider, for instance, the radial momentum $p_r = (q\cdot p)/\|q\|$. It is $\mathrm{SO}(n)$-invariant and the reduced observable is $i_0^*p_r = s$. Since p_r preserves V both p_r and s are quantizable. From (3.9) and (3.4)–(3.6) we compute

$$\mathscr{Q}p_r[\Psi] = -\left(i\hbar\left\{\frac{1}{\|q\|}(q\cdot\nabla) + \frac{n-1}{2\|q\|}\right\}\psi(q)\right)\lambda\otimes\nu_{\tilde{I}},$$

for Ψ given by (5.15). On $\mathscr{h}_0 = L^2(\mathbb{R}^+,r^{n-1})$ in hyperspherical coordinates, $\mathscr{Q}p_r$ takes the form

$$\mathscr{Q}p_r[\psi(r)] = -i\hbar\left(\frac{d}{dr}+\frac{n-1}{2r}\right)\psi(r)\,.$$

Likewise, on $\bar{\mathscr{h}}_0 = L^2(\mathbb{R}^+)$,

$$\bar{\mathscr{Q}}s[\bar{\psi}(r)] = -i\hbar\frac{d}{dr}\bar{\psi}(r)\,.$$

It is routine to verify that the unitary map (5.20) intertwines $\mathscr{Q}p_r$ and $\bar{\mathscr{Q}}s$ according to

$$-i\hbar\left(\frac{d}{dr}+\frac{n-1}{2r}\right) = \mathscr{U}^{-1}\left(-i\hbar\frac{d}{dr}\right)\mathscr{U}\,.$$

Theorem (4.3) therefore holds when $n>2$, at least for the radial momentum observable.

When $n=2$ the action Φ is free and we may apply all our previous results. Other than this, the main difference between the cases $n=2$ and $n>2$ is that, since $\mathrm{SO}(2)$ is Abelian, *every* $\mu\in\mathrm{so}(2)^*\approx\mathbb{R}$ is invariant.

Consider the standard connection

$$\alpha = (1/\|q\|^2)\,(q\wedge dq)\,, \tag{5.21}$$

on $\dot{\mathbb{R}}^2 = \mathrm{SO}(2)\times\mathbb{R}^+$. Since α is flat, Kummer–Marsden–Satzer reduction implies that the reduced phase space is $(\mathbb{R}^+\times\mathbb{R},\bar{\omega})$ as before. Composing i_0: $\mathbb{R}^+\times\mathbb{R}\to J^{-1}(0)$ given by (5.14) with the $\mathrm{SO}(2)$-equivariant diffeomorphism $\delta_\mu: J^{-1}(0)\to J^{-1}(\mu)$ given by (2.5) defines a global section

$$i_\mu(r,s) = (r\check{q}, s\check{q} + \mu\alpha(r\check{q})) \tag{5.22}$$

of $\pi_\mu = \pi_0\circ\delta_\mu^{-1}$. Hence

$$J^{-1}(\mu) = \mathrm{SO}(2)\times\mathbb{R}^+\times\mathbb{R}$$

is trivial as a principal $\mathrm{SO}(2)$-bundle.

Now, $H^2(\dot{\mathbb{R}}^2\times\mathbb{R}^2,\mathbb{Z}) = 0$ so that again the prequantization line bundle is unique and trivial. Letting $[c(t)]$ be the generator of $\pi_1(\mathrm{SO}(2)) = \mathbb{Z}$ and using (3.12) and (3.13), we find that the holonomy of L_μ is $\exp((2\pi i/\hbar)\mu)$. Proposition (3.4) then gives rise to a *quantization condition*: L_μ is reducible iff $\mu = m\hbar$ for some integer m. When $\mu = m\hbar$, $\bar{L}_\mu$ is trivial as before.

We must also be careful with the metalinear structures. Since $H^1(\dot{\mathbb{R}}^2\times\mathbb{R}^2,\mathbb{Z}_2) = \mathbb{Z}_2$ there are *two* metalinear frame bundles for V. On the other hand, there is exactly *one* (necessarily trivial) metalinear structure on the reduced phase space. This indicates that one of the metalinear frame bundles on $\dot{\mathbb{R}}^2\times\mathbb{R}^2$ will not project to $\mathbb{R}^+\times\mathbb{R}$ and hence will lead to a spurious quantization.

To construct these metalinear structures, introduce polar coordinates (r,θ) on $\dot{\mathbb{R}}^2$ and set

$$U_+ = \{(r,\theta)\,|\,0<\theta<2\pi\},\quad U_- = \{(r,\theta)\,|\,-\pi<\theta<\pi\}$$

and

$$W_+ = \{(r,\theta)\,|\,0<\theta<\pi\},\quad W_- = \{(r,\theta)\,|\,-\pi<\theta<0\}\,.$$

Since FV is trivial, the transition functions $\tilde{M}_\pm$: $W_\pm\times\mathbb{R}^2\to\mathrm{ML}(2,\mathbb{R})$ for the two $\widetilde{FV}$ are

$$\tilde{M}_+ = \tilde{I},\quad \tilde{M}_- = \tilde{I}\,, \tag{5.23}$$

corresponding to the identity of $H^1(\dot{\mathbb{R}}\times\mathbb{R}^2,\mathbb{Z}_2)$, and

$$\tilde{M}_+ = \tilde{I},\quad \tilde{M}_- = \epsilon\,, \tag{5.24}$$

corresponding to its generator, where

$$\epsilon = \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}.$$

The metalinear frame bundle defined by (5.23) is trivial and the SO(2)-action on FV lifts to this $\widetilde{F}V$ just as when $n > 2$. It is this metalinear structure which projects to the reduced phase space. For the second metalinear frame bundle, it is clear from (5.24) that both $\widetilde{F}V$ and the natural homomorphism $\chi_{SO(2)}: \mathbb{Z} \to \mathbb{Z}_2$ are nontrivial. It follows from Proposition (3.7) that the SO(2)-action does *not* lift to this $\widetilde{F}V$ which therefore does *not* project to $\mathbb{R}^+ \times \mathbb{R}$.

Let $\underline{b} = (\partial/\partial p_r, \partial/\partial p_\theta)$ be a global frame field for V and fix a lift $\underline{\tilde{b}}$ of $\underline{b}$ to the trivial metalinear frame bundle. From (4.7) every polarized section Ψ of $L \otimes \sqrt{\wedge^2 V}$ can be written

$$\Psi = \check{\psi}(r,\theta)\lambda \otimes \nu_{\tilde{b}} . \tag{5.25}$$

Using (5.16) with $\xi = 1$, the quantum constraint $\mathscr{Q}J[\Psi] = m\hbar\Psi$ for $\mu = m\hbar$ becomes

$$-i\hbar \frac{\partial}{\partial\theta}\check{\psi}(r,\theta) = m\hbar\check{\psi}(r,\theta) . \tag{5.26}$$

Thus the physically admissible states take the form $\check{\psi}(r,\theta) = k(r)e^{im\theta}$ consistent with (4.9).

The reduced phase space quantization proceeds exactly as before. The reduced wave functions are given by (5.19) and the isomorphism (4.11) becomes $k(r)e^{im\theta} \mapsto k(r)$.

When $\mu = 0$ these results correlate exactly with those obtained earlier for $n > 2$. The only difference is that here we have used the half-form $\nu_{\tilde{b}}$ rather than $\nu_{\tilde{l}}$ which, according to (3.2), satisfies $\nu_{\tilde{l}} = \sqrt{r}\,\nu_{\tilde{b}}$. Writing Ψ given by (5.25) in the form (5.15) we have that $\psi(r,\theta) = \sqrt{r}\check{\psi}(r,\theta)$. With this change of notation, (5.20) is just (4.19) and all our previous results immediately carry over to the case $n = 2$.

Remarks: (1) It is interesting to see what happens when we quantize the extended phase space using the *nontrivial* metalinear structure. Let $\underline{\tilde{b}}_\pm$ be lifts of $\underline{b}$ to this $\widetilde{F}V$ over $U_\pm \times \mathbb{R}^2$; then from (5.24),

$$\underline{\tilde{b}}_- = \begin{cases} \underline{\tilde{b}}_+, & \text{on } W_+ \times \mathbb{R}^2, \\ \underline{\tilde{b}}_+\epsilon, & \text{on } W_- \times \mathbb{R}^2. \end{cases} \tag{5.27}$$

Defining local sections $\nu_\pm$ of $\sqrt{\wedge^2 V}$ over $U_\pm \times \mathbb{R}^2$ by $\nu_\pm^\#(\underline{\tilde{b}}_\pm) = 1$, it follows from (5.27) and (3.2) that $\nu_- = \pm\nu_+$ on $W_\pm \times \mathbb{R}^2$, respectively. The quantum wave functions are now

$$\Psi|(U_\pm \times \mathbb{R}^2) = \check{\psi}_\pm(r,\theta)\lambda \otimes \nu_\pm ,$$

where

$$\check{\psi}_-(r,\theta) = \pm\check{\psi}_+(r,\theta) \tag{5.28}$$

on $W_\pm$. Such a Ψ satisfies the quantum constraint (5.26) iff

$$\check{\psi}_\pm(r,\theta) = k_\pm(r)e^{im\theta}$$

and (5.28) then implies that $k_-(r) = \pm k_+(r)$ on $W_\pm$, which forces $k_\pm(r) \equiv 0$. Thus, when the nontrivial metalinear structure is used, $\mathscr{H}_\mu = \{0\}$ and we have a spurious quantization.

(2) Earlier we showed that the extended and reduced phase space quantizations of the radial momentum p_r were unitarily related. Of course, this is not really surprising and was in fact guaranteed by our theorems when $n = 2$. It is therefore very curious that the same is *not* true for a rotationally invariant Hamiltonian *except* when $n = 3$.

Set $\mu = 0$ and let the Hamiltonian be

$$h(\underline{q},\underline{p}) = \tfrac{1}{2}\|\underline{p}\|^2 + V(\|\underline{q}\|) .$$

The reduced Hamiltonian on $\mathbb{R}^+ \times \mathbb{R}$ is

$$\bar{h}_0(r,s) = \tfrac{1}{2}s^2 + V(r) .$$

Neither of these is polarization preserving but may nonetheless be quantized using Blattner–Kostant–Sternberg kernels (cf. Chap. 6 of Ref. 21).

On $L^2(\dot{\mathbb{R}}^n)$ we compute

$$\mathscr{Q}h[\Psi] = \{[-(\hbar^2/2)\Delta + V(r)]\psi(\underline{q})\}\lambda \otimes \nu_{\tilde{l}} .$$

In hyperspherical coordinates this reduces to

$$\mathscr{Q}h[\Psi] = \left\{\left[-\frac{\hbar^2}{2}\left(\frac{d^2}{dr^2} + \frac{n-1}{r}\frac{d}{dr}\right) + V(r)\right]\psi(r)\right\}\lambda \otimes \nu_{\tilde{l}} ,$$

for $\Psi \in \mathscr{H}_0$. Similarly,

$$\overline{\mathscr{Q}}\bar{h}_0[\bar{\Psi}] = \left\{\left[-\frac{\hbar^2}{2}\frac{d^2}{dr^2} + V(r)\right]\bar{\psi}(r)\right\}\bar{\lambda}_0 \otimes \nu_{\tilde{\partial}/\partial s}$$

on $L^2(\mathbb{R}^+)$. Using (5.20) it follows that

$$\mathscr{U}^{-1}(\overline{\mathscr{Q}}\bar{h}_0)\mathscr{U} = \mathscr{Q}h - \hbar^2(n-1)(n-3)/8r^2 .$$

Consequently, these two quantizations do not intertwine $\overline{\mathscr{Q}}\bar{h}_0$ and $\mathscr{Q}h$ unless $n = 3$. It would be nice to understand the underlying geometric reason for this.

When $n = 2$ and $\mu = m\hbar$, (5.22) and (5.21) imply that the amended Hamiltonian $\bar{h}_\mu = i_\mu^* h$ is

$$\bar{h}_\mu = \bar{h}_0 + \mu^2/2r^2 .$$

Since now $\Psi = k(r)e^{im\theta}\lambda \otimes \nu_{\tilde{l}}$, the above expressions for the quantum Hamiltonians must be modified by replacing $V(r)$ by $V(r) + m^2\hbar^2/2r^2$. But $\mathscr{Q}h$ and $\overline{\mathscr{Q}}\bar{h}_\mu$ are still not unitarily related.

(3) We have punctured $\mathbb{R}^n$ in order to avoid pathologies. If the origin is not excluded Φ is no longer even effective, $\mathbb{R}^n/SO(n)$ is not a manifold and $J^{-1}(0)$ will be singular. Our entire formalism then fails to apply. For a discussion of this case, see Refs. 25 and 26.

C. Kaluza–Klein electrodynamics

The Kaluza–Klein theory of a relativistic charged particle provides another nice illustration of our formalism. We present only a brief account here; for more details on this and related topics see Refs. 16, 18, 19, 21, and 27.

Let $\bar{Q}$ represent four-dimensional space-time. The configuration space for our charged particle is a left principal T-bundle $\pi_Q: Q \to \bar{Q}$, T being the multiplicative group of complex numbers of modulus one. We identify the Lie algebra of T with $\mathbb{R}$ by associating to each $e \in \mathbb{R}$ the one-parameter group

$$z \mapsto \exp\{i(e_0/\hbar)et\}z , \tag{5.29}$$

where e_0 is a parameter which we interpret as the "elementary" charge.

Suppose Q carries a T-invariant metric g of signature $(+ + + - -)$. Define a connection form α on Q by

$$\alpha(v) = g(1_Q, v) ,$$

for all $v \in TQ$, where 1_Q is the fundamental vector field on Q

corresponding to $1 \in \mathbb{R}$. By Lemma (2.6) there exists a closed two-form F on $\overline{Q}$ such that

$$\pi_{\overline{Q}}^{*} F = d\alpha .$$

We construe F as the electromagnetic field. Since F is the curvature of a circle bundle, the de Rham class $(e_0/h)\ [F]_{\overline{Q}}$ must be integral. This condition may be viewed as a restriction on the allowable interactions of a particle of charge e_0 with the electromagnetic field F in the Kaluza–Klein formalism. Finally we define the space-time metric $\bar{g}$ via

$$\operatorname{hor} g = \pi_{\overline{Q}}^{*}\, \bar{g} .$$

The group action is by construction free and proper and every $e \in \mathbb{R}$ is invariant. We fix the charge of our particle by imposing the *charge constraint* $J = e$ on T^*Q. Kummer–Marsden–Satzer reduction then identifies the reduced phase space $(J^{-1}(e)/T, \bar{\omega}_e)$ with $(T^*\overline{Q}, \overline{\Omega}_e)$; here

$$\overline{\Omega}_e = \bar{\omega} + e\tau_{\overline{Q}}^{*} F$$

is just the *charged symplectic structure* on $T^*\overline{Q}$.

Let the prequantization line bundle be $L = T^*Q \times \mathbb{C}$. Mimicking the calculation in the previous example while taking the precise form of (5.29) into account, we find that the holonomy of L_e is $\exp(2\pi i(e/e_0))$. In the Kaluza–Klein framework, then, the lifting criteria become *superselection rules*: L_e is reducible iff the particle's charge e is an integral multiple of the elementary charge e_0. When $e = ne_0$, the induced line bundle $\overline{L}_e$ is also trivial. As an aside, notice how the integrality condition on $(e_0/h)\ [F]_{\overline{Q}}$ and the superselection rule $e = ne_0$ combine to guarantee the integrality of $(1/h)[\overline{\Omega}_e]_{T^*\overline{Q}} = (e/h)[F]_{\overline{Q}}$ as required for the quantizability of the reduced phase space.

Now assume that $\overline{Q}$, and hence Q, is orientable. Using Proposition (3.6) and the metric g, we reduce the structure group $GL(5,\mathbb{R})$ of FV to $SO(3,2)$. Now $SO(3,2)$ is isomorphic to the intersection of $\sigma^{-1}(SO(3,2))$ with the component of the identity in $ML(5,\mathbb{R})$. Thus the transition functions for FV, valued in $SO(3,2)$, can be lifted to $\sigma^{-1}\ (SO(3,2)) \subset ML(5,\mathbb{R})$ thereby defining a metalinear frame bundle $\widetilde{F}V$. The characteristic homomorphism of $\widetilde{F}V$ so defined is obviously trivial. This and the orientability of Q imply that the associated bundle $\surd \wedge^5 V$ of half-forms is trivial.

Proposition (3.7) and Theorem (3.8) guarantee that $\widetilde{F}V$ projects to a metalinear frame bundle $\widetilde{F}\overline{V}$ on $T^*\overline{Q}$, which is exactly that constructed in a similar fashion to $\widetilde{F}\overline{V}$ by reducing the structure group $GL(4,\mathbb{R})$ of $F\overline{V}$ to $SO(3,1)$ using the space-time metric $\bar{g}$. The half-form bundle $\surd \wedge^4 \overline{V}$ is likewise trivial.

Set $\Psi = \psi\lambda \otimes \nu_g$, where ν_g is defined as follows. Fix a positively oriented orthonormal frame $\underline{b}$ for FQ, where b_5 is tangent to the fibers of π_Q, and denote also by $\underline{b}$ the corresponding frame in FV (cf. Sec. III E). Then let ν_g be such that $\nu_g^{\#}(\tilde{\underline{b}}) = 1$. It follows from Sec. IV C and the triviality of both L and $\surd \wedge^5 V$ that $\mathscr{H} = L^2(Q, \sqrt{\det g})$. Similarly, we have $\overline{\Psi} = \bar{\psi}\bar{\lambda}_e \otimes \bar{\nu}_{\bar{g}}$ for $\overline{\Psi} \in \bar{\mathscr{H}}_e = L^2(\overline{Q}, \sqrt{\det \bar{g}})$.

In a chart (q,z) on Q, where the q are space-time coordinates and z is the T-coordinate, the quantum constraint

$$\mathscr{Q}J[\Psi] = ne_0\Psi$$

becomes

$$-i\hbar \frac{\partial}{\partial z}\psi(q,z) = ne_0\psi(q,z) .$$

Thus the Kaluza–Klein quantum state space for a particle with charge $e = ne_0$ consists of wave functions

$$\Psi = k(q)\ e^{(i/\hbar)ne_0 z}\lambda \otimes \nu_g .$$

Since T is compact, the Unitary Equivalence Theorem asserts that the correspondence

$$k(q)e^{(i/\hbar)ne_0 z}\lambda \otimes \nu_g \mapsto k(q)\bar{\lambda}_e \otimes \bar{\nu}_{\bar{g}}$$

defines a unitary isomorphism of $L^2(Q, \sqrt{\det g})$ with $L^2(\overline{Q}, \sqrt{\det \bar{g}})$.

Our theorems therefore guarantee that the quantizations of a relativistic charged particle with $e = ne_0$ in both the Kaluza–Klein formalism and the conventional space-time-based approach are unitarily equivalent. Since all ordinary polarization-preserving observables—viz., the position, linear and angular momenta—are T-invariant, Theorem (4.3) shows that they may be equally well quantized in either formalism.

VI. DISCUSSION

We have proven theorems to the effect that one can quantize either the extended or reduced phase space of a constrained cotangent system with unitarily equivalent results. The examples in the previous section demonstrate the utility of our formalism. Here we briefly overview our constructions and conclusions with an eye to possible generalizations and improvements.

We begin by reexamining the conditions under which our formalism operates. These are (1) G must admit a bi-invariant metric and the action of G on Q must be free and proper, (2) $\mu \in \mathscr{g}^*$ must be invariant, and (3) the geometric quantization structures must be G-invariant.

Regarding (1), the only really severe restriction is that the action be free. In fact, virtually all our results are predicated upon this assumption although, as the $n > 2$ angular momentum example shows, our theorems *may* be valid without it. One might try to weaken this hypothesis as in Montgomery,[20] but it is not clear to what extent this is workable.

As noted earlier, condition (2) serves a dual purpose. It guarantees classically that the reduction of a cotangent bundle is again a cotangent bundle and quantum mechanically that one obtains a representation of $\mathscr{g}$ on $\mathscr{H}$. The possibility that the reduced phase space is not a cotangent bundle is not a problem in principle, although one then of course loses much of the structure that so simplified our formalism. On the quantum level the noninvariance of μ would not necessarily be a disaster either, since one can *always* find another extended phase space in which the constraint set is imbedded coisotropically.[7] One can then consistently quantize this new constrained system, but the price is that one will lose both the group-theoretical and cotangent bundle structures in the process.

Our last condition (3) on the invariance of the quantization structures is vital. As the examples show, one either cannot quantize or obtains spurious quantizations if the G-action does not lift appropriately to both the prequantization line and metalinear frame bundles. It would be interesting to know if every such structure on the reduced phase space arises by projection from an invariant one on the extended phase space and, conversely, whether every invariant such structure on T^*Q is the pullback of a compatible one on $T^*\bar{Q}$.

To what extent can our results be expected to carry over to more general settings? The simplest modification to our framework is to allow for other types of polarizations. This should not cause too much difficulty provided P is G-invariant, has simply connected leaves and intersects $J^{-1}(\mu)$ sufficiently regularly. Two distinguished possibilities are polarizations P which satisfy either

$$P \cap TJ^{-1}(\mu) = \{0\}$$

or

$$P|J^{-1}(\mu) \subseteq TJ^{-1}(\mu).$$

One could also consider nonreal polarizations.

Of critical importance, however, is that the polarization be chosen in such a way that every quantum wave function is uniquely determined by its restriction to the constraint set. In essence, this means that the extended phase space quantization must be totally insensitive to what happens "off" $J^{-1}(\mu)$. This requirement seems reasonable, since in principle only those classical states contained in the constraint set are physically permissible and/or relevant. For further discussion of these matters, see Ref. 7. In any case, this condition played a key role in the proof of the Smooth Equivalence Theorem. Without it there is no effective way to properly correlate the extended and reduced phase space quantizations which will then, in general, be wildly incompatible. An interesting—and physically meaningful—illustration of the consequences of violating this condition is given by Ashtekar and Horowitz.[5] An even more bizarre example is studied in Gotay.[28]

The next step is to consider arbitrary constrained systems with or even without symmetry. The problem is now much more difficult since we cannot explicitly construct anything and no longer necessarily have at our disposal well-behaved polarizations. What is known in this general case is summarized in Refs. 7, 8, and 10–12.

ACKNOWLEDGMENTS

This work was supported by a grant from the United States Naval Academy Research Council.

[1]M. J. Gotay, J. M. Nester, and G. Hinds, "Presymplectic manifolds and the Dirac–Bergmann theory of constraints," J. Math. Phys. **19**, 2388 (1978).

[2]K. Sundermeyer, *Constrained Dynamics, Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982).

[3]M. J. Gotay, J. A. Isenberg, J. E. Marsden, R. Montgomery, J. Sniatycki, and P. Yasskin, "Constraints and momentum mappings in relativistic field theory," in preparation, 1986.

[4]J. E. Marsden and A. Weinstein, "Reduction of symplectic manifolds with symmetry," Rep. Math. Phys. **5**, 121 (1974).

[5]A. Ashtekar and G. T. Horowitz, "On the canonical approach to quantum gravity," Phys. Rev. D **26**, 3342 (1982).

[6]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Academic, New York, 1964).

[7]M. J. Gotay and J. Sniatycki, "On the quantization of presymplectic dynamical systems via coisotropic imbeddings," Commun. Math. Phys. **82**, 377 (1981).

[8]V. Guillemin and S. Sternberg, "Geometric quantization and multiplicities of group representations," Invent. Math. **67**, 515 (1982).

[9]M. Puta, "On the reduced phase space of a cotangent bundle," Lett. Math. Phys. **8**, 189 (1984).

[10]J. Sniatycki, "Constraints and quantization," in *Lecture Notes in Mathematics*, Vol. 1037 (Springer, Berlin, 1983), pp. 301–334.

[11]I. Vaisman, "Geometric quantization of presymplectic manifolds," Mh. Math. **96**, 293 (1983).

[12]N. M. J. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).

[13]M. Kummer, "On the construction of the reduced phase space of a Hamiltonian system with symmetry," Indiana Univ. Math. J. **30**, 281 (1981).

[14]W. J. Satzer, "Canonical reduction of mechanical systems invariant under Abelian group actions with an application to celestial mechanics," Indiana Univ. Math. J. **26**, 951 (1977).

[15]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin–Cummings, Reading, MA, 1978).

[16]J. E. Marsden, *Lectures on Geometric Methods in Mathematical Physics, NSF–CBMS Regional Conference Series in Mathematics*, Vol. 37 (SIAM, New York, 1981).

[17]J. Sniatycki and A. Weinstein, "Reduction and quantization for singular momentum mappings," Lett. Math. Phys. **7**, 155 (1983).

[18]R. Montgomery, "Canonical formulations of a classical particle in a Yang–Mills field and Wong's equations," Lett. Math. Phys. **8**, 59 (1984).

[19]J. Sniatycki, "On Hamiltonian dynamics of particles with gauge degrees of freedom," Hadronic J. **2**, 642 (1979).

[20]R. Montgomery, "The structure of reduced cotangent phase spaces for non-free group actions," Univ. of California at Berkeley Math. Dept. preprint. 1983.

[21]J. Sniatycki, *Geometric Quantization and Quantum Mechanics*, in *Applied Mathematical Sciences Series*, Vol. 30 (Springer, Berlin, 1980).

[22]B. Kostant, *Symplectic Spinors*, in *Symposia Mathematica*, Vol. 14 (Academic, London, 1974), pp. 139–152.

[23]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, NJ, 1951).

[24]R. C. Robinson, "Fixing the center of mass in the N-body problem by means of a group action," Colloquium International C.N.R.S., Vol. 237 (C.N.R.S., Paris, 1975).

[25]L. Bos and M. J. Gotay, *Reduced Canonical Formalism for a Particle with Zero Angular Momentum, Proceedings of the XIII Colloquium on Group Theoretical Methods in Physics* (World Scientific, Singapore, 1984), pp. 83–91.

[26]L. Bos and M. J. Gotay, "Singular angular momentum mappings," to appear in J. Diff. Geom.

[27]J. Sniatycki, "Kinematics of particles with isotopic spin," Hadronic J. **3**, 743 (1980).

[28]M. J. Gotay, "Negative Energy States in Quantum Gravity?," to appear in Class. Quantum Grav.

Wigner quantum systems. Two particles interacting via a harmonic potential. I. Two-dimensional space

A. H. Kamupingene
Institute of Nuclear Research and Nuclear Energy, 1184 Sofia, Bulgaria
T. D. Palev[a)]
International Centre for Theoretical Physics, Trieste, Italy
S. P. Tsaneva
Higher Pedagogical Institute, Shumen, Bulgaria

(Received 19 November 1985; accepted for publication 30 January 1986)

A noncanonical quantum system, consisting of two nonrelativistic particles, interacting via a harmonic potential, is considered. The center-of-mass position and momentum operators obey the canonical commutation relations, whereas the internal variables are assumed to be the odd generators of the Lie superalgebra sl(1,2). This assumption implies a set of constraints in the phase space, which are explicitly written in the paper. All finite-dimensional irreducible representations of sl(1,2) are considered. Particular attention is paid to the physical representations, i.e., the representations corresponding to Hermitian position and momentum operators. The properties of the physical observables are investigated. In particular, the operators of the internal Hamiltonian, the relative distance, the internal momentum, and the orbital momentum commute with each other. The spectrum of these operators is finite. The distance between the constituents is preserved in time. It can take no more than three different values. For any non-negative integer or half-integer l there exists a representation, where the orbital momentum is l (in units of $2\hbar$). The position of any one of the particles cannot be localized, since the operators of the coordinates do not commute with each other. The constituents are smeared with a certain probability within a finite surface, which moves with a constant velocity together with the center of mass.

I. INTRODUCTION

In the last years there seems to be slow but increasing interest in the study of noncanonical quantum systems mainly in the frame of quantum mechanics. The interest in this field stems from the observations that it opens new, uninvestigated-up-to-now possibilities for theoretical explanation of certain subnuclear phenomena, which have not been properly understood so far. Yet, one has to acknowledge the results in this approach are still modest, more of a philosophical, purely theoretical interest. They have not reached the level to propose any convincing explanation, overcoming some of the difficulties in the contemporary high energy physics.

A common feature of the various approaches to noncanonical quantum mechanics is the assumption that the position and the momentum operators x_i, p_i, $i = 1,...,n$, of the particles do not satisfy the canonical commutation relations (CCR's)

$$[x_j,p_k] = i\hbar\delta_{jk}, \quad [x_i,x_j] = [p_i,p_j] = 0. \qquad (1.1)$$

The main ideas of the current approaches, which fall essentially into three groups, were first formulated and worked out by Schrödinger,[1] Weyl,[2] and Wigner.[3] The Schrödinger approach is intensively developed by Barut and co-workers.[4] Preserving the essence of the Schrödinger ideas, these authors consider the Dirac electron as a composite system. The center of charge of the system $\mathbf{x}(t)$ does not necessarily coincide with the center of mass $\mathbf{X}(t)$. The vector $\mathbf{Q}(t) = \mathbf{x}(t) - \mathbf{X}(t)$ is describing the high oscillation of the charge around the center of mass (the *Zitterbewegung*). Assuming that $Q_i(t)$ and their conjugate momenta $P_i(t)$, $i = 1,2,3$, are noncanonical operators and more precisely that they generate under commutation the four-dimensional spinor representation of the Lie algebra so(3 + 2), the authors derive the Dirac equation. The spin of the electron is interpreted as the angular momentum of the relative motion. Other representations or realizations of so(3 + 2) or of its other real forms lead to systems like an H atom or a hadron.[5–7] Enlargements of the algebra to so(n + 2) for n degrees of freedom allow kinematical description of composite atoms or hadrons,[5] relativistic oscillator or rotator,[6] etc.

The Weyl approach to quantization was discovered after putting the CCR's in a mathematically rigorous form.[8] It is well known that the relations (1.1) have only one representation in the case of a finite or countable number of operators. This is the infinite-dimensional Fock representation. Since, however, the (unbounded) operators x_i, p_k are not defined everywhere in the Fock space, one concludes that $[x_j,p_k] \subset i\hbar$, i.e., the relations (1.1) do not hold in the strict sense. To be rigorous, Weyl replaced the CCR's with a new, equivalent (however, no difficulties with the domains) set of relations

$$\exp\Big(\frac{1}{\hbar}\alpha_j p_j\Big)\exp\Big(\frac{i}{\hbar}\beta_k q_k\Big) = \exp\Big(\frac{i}{\hbar}\alpha_j\beta_k\Big)\exp\Big(\frac{i}{\hbar}\beta_k q_k\Big)\exp\Big(\frac{i}{\hbar}\alpha_j p_j\Big). \qquad (1.2)$$

[a)] Permanent address: Institute of Nuclear Research and Nuclear Energy, Boul. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/86/082067-09$02.50

It turned out, however, that the form (1.2) of the canonical variables meant more than a sole mathematical rigor. As Weyl pointed out, one can satisfy Eqs. (1.2) (and this is now a definition for a representation of the position and the momentum operators) with x_j, p_k acting as Hermitian operators in finite-dimensional spaces. For each i the numbers α_i, β_i run over a finite set. In the recent years the Weyl quantization has been successfully developed by Santhanam, Jagannathan, Tekumalla, and Vasudevan.[9] Similar ideas are worked out in Refs. 10 and 11.

The Wigner noncanonical approach originates from a remark of Ehrenfest,[12] who observed that (in the Heisenberg picture), if one assumes the CCR's (1.1) to be valid, then the Heisenberg equations of motion

$$\dot{p}_k = -(i/\hbar)[p_k, H], \quad \dot{q}_k = -(i/\hbar)[q_k, H] \tag{1.3}$$

hold if and only if the Hamiltonian equations

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}, \quad \dot{q}_k = \frac{\partial H}{\partial p_k} \tag{1.4}$$

are fulfilled. To come to the generalization, discussed by Wigner,[3] let us formulate the result of Ehrenfest in the following way. Consider the following statements.

(1) The Heisenberg equations (1.3) hold.

(2) The Hamiltonian equations (1.4) hold.

(3) The canonical commutation relations (1.1) hold.

Then (Ehrenfest[12]) each one of the first two statements is a consequence of the other two. Observing this and noting that Eqs. (1.3) and (1.4) have a more immediate physical significance than the CCR's (1.1), Wigner asked the question[3]: Is each one from the above statements a consequence of the other two? In an example of a one-dimensional harmonic oscillator he has shown that the answer to the above question was negative. The CCR's are not a consequence of the Hamiltonian and the Heisenberg equations. These equations are equivalent for a much larger class of position and momentum operators (and in particular, for those satisfying the CCR's). The quantum pictures corresponding to such operators may turn out to be of interest also. The quantum systems corresponding to such generalized position and momentum $q_i, p_i, i = 1,...,n$, will be called Wigner quantum systems. More precisely, by a Wigner quantum system (WQS) we understand a quantum system with q_i, p_i defined as self-adjoint operators in such a way that the Heisenberg and the Hamiltonian equations (1.3) and (1.4)] are equivalent, they appear as different forms of one and the same equation. As we shall also see in this paper, one can find a WQS with position and momentum operators, defined in a finite-dimensional state space, i.e., also in this case one can consider finite quantum systems.

One way to construct a WQS is to use the correspondence principle. To this end consider a classical Hamiltonian

$$H = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + U(q_1,...,q_n) \tag{1.5}$$

with an even potential $U(-q_1,..., -q_n) = U(q_1,...,q_n)$, which we assume for simplicity to be a polynomial of the coordinates. Represent $U(q_1,...,q_b)$ (in one of the several possible ways) as a function of anticommutators $\{q_i, q_j\}$, $i,j = 1,...,n$, and replace all p_i, q_i by operators, defined as follows:

$$\hat{q}_i = (1/\sqrt{2})(a_i^+ + a_i^-), \quad \hat{p}_i = (i/\sqrt{2})(a_i^+ - a_i^-), \tag{1.6}$$

where

$$[\{a_i^\xi, a_j^\eta\}, a_k^\epsilon] = \delta_{ik}(\epsilon - \xi)a_j^\eta + \delta_{jk}(\epsilon - \eta)a_i^\xi. \tag{1.7}$$

Here and throughout the paper $\xi,\eta,\epsilon,\delta = \pm$ or ± 1, $[x,y] = xy - yx$, and $\{x,y\} = xy + yx$. In this way one obtains a WQS.

The operators (1.7) are known in quantum field theory. They were introduced by Green[13] as a possible generalization of the statistics of integer-spin fields and are called para-Bose operators. Their irreducible representations, corresponding to Hermitian position and momentum operators and a nondegenerate ground state $|0\rangle$, are infinite dimensional and are labeled by one non-negative integer p, the order of the statistics.[14] Only for $p = 1$ the position and the momentum operators are canonical. For definiteness we call the WQS's with operators (1.6) para-Bose quantum systems (PBQS's).

The noncanonical one-dimensional oscillator, considered by Wigner, was a PBQS with one q and one p. The infinite number of solutions he found gave different representations of the same operators.[15] In recent years some PBQS's were studied in Refs. 16 and 17.

In the present paper, we consider a model of a Wigner quantum system, which is not a PBQS. In a two-dimensional space we study the behavior of two nonrelativistic point particles interacting via a harmonic potential. Such a system in the more realistic three-dimensional space was introduced by one of the authors (T.D.P.) in Refs. 15 (hereafter referred to as I) and (18). The physical properties of this system were investigated for a certain, in fact, very poor class of representations of the position and the momentum operators. Here we simplify the model, going to a two-dimensional space. On the other hand, however, we investigate the physical consequences for all irreducible representations of the underlying algebraic structure.

Mathematically the problem reduces to the determination of the so-called star representation[19] of the special linear Lie superalgebra (LS) sl(1,2). This LS is generated from the internal position and momentum operators. It does not appear as a result of some other physical assumptions. We postulate it more on a logical background. To see where the idea comes from consider n pairs $a_1^\pm,...,a_n^\pm$ of Bose creation and annihilation operators (CAO's) or the corresponding to them [according to (1.6)] position and momentum operators $q_i, p_i, i = 1,...,n$. Then (Ref. 15) the linear span of $a_i^\pm$ and their anticommutators $\{a_i^\xi, a_j^\eta\}$, $i,j = 1,...,n$; $\xi,\eta = \pm$, give rise to one particular infinite-dimensional representation of the orthosymplectic LS, which in the Kac notation[20] is $B(0,n)$. Other representations of $B(0,n)$ lead to different-order para-Bose operators.[21] In the same way n pairs of Fermi or para-Fermi operators give rise to representations of the orthogonal Lie algebra B_n (see Ref. 22). There exists a one-to-one correspondence between the representations of n pairs of para-Bose (resp. para-Fermi) operators and the representations of the LS $B(0,n)$ [resp. of the Lie algebra B_n, which in the Kac classification belongs to the class of the orthosymplectic LS's (see Ref. 23)]. The understanding of the Lie algebraical structure of the canonical quantum theory and its generalization along a parastatistics line suggests immediately some new ideas. To this end we first observe that the orthosymplectic LS's constitute one of the (essentially) four infinite classes of basis Lie superalgebras A,B,C,D (see Ref. 23), namely the class B. The physical state spaces are representation spaces of LS's from B. The dynamical variables (in the canonical and parastatistical cases) are functions of the generators of LS's from the class B. This observation states in a natural way the question of whether one can construct quantum systems and in particular a WQS with dynamical variables, which are functions of other LS's and in particular of other basic LS's. In the frame of the quantum field theory (QFT) a positive answer to this question was given in Refs. 24, where it was shown that to every simple class of Lie algebras one can put in correspondence a field quantization. In case of proper LS's only the class A, consisting of the special linear LS's, was investigated. In the frame of the quantum mechanics (QM) we refer to Refs. 15 and 18. The present paper is also an example of a WQS of the type A. The quantum systems considered in Refs. 4–7 are B-type quantum systems. In the finite-dimensional cases, they can be expressed as para-Fermi realizations. Also of this type is the generalization of the Heisenberg commutation relations, which in the nonrelativistic limit reduces to the CCR's (see Ref. 25). Also interesting, although in a somewhat different spirit, is the high-energy generalization of the CCR's, proposed recently by Saavedra and Utreras.[26]

II. A TWO-PARTICLE OSCILLATOR AS A WIGNER QUANTUM SYSTEM

A. Formulation of the problem

We wish to define in a rigorous way a noncanonical Wigner quantum system in a two-dimensional space with a Hamiltonian

$$H_{\text{tot}} = \frac{\mathbf{p}_1^2}{2m_j} + \frac{\mathbf{p}_2^2}{2m_2} + \frac{m\omega^2}{2}(\mathbf{r}_1 - \mathbf{r}_2)^2. \tag{2.1}$$

This Hamiltonian corresponds to two nonrelativistic point particles with masses m_1 and m_2, interacting via a harmonic potential. The requirement that the oscillator is a quantum system puts certain restrictions on the operators involved. First of all the radius vectors of the particles $r_i = (x_i, y_i)$, the corresponding momenta, and, more generally, any dynamical variable F have to be defined as self-adjoint operators in a Hilbert space, which is the physical state space. Second, the Hamiltonian and the total momentum $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$ should be infinitesimal generators of the unitary groups of time and space translations

$$U(t) = e^{-(i/\hbar)Ht}, \quad U(\mathbf{x}) = e^{(i/\hbar)\mathbf{P}\mathbf{x}}, \tag{2.2}$$

respectively. Therefore, in the Heisenberg picture (which we shall use) the time evolution of any observable F is governed by the Heisenberg equation of motion

$$\dot{F} = -\frac{i}{\hbar}[F,H], \quad \dot{F} = \frac{dF}{dt}. \tag{2.3}$$

In particular,

$$\dot{\mathbf{p}}_k = \frac{i}{\hbar}[\mathbf{p}_k, H], \quad \dot{\mathbf{r}}_k = -\frac{i}{\hbar}[\mathbf{r}_k, H], \quad i = 1,2. \tag{2.4}$$

We require, moreover, the oscillator to be a Wigner quantum system, i.e., the Hamiltonian equations should also hold:

$$\dot{\mathbf{p}}_k = -\frac{\partial H}{\partial \mathbf{x}_k}, \quad \dot{\mathbf{r}}_k = \frac{\partial H}{\partial \mathbf{p}_k}. \tag{2.5}$$

Introduce the center-of-mass (CM) coordinates

$$\mathbf{R} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1 + m_2), \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \tag{2.6}$$

and let $\mu = m_1 + m_2$ and m be the total and the reduced masses, $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$ and $\mathbf{P}$ be the total and the internal (the conjugate to $\mathbf{r}$) momentum, respectively, $r = |\mathbf{r}_1 - \mathbf{r}_2|$. Then the Hamiltonian is a sum of the CM Hamiltonian H_{CM} and the internal Hamiltonian H,

$$H_{\text{tot}} = H_{\text{CM}} + H, \tag{2.7}$$

where

$$H_{\text{CM}} = \mathbf{P}^2/2\mu, \quad H = \mathbf{p}^2/2m + m\omega^2 r^2/2. \tag{2.8}$$

Similarly, the angular momentum splits, $M_{\text{tot}} = M_{\text{CM}} + M$. In this notation one has Heisenberg equation,

$$\dot{\mathbf{P}} = (i/\hbar)[\mathbf{P}, H_{\text{CM}} + H], \quad \dot{\mathbf{R}} = -(i/\hbar)[\mathbf{R}, H_{\text{CM}} + H], \tag{2.9}$$

$$\dot{\mathbf{p}} = -(i/\hbar)[\mathbf{p}, H_{\text{CM}} + H], \quad \dot{\mathbf{r}} = -(i/\hbar)[\mathbf{r}, H_{\text{CM}} + H]; \tag{2.10}$$

and Hamiltonian equations,

$$\dot{\mathbf{P}} = 0, \quad \dot{\mathbf{R}} = \mathbf{P}/\mu, \tag{2.11}$$

$$\dot{\mathbf{p}} = -m\omega^2\mathbf{r}, \quad \dot{\mathbf{r}} = \mathbf{p}/m. \tag{2.12}$$

Following I, we assume first that the center-of-mass variables $\mathbf{R}$ and $\mathbf{P}$ commute with the internal variables $\mathbf{r}$ and $\mathbf{p}$. Second, we postulate that $\mathbf{R}$ and $\mathbf{P}$ satisfy the CCR's (1.1). Then we are left with the Heisenberg equations (2.13) and the Hamiltonian equations (2.14) for the internal $\mathbf{r}$ and $\mathbf{p}$

$$\dot{\mathbf{p}} = -(i/\hbar)[\mathbf{p}, H], \quad \dot{\mathbf{r}} = -(i/\hbar)[\mathbf{r}, H], \tag{2.13}$$

$$\dot{\mathbf{p}} = -m\omega^2\mathbf{r}, \quad \dot{\mathbf{r}} = \mathbf{p}/m. \tag{2.14}$$

Equations (2.13) and (2.14) are compatible if at any time

$$[H, p_k] = i\hbar m\omega^2 r_k,$$
$$[H, r_k] = -(i\hbar/m)p_k, \quad k = 1,2. \tag{2.15}$$

These equations have to be fulfilled, if one wants the oscillator to be a Wigner quantum system.

B. Lie superalgebraical realization

We now proceed to establish a realization of the internal variables in terms of the generators of the Lie superalgebra sl(1,2). This realization will satisfy the requirements (2.15) for the system to be a WQS. It is, however, by no means the most general one. One can study realizations with generators of other Lie algebras or Lie superalgebras. Nevertheless, we shall see that the sl(1,2) realization leads to infinitely many nonequivalent Wigner quantum systems.

To begin with we recall the definition of the special linear Lie superalgebra sl(1,2). Let e_{AB}, $A,B = 0,1,2$, be a 3×3 matrix with 1 on the A th row and B th column and 0 elsewhere. Then sl(1,2) is the linear span of the odd generators

$$e_{0k}, e_{k0}, \quad k=1,2, \tag{2.16}$$

and all their anticommutators

$$E_{ij} = \{e_{i0}, e_{0j}\} = e_{ij} + \delta_{ij} e_{00}. \tag{2.17}$$

The matrices E_{ij}, $i,j=1,2$, constitute a basis in the even subalgebra, which is isomorphic to the Lie algebra gl(2). The product $[\![\,,\,]\!]$ on sl(1,2) is defined in terms of commutators or anticommutators

$$[\![a,b]\!] = ab - (-1)^{\alpha\beta} ba, \tag{2.18}$$

where a and b are any two homogeneous elements of degree α and β, respectively. Using the relation

$$e_{AB}\, e_{CD} = \delta_{BC}\, e_{AD} \tag{2.19}$$

and the defining relations (2.16) and (2.17), one easily derives the (anti)commutation relations in sl(1,2)

$$\begin{aligned}
&[E_{ij}, e_{k0}] = \delta_{jk} e_{i0} - \delta_{ij} e_{k0},\\
&[E_{ij}, e_{0k}] = -\delta_{ik} e_{0j} + \delta_{ij} e_{0k},\\
&[e_{i0}, e_{0j}] = E_{ij}, \quad \{e_{i0}, e_{j0}\} = \{e_{0i}, e_{0j}\} = 0.
\end{aligned} \tag{2.20}$$

We now give the following realization of the internal coordinates and momenta:

$$\begin{aligned}
&r_1 = (\hbar/2m\omega)^{1/2}(e_{10}+e_{01}+e_{20}+e_{02}),\\
&p_1 = i(m\omega\hbar/2)^{1/2}(-e_{10}+e_{01}-e_{20}+e_{02}),\\
&r_2 = i(\hbar/2m\omega)^{1/2}(e_{10}-e_{01}-e_{20}+e_{02}),\\
&p_2 = (m\omega\hbar/2)^{1/2}(e_{10}+e_{01}-e_{20}-e_{02}).
\end{aligned} \tag{2.21}$$

The inverse relations read

$$\begin{aligned}
e_{k0} &= (m\omega/8\hbar)^{1/2}[r_1 + i(-1)^k r_2]\\
&\quad + i(8m\omega\hbar)^{-1/2}[p_1 + i(-1)^k p_2],\\
e_{0k} &= (m\omega/8\hbar)^{1/2}[r_1 - i(-i)^k r_2]\\
&\quad + i(8m\omega\hbar)^{-1/2}[-p_1 + i(-1)^k p_2].
\end{aligned}$$

In terms of these generators one obtains the following expressions for the internal Hamiltonian H and the compatibility conditions (2.15):

$$H = \omega\hbar(E_{11}+E_{22}), \tag{2.22}$$

$$[E_{11}+E_{22}, e_{k0}] = -e_{k0}, \quad [E_{11}+E_{22}, e_{0k}] = e_{0k}. \tag{2.23}$$

From the sl(1,2) (anti)commutation relations (2.20) one immediately concludes that Eqs. (2.23) hold. We recall, however, that our considerations are in the Heisenberg picture. The position and the momentum operators and, hence, also the generators of sl(1,2) (considered as dynamical variables) depend on time. The Heisenberg equations (2.13) and the Hamiltonian equations (2.14) will be equivalent if the compatibility relations (2.22) hold at any time, i.e., if they are fulfilled as equal time commutation relations. One easily checks that this is the case. To show it we observe that in terms of the generators the Hamiltonian equations (2.14) are

$$\dot{e}_{k0}(t) = -i\omega e_{k0}(t), \quad \dot{e}_{0k}(t) = i\omega e_{0k}(t). \tag{2.24}$$

Hence

$$e_{k0}(t) = \exp(-i\omega t) e_{k0}(0), \quad e_{0k}(t) = \exp(i\omega t) e_{0k}(0). \tag{2.25}$$

Inserting this solution in (2.20) we see that the (anti)commutation relations of sl(1,2) can be considered as equal time relations. Hence, also Eqs. (2.23) hold at any time. Observe that the even generators and, therefore, all the gl(2) subalgebras are time independent. In particular the Hamiltonian (as it should be) does not depend on time.

To turn the two-particle oscillator into a quantum system it remains to determine the physical state space of the system as a Hilbert space V and to represent the position and the momentum operators as self-adjoint operators in V. The relations between r_i and p_i, which have to hold in any state space, follow from (2.20). The three linear relations from (2.20) read

$$\begin{aligned}
&[\{r_i,r_j\},r_k] = (i\hbar/m^2\omega^2)(\delta_{ik}p_j + \delta_{jk}p_i - 2\delta_{ij}p_k),\\
&[\{r_i,r_j\},p_k] = i\hbar(-\delta_{ik}r_j - \delta_{jk}r_i + 2\delta_{ij}r_k),\\
&[\{r_i,p_j\},r_k] = i\hbar(\delta_{jk}r_i - \delta_{ik}r_j),\\
&[\{r_i,p_j\},p_k] = i\hbar(\delta_{jk}p_i - \delta_{ik}p_j),\\
&[\{p_i,p_j\},r_k] = i\hbar(\delta_{ik}p_j + \delta_{jk}p_i - 2\delta_{ij}p_k),\\
&[\{p_i,p_j\},p_k] = i\hbar m^2\omega^2(-\delta_{jk}r_i - \delta_{ik}r_j + 2\delta_{ij}r_k).
\end{aligned} \tag{2.26}$$

For the anticommutation relations in (2.20) one obtains

$$\begin{aligned}
&\{r_k,p_k\} = 0, \quad k=1,2,\\
&\{r_1,p_2\} + \{r_2,p_1\} = 0,\\
&(1/2m)\{p_i,p_j\} - (m\omega^2/2)\{r_i,r_j\} = 0, \quad i,j=1,2.
\end{aligned} \tag{2.27}$$

The relations (2.26) and (2.27) should not be considered as a minimal set. Some of them are dependent [in view of the third relation (2.27), for instance, the first two and the last two relations in (2.26) are equivalent]. Of particular interest is the last relation (2.27) if $i=j=1,2$. Then

$$(1/2m)p_i^2 - (m\omega^2/2)r_i^2, \quad i=1,2.$$

In the classical case these equations have no bound state solutions. We shall see that in the quantum case the behavior of the particles is different. In view of all constraints (2.26) and (2.27) the spectrum of the internal oscillator is discrete and even finite.

As in Refs. 4, 5, and 25, in our case the internal coordinates do not commute with each other:

$$[r_1,r_2] \neq 0, \quad [p_1,p_2] \neq 0.$$

Indeed, if $[r_1,r_2]=0$, then the right-hand side of (2.26a) would have been equal to zero.

Eqs. (2.26) and (2.27) are equivalent to the (anti)commutation relations (2.20), defining the Lie superalgebra sl(1,2). Therefore, the problem of determining all operators p_i and r_i that satisfy (2.26) and (2.27) means that one has to consider all representations of sl(1,2). The requirement that

r_i and p_i are self-adjoint operators imposes certain limitations. To show this we observe that (A^+ denotes the adjoint to the A operator) the conditions

$$p_i^+ = p_i, \quad r_i^+ = r_i, \quad i = 1,2,$$

hold if and only if

$$(e_{k0})^+ = e_{0k}, \quad k = 1,2. \tag{2.28}$$

As a consequence of (2.28) one has for the even generators

$$(E_{ij})^+ = E_{ji}, \tag{2.29}$$

i.e., the representations of the real form U(2) have to be anti-Hermitian. As is well known all such irreducible representations are finite dimensional.

Let V be a physical state space of the (internal) oscillator. For physical reasons we assume that the energy spectrum is bounded from below. Let $|E_0\rangle$ be a state corresponding to the ground energy E_0,

$$H|E_0\rangle = E_0|E_0\rangle. \tag{2.30}$$

Then [see (2.23)]

$$Hl_{k0}|E_0\rangle = (E_0 - \omega\hbar)e_{k0}|E_0\rangle \tag{2.31}$$

and, therefore,

$$e_{k0}|E_0\rangle = 0, \quad k = 1,2. \tag{2.32}$$

In order to proceed further we observe that the representations of sl(1,2) for which (2.28) holds are star representations.[27] An important feature of the star representations is that they are always completely reducible. Let V be an irreducible sl(1,2) representation space and U be the algebra of all polynomials of the sl(1,2) generators. Then the irreducibility of V implies that for any $|x\rangle \in V$ and in particular for $|E_0\rangle \in V$,

$$U|E_0\rangle = V. \tag{2.33}$$

Using the structure relations (2.20), one can represent any element $u \in U$ as a linear combination of ordered monomials of the sl(1,2) generators (Poincaré–Birkhoff–Witt theorem[23])

$$(e_{01})^{\theta_1}(e_{02})^{\theta_2}(E_{11})^{n_{11}}(E_{22})^{n_{22}}(E_{12})^{n_{12}}(E_{21})^{n_{21}}(e_{10})^{\eta_1}(e_{20})^{\eta_2}. \tag{2.34}$$

Since $(e_{0k})^2 = \frac{1}{2}\{e_{0k}, e_{0k}\} = 0$ and $(e_{k0})^2 = 0$, in (2.34), $\theta_1, \theta_2, \eta_1, \eta_2 = 0,1$, whereas η_{ij} are any non-negative integers, $n_{ij} \in \mathbf{Z}_+$. In view of (2.32) from (2.34) one concludes that

$$V = \sum_{\theta_1,\theta_2 = 0,1} (e_{01})^{\theta_1}(e_{02})^{\theta_2}\mathscr{P}|E_0\rangle, \tag{2.35}$$

where $\mathscr{P}$ denotes all polynomials of the even generators. Clearly, $\mathscr{P}|E_0\rangle = V_0$ is an irreducible finite-dimensional representation space for gl(2). Therefore, the space

$$V = V_0 + e_{01}V_0 + e_{02}V_0 + e_{01}e_{02}V_0 \tag{2.36}$$

is also finite dimensional. We came to an important conclusion: the state space V of the internal variables of the oscillator is a finite-dimensional irreducible representation space of the Lie superalgebra sl(1,2). Hence, in order to determine the possible physical state spaces, it suffices to consider only the finite-dimensional irreducible representations of sl(1,2) and subsequently select those for which the star condition (2.28) holds. All finite-dimensional irreducible representations of sl(1,2) are known. They have been constructed in Ref. 28. In the next subsection we write them down in a Gel'fand–Zetlin basis.[29]

C. Finite-dimensional irreducible representations of sl(1,2)[29]

Everywhere in the rest of the paper by a representation we understand a finite-dimensional representation. The irreducible representations (IR's) of sl(1,2) are labeled with all complex pairs $[m_{13}, m_{23}]$ such that $m_{13} - m_{23}$ are non-negative integers. An orthonormed basis in the sl(1,2) representation space [= sl(1,2) module] $W([m_{13}, m_{23}])$ is given with all possible Gel'fand–Zetlin patterns, i.e., with all possible complex tables

$$\left| \begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix} \right\rangle \tag{2.37}$$

such that ($\mathbf{Z}_+$ = all non-negative integers)

$$\begin{aligned} &(1)\ m_{13} - m_{23},\ m_{12} - m_{11},\ m_{11} - m_{22} \in \mathbf{Z}_+; \\ &(2)\ m_{i3} - m_{i2} = 0,1, \quad i = 1,2; \\ &(3)\ \text{if } m_{k3} = k - 1, \text{ then } m_{i3} = m_{i2}, \quad i \neq k = 1,2. \end{aligned} \tag{2.38}$$

Let

$$l_{ij} = m_{ij} - i, \quad \xi_i = m_{i3} - m_{i2}, \quad i = 1,2, \quad \xi = \xi_1 + \xi_2,$$

$$\theta(x) = \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{for } x < 0. \end{cases} \tag{2.39}$$

The transformation of the Gel'fand–Zetlin basis (GZ basis) under the action of the generators reads ($k = 1,2$)

$$e_{0k}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix}\right\rangle = \sum_{i=1}^{2} (-1)^{\xi_i\xi+(i-1)(k-1)}\theta(\xi-1)\theta(\xi-\xi_i-1)(l_{13}+1)^{1/2}$$

$$\times\left|\frac{\delta_{1k}l_{3-i,2}+\delta_{2k}l_{i2}-l_{11}-k-2}{l_{13}-l_{23}}\right|^{1/2}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}-\delta_{1i}+1, m_{22}-\delta_{2i}+1 \\ m_{11}+k-1 \end{matrix}\right\rangle,$$

$$e_{k0}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix}\right\rangle = \sum_{i=1}^{2} (-1)^{\xi_2+(i-1)k}\theta(\xi_i-\xi)(l_{i3}+1)^{1/2}$$

$$\times\left|\frac{\delta_{1k}l_{3-i,2}+\delta_{2k}l_{i2}-l_{11}+k-1}{l_{13}-l_{23}}\right|^{1/2}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}+\delta_{1i}-1, m_{22}+\delta_{2i}-1 \\ m_{11}-k+1 \end{matrix}\right\rangle, \tag{2.40}$$

$$(E_{11}+E_{22})\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix}\right\rangle = (l_{12}+l_{22}+3)\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix}\right\rangle,$$

$$(E_{11}-E_{22})\left|\begin{matrix} m_{13},m_{23} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle = (2l_{11}-l_{12}-l_{22}-1)\left|\begin{matrix} m_{13},m_{23} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle,$$

$$E_{12}\left|\begin{matrix} m_{13},m_{23} \\ m_{12},m_{22} \\ m_{11} \end{matrix}\right\rangle = |(l_{12}-l_{11})(l_{22}-l_{11})|^{1/2}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11}+1 \end{matrix}\right\rangle,$$

$$E_{21}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11} \end{matrix}\right\rangle = |(l_{12}-l_{11}+1)(l_{22}-l_{11}+1)|^{1/2}\left|\begin{matrix} m_{13}, m_{23} \\ m_{12}, m_{22} \\ m_{11}-1 \end{matrix}\right\rangle.$$

As in the case of the simple Lie algebras [the LS sl(1,3) is also simple], the rows labeling the GZ pattern (2.37) have well-defined meaning. The second row $[m_{12},m_{22}]$ indicates that (2.37) is a vector from the gl(2) irreducible representation subspace $V([m_{12},m_{22}])\subset W([m_{13},m_{23}])$ in the decomposition sl(1,2)$\supset$gl(2). The third row label m_{11} indicates that the pattern belongs to the gl(1) irreducible subspace of $V([m_{12},m_{22}])$ in the decomposition gl(2)$\supset$gl(1). In view of the second condition in (2.38), the space $W([m_{13},m_{23}])$ decomposes into a direct sum of no more than four gl(2) irreducible subspaces $V([m_{12},m_{22}])$. More precisely,

$$\begin{aligned} &W([m_{13},m_{23}]) \\ &\quad= V([m_{13},m_{23}])\oplus V[(m_{13}-1,m_{23}]) \\ &\qquad\oplus V([m_{13},m_{23}-1])\oplus V([m_{13}-1,m_{23}-1]). \end{aligned} \tag{2.41}$$

The terms in the above sum that do not satisfy the conditions (2.38) have to be replaced by zero. For instance, if $m_{13}=0$, then according to the third condition in (2.38), $m_{23}=m_{22}$, and, therefore, $V([0,m_{23}-1]) = V([-1,m_{23}-1]) = 0$, i.e.,

$$W([0,m_{23}]) = V([0,m_{23}])\oplus V([0,m_{23}-1]). \tag{2.42}$$

Similarly,

$$W([m_{13},1]) = V([m_{13},1])\oplus V([m_{13},0]). \tag{2.43}$$

The representations (2.42) and (2.43), corresponding to $m_{13}=0$ and $m_{23}=1$, respectively, are exceptional in a certain sense. Only for these two classes of representations the space $W([m_{13},m_{23}])$ is always a sum of no more than two gl(2) irreducible subspaces. In the Kac terminology[23] these representations are called nontypical; all other representations are typical.

D. Spectrum and eigenstates of the physical observables

Denote by (,) the scalar product in $W([m_{13},m_{23}])$ corresponding to the orthonormed GZ basis. It is a simple task to check that for any two vectors $x,y\in W([m_{13},m_{23}])]$ the relation

$$(x,e_{k0}y) = (e_{0k}x,y), \quad k=1,2, \tag{2.44}$$

holds if and only if m_{23} is real and $m_{23}\geqslant 1$. It is slightly more difficult to prove that one cannot enlarge this class of the star representations[30] by a change of the metric in $W([m_{13},m_{23}])$. Thus, the internal position and the momentum operators (and hence also the initial $\mathbf{r}_1,\mathbf{r}_2,\mathbf{p}_1,\mathbf{p}_2$) are Hermitian operators[31] only in those sl(1,2) modules for which $m_{23}\geqslant 1$. In other words, the two-particle oscillator, corresponding to the Hamiltonian (2.1), is a Wigner quantum system only in the sl(1,2) irreducible representation spaces

$$W([m_{13},m_{23}]),\quad m_{23}\geqslant 1. \tag{2.45}$$

For definiteness we call the representations corresponding to (2.45) physical representations. This class resolves into three subclasses, describing essentially different physical systems: (A) nontypical representations with $m_{13}\geqslant m_{23}=1$, (B) typical representations with $m_{13}=m_{23}>1$, and (C) typical representations with $m_{13}>m_{23}>1$.

We shall determine the spectrum of H, $\mathbf{r}^2$, $\mathbf{p}^2$, and M in each of these cases. We have

$$\hat{e}=E_{11}+E_{22},\quad \hat{l}=\tfrac{1}{2}(E_{11}-E_{22}). \tag{2.46}$$

The operator $\hat{e}$ is the central element (= the first-order Casimir operator) of gl(2). On each gl(2) irreducible subspace $V([m_{12},m_{22}])\subset W([m_{13},m_{23}])$ [as one can also see from (2.40)] it is a constant. The second operator $\hat{l}$ labels the GZ basis vectors within each $V([m_{12},m_{23}])$. From (2.21) and (2.22) one easily derives that (we put a caret over the operators; A denotes an eigenvalue of $\hat{A}$)

$$\hat{H}=\omega\hbar\hat{e},\quad \hat{\mathbf{r}}^2=(\hbar/m\omega)\hat{e},\quad \hat{\mathbf{p}}^2=m\omega\hbar\hat{e},\quad \hat{M}=2\hbar\hat{l}. \tag{2.47}$$

Clearly, all operators commute. Moreover, in the GZ basis (2.37) they are diagonal [see (2.40)]

$$\hat{e}\left|\begin{matrix} m_{13},m_{23}\\ m_{12},m_{22}\\ m_{11}\end{matrix}\right\rangle=(m_{12}+m_{22})\left|\begin{matrix} m_{13},m_{23}\\ m_{12},m_{22}\\ m_{11}\end{matrix}\right\rangle, \tag{2.48}$$

$$\hat{l}\left|\begin{matrix} m_{13},m_{23}\\ m_{12},m_{22}\\ m_{11}\end{matrix}\right\rangle=\left(m_{11}-\frac{m_{12}+m_{22}}{2}\right)\left|\begin{matrix} m_{13},m_{23}\\ m_{12},m_{22}\\ m_{11}\end{matrix}\right\rangle. \tag{2.49}$$

1. The class $m_{13}\geqslant m_{23}=1$

According to (2.42), $W([m_{13},1])$ is a direct sum of two gl(2) irreducible subspaces $V([m_{13},\theta]),\theta=0,1$. Therefore,

$$\begin{aligned}\hat{e}\left|\begin{matrix} m_{13},1\\ m_{13},\theta\\ m_{11}\end{matrix}\right\rangle&=(m_{13}+\theta)\left|\begin{matrix} m_{13},1\\ m_{13},\theta\\ m_{11}\end{matrix}\right\rangle,\\ \hat{l}\left|\begin{matrix} m_{13},1\\ m_{13},\theta\\ m_{11}\end{matrix}\right\rangle&=\left(m_{11}-\frac{m_{13}+\theta}{2}\right)\left|\begin{matrix} m_{13},1\\ m_{13},\theta\\ m_{11}\end{matrix}\right\rangle.\end{aligned} \tag{2.50}$$

The operators $\hat{H},\hat{r}=(\hat{\mathbf{r}}^2)^{1/2},\hat{p}=(\hat{\mathbf{p}}^2)^{1/2}$ have two different eigenvalues

$$H=\omega\hbar(m_{13}+\theta),\quad r=(\hbar(m_{13}+\theta)/m\omega)^{1/2},\quad p=[m\omega\hbar(m_{13}+\theta)]^{1/2},\quad \theta=0,1. \tag{2.51}$$

The angular momentum takes simultaneously $2m_{13}+1$ different integer and half-integer values

$$\begin{aligned}&-m_{13}/2\leqslant l\leqslant m_{13}/2\quad\text{in}\quad V([m_{13},0]),\\ &-(m_{13}-1)/2\leqslant l\leqslant (m_{13}-1)/2\quad\text{in}\quad V([m_{13},1]).\end{aligned} \tag{2.52}$$

The sl(1,2) module from this class, corresponding to the minimal energy $H_{\min}$ and, hence, to the minimal distance between the particles $r_{\min}$, is $W([1,1])$. The corresponding eigenvector

$$\left|\begin{matrix} 1,1\\ 1,0\\ m_{11}\end{matrix}\right\rangle$$

spans a basis in the two-dimensional gl(2) irreducible subspace $V([1,0])$. Its eigenvalues are

$$H_{\min}=\omega\hbar,\quad r_{\min}=(\hbar/m\omega)^{1/2},\quad M=-\hbar,\hbar. \tag{2.53}$$

2. The class $m_{13}=m_{23}>1$

Each sl(1,2) module $W([m_{13},m_{13}])$ is a direct sum of three gl(2) irreducible representation spaces

$$W([m_{13},m_{13}])=V([m_{13},m_{13}])\oplus V([m_{13},m_{13}-1])\oplus V([m_{13}-1,m_{13}-1]). \tag{2.54}$$

All these spaces are four dimensional $(1+2+1)$. The operators $\hat{H},\hat{r},\hat{p}$ have three different eigenvalues $(k=0,1,2)$

$$H=\omega\hbar(2m_{13}-k),\quad r=[\hbar(2m_{13}-k)/m\omega]^{1/2},\quad p=[m\omega\hbar(2m_{13}-k)]^{1/2}. \tag{2.55}$$

The eigenvalues l of $\hat{l}$ are

$$\begin{aligned}&l=0\quad\text{on}\quad V([m_{13},m_{13}])\quad\text{and}\quad V([m_{13}-1,m_{13}-1]),\\ &l=-\tfrac{1}{2},\tfrac{1}{2}\quad\text{on}\quad V([m_{13},m_{13}-1]).\end{aligned} \tag{2.56}$$

The representation space from this class, carrying a minimal energy $H_{\min}$ (resp. $r_{\min}$ and $p_{\min}$) is $W([2,2])$. The state corresponding to them is

$$\left|\begin{matrix} 2,2\\ 1,1\\ 1\end{matrix}\right\rangle$$

with eigenvalues

$$H_{\min}=2\omega\hbar,\quad r_{\min}=(2\hbar/m\omega)^{1/2},\quad M_{\min}=0. \tag{2.57}$$

3. The class $m_{13}>m_{23}>1$

In this case every $W([m_{13},m_{23}])$ is a direct sum of four nonzero subspaces $V([m_{12},m_{22}])$ [see (2.41)]. The spectrum of $\hat{H}$, $\hat{r}$, and $\hat{p}$ consists of three different points $(k=0,1,2)$

$$\begin{aligned}&H=\omega\hbar(m_{13}+m_{23}-k),\\ &r=[\hbar(m_{13}+m_{23}-k)/m\omega]^{1/2},\\ &p=[m\omega\hbar(m_{13}+m_{23}-k)]^{1/2}.\end{aligned} \tag{2.58}$$

For l one has

$$-\tfrac{1}{2}(m_{13}-m_{23}+1-k)\leqslant l\leqslant\tfrac{1}{2}(m_{13}-m_{23}+1-k),\quad k=0,1,2. \tag{2.59}$$

The spectrum of $\hat{H},\hat{r},\hat{p}$, and $\hat{l}$ is not simple. In all cases the eigenspaces of $\hat{l}$ are of dimension 1 or 2. The subspace $V([2,1])$ of $W([3,2])$ corresponds to the minimal values of H, r, and p

$$H_{\min}=3\omega\hbar,\quad r_{\min}=(3\hbar/m\omega)^{1/2},\quad p_{\min}=(3m\omega\hbar)^{1/2}. \tag{2.60}$$

The angular momentum of the minimum energy states takes values

$$-\tfrac{1}{2}(m_{13}-m_{23})\leq l\leq\tfrac{1}{2}(m_{13}-m_{23}). \qquad (2.61)$$

Comparing the expressions (2.53), (2.57), and (2.61), we conclude that the nearest allowed distance between the particles is reached, if the state space is $W([1,1])$. The subspace $V([1,0])\subset W([1,1])$ is an eigenspace of $\hat{H}$, $\hat{r}$, and $\hat{p}$, corresponding to the absolute minimum of H, r, and p

$$H_{\text{abs min}}=\omega\hbar,\quad r_{\text{abs min}}=(\hbar/m\omega)^{1/2},$$
$$p_{\text{abs min}}=(m\omega\hbar)^{1/2}. \qquad (2.62)$$

We see that in all cases the particles are bound to each other. Since $[\hat{H},\hat{r}^2]=0$, the distance r between them is preserved in time. The position, however, of any one of the constituents cannot be localized, since the coordinates do not commute. The particles are moving in a plane as the ends of a stick. The latter is rotating around the center of the mass, however; its orientation in the plane cannot be localized. The particles are smeared with a certain probability on a circumference, which centrum is the CM of the system. In the limit $\hbar\to 0$, also $r\to 0$, $H\to 0$, and $p\to 0$. Therefore, in the classical limit the composite system collapses into a point, which moves as a free classical point particle with a mass m_1+m_2 together with the center of the mass.

III. CONCLUDING REMARKS

Apart from the general axioms, which have to be satisfied by any quantum system, following Wigner and having also in mind the canonical quantum mechanics, we have demanded that (in the Heisenberg picture) the position and the momentum operators of the particles involved should satisfy the equations of motion of the classical mechanics. This is the point at which the main difference occurs between the Wigner approach, accepted in Refs. 15–17 and also in the present paper, and the other related noncanonical approaches, mentioned in the Introduction. In a case of Wigner quantum systems one can consider the particles involved as real quantum objects, discuss their masses (the equation $\bar{p}=m\dot{r}$ holds), the forces ($m\ddot{r}=-\partial u/\partial r$ also holds), etc. At the same time, the compatibility equations (2.15) put strong limitations on the possible choice of the position and the momentum operators $\mathbf{r}_i$ and $\mathbf{p}_i$, respectively. It may even happen that the Hamiltonian and the Heisenberg equations are not compatible for certain potentials, i.e., the operators $\mathbf{r}_i$ and $\mathbf{p}_i$ do not exist. On the other hand, the example we have considered allows the existence of infinitely many nonequivalent WQS's.

We have chosen the Lie superalgebra $sl(1,2)\equiv A(0,1)$ because in the Kac classification of the basic LS's[20] it is the nearest neighbor to the canonical LS of the oscillator, which is the orthosymplectic algebra $B(0,2)$. Moreover, the representations of $sl(1,2)$ are known. One may also construct Wigner quantum systems (in a two- or three-dimensional space) with internal variables generating basic LS's from the classes C and D or some of the exceptional LS's. One may go over to consider also some of the strange or the Cartan simple LS's. Not much is known, however, about the representations of these LS's and this is the main difficulty that one will encounter when dealing with them.

If one abandons the requirement of the system to be a Wigner one and goes to general quantum systems, the freedom is much larger. Even in the frame of $sl(1,2)$ [or $sl(1,3)$ in the three-dimensional space[15]] one can consider other interactions between the particles. Another approach is to assume that the coordinates and the momenta generate Lie algebras. In this way one considers relativistic models of composite systems,[5] relativistic oscillator and rotator,[6,7] etc.

Our next step will be to come back to the two-particle oscillator in a three-dimensional space and to study its properties in the frame of all finite-dimensional irreducible representations of the LS $sl(1,3)$, which are by now known.[32] As an important future problem it remains to develop a manifestly covariant approach to the ideas of Wigner, considered in I and in this paper.

ACKNOWLEDGMENTS

One of the authors (T.D.P.) would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for the kind hospitality at the International Centre for Theoretical Physics, Trieste. The authors are thankful to Professor J. Lukierski for discussions.

T. D. P. is also grateful to Professor Paolo Budinich and the International School for Advanced Studies for financial support.

[1]E. Schrödinger, Sitz. Preuss. Akad. Wiss. Phys.-Math. Kl. **24**, 418 (1930); **3**, 1 (1931).

[2]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), pp. 272–280.

[3]E. P. Wigner, Phys. Rev. **77**, 711 (1950).

[4]A. O. Barut and A. J. Bracken, in *Group Theoretical Methods in Physics, Lectures Notes in Physics*, Vol. 135, edited by K. B. Wolf (Springer, New York, 1980), p. 206; Phys. Rev. D **23**, 2454 (1981); Austr. J. Phys. **35**, 353 (1982); Phys. Rev. D **24**, 3333 (1981); A. O. Barut, A. J. Bracken, and W. D. Thacker, Lett. Math. Phys. **8**, 477 (1984); A. O. Barut and N. Zanghi, Phys. Rev. Lett. **52**, 2009 (1984); A. O. Barut and W. D. Thacker, Phys. Rev. D **31**, 1386, 2076 (1985); A. O. Barut and A. J. Bracken, J. Math. Phys. **26**, 2515 (1985).

[5]A. O. Barut, "Dynamical groups for the motion of relativistic composite systems," in *Groups, Systems and Many-Body Physics*, edited by P. Kramer and M. Dal Cin (Vieweg, Braunschweig, 1980), pp. 285–313.

[6]R. R. Aldinger, A. Bohm, P. Kielanowski, M. Loewe, P. Magnollay, N. Mukunda, W. Drechsler, and S. R. Komy, Phys. Rev. D **28**, 3020 (1983); A. Bohm, M. Loewe, L. C. Biedenharn, and H. van Dam, *ibid.* **28**, 3032 (1983).

[7]E. Celeghini and M. Tarlini, Nuovo Cimento **68B**, 133 (1982); A. Bohm, M. Loewe, and P. Magnollay, "The quantum relativistic oscillator. I. Modifying the Hamiltonian formalism of the relativistic string," University of Texas preprint DOE-ER-03992-574, 1984; "II. Non-relativistic limit and phenomenological justification," University of Texas preprint DOE-ER-03992-576, 1984; A. Bohm, M. Loewe, P. Magnollay, M. Tarlini, and R. R. Aldinger, "III. Contraction between the algebras SO(3,2) and the three-dimensional harmonic oscillator," University of Texas preprint DOE-ER-40200-583, 1985; L. S. Biedenharn, A. Bohm, M. Tarlini, H. van Dam, and N. Mukunda, "On a spinorial relativistic rotator: the transformating from quasi-Newtonian to Minkowski coordinates," University of Texas preprint DOE-ER-03992-539, 1984.

[8]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), pp. 272–280.

[9]T. S. Santhanam and A. R. Tekumalla, Found. Phys. **6**, 583 (1976); T. S. Santhanam, *ibid.* **7**, 121 (1977); in *Uncertainty Principle and Foundations of Quantum Mechanics*, edited by W. Price and S. S. Chissick (Wiley, New York, 1977), p. 277; in *Proceedings of the 1978 International Meeting of Frontiers of Physics*, edited by K. K. Phua, C. K. Chew, and Y. K. Lim (Singapore National Academy of Sciences, Singapore, 1978), p. 1167; R. Jagannathan, T. S. Santhanam, and R. Vasudevan, Int. J. Theor. Phys. **20**,

755 (1981); R. Jagannathan and T. S. Santhanam, *ibid.* **21**, 351 (1982); R. Jagannathan, *ibid.* **22**, 1105 (1983).
[10]S. Gudder and V. Naroditsky, Int. J. Theor. Phys. **20**, 619 (1981).
[11]P. Stovicek and J. Tolar, ICTP, Trieste, preprint IC/79/147, 1979.
[12]P. Ehrenfest, Z. Phys. **45**, 455 (1927).
[13]H. S. Green, Phys. Rev. **90**, 270 (1953).
[14]O. W. Greenberg and A. M. L. Messiah, Phys. Rev. B **138**, 1155 (1965).
[15]T. D. Palev, J. Math. Phys. **23**, 1778 (1982).
[16]Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **19**, 67 (1978).
[17]J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, J. Math. Phys. **19**, 2089 (1978); N. Mukunda, E. C. G. Sudarshan, J. K. Sharma, and C. L. Mehta, *ibid.* **21**, 2386 (1980); J. K. Sharma, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, *ibid.* **22**, 78 (1981).
[18]T. D. Palev, Communication JINR E2-81-419, Dubna, 1981; in *Proceedings of the 1981 Conference on Differential Geometric Methods in Theoretical Physics*, edited by G. Denardo and H. D. Doebner (World Scientific, Singapore, 1983), p. 109; Czech. J. Phys. B **32**, 680 (1982).
[19]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[20]V. G. Kac, Adv. Math. **26**, 8 (1977).
[21]M. Omote, Y. Ohnuki, and S. Kamefuchi, Prog. Theor. Phys. **56**, 1948 (1976); A. Ganchev and T. D. Palev, J. Math. Phys. **21**, 797 (1980).
[22]S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1960); C. Ryan and E. C. Q. Sudarshan, *ibid.* **47**, 207 (1963).
[23]V. G. Kac, "Representations of classical Lie superalgebras," in *Lecture Notes in Mathematics*, Vol. 676, edited by J. Polis and M. do Carmo (Springer, Berlin, 1977), p. 597.
[24]T. D. Palev, "Lie Algebraic aspects of quantum statistics," Habil. thesis, Institute of Nuclear Research and Nuclear Energy, Sofia, 1976; C. R. Acad. Bulg. Sci. **30**, 993 (1977); "Unitary quantization (*A*-quantization)," JINR preprint E17-10550, 1977; "Lie superalgebraic aspects of quantum statistics," Communication JINR E2-11929, 1978; "*A*-superquantization," Communication JINR E2-11942; C. R. Acad. Bulg. Sci. **32**, 159 (1979); Czech. J. Phys. B **29**, 91 (1979); A. B. Govorkov, Teor. Mat. Fiz. **41**, 318 (1979) [Theor. Math. Phys. 1048 (1980)]; J. Phys. A **13**, 1673 (1980); T. D. Palev, Rep. Math. Phys. **18**, 177, 129 (1980); J. Math. Phys. **21**, 2560 (1980).
[25]A. Bohm, M. Loewe, P. Magnollay, M. Tarlini, R. R. Aldinger, and P. Kielanowski, "A new relativistic generalization of the Heisenberg commutation relations," University of Texas preprint DOE-ER-03992-571, 1984.
[26]I. Saavedra and C. Utreras, Phys. Lett. B **98**, 74 (1981). See also B. Talukdar, Phys. Lett. B **119**, 155 (1982); M. Giffon and E. Predazzi, Lett. Nuovo Cimento **37**, 430 (1983).
[27]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[28]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[29]A. H. Kamupingene and T. D. Palev, "Irreducible finite-dimensional representations of the Lie superalgebra sl(1,2) in a Gel'fand–Zetlin basis," ICTP, Trieste, preprint IC/85/146, 1985.
[30]There is another class of representations, which are called star representations. It corresponds to $(e_{k0})^+ = -e_{0k}$. This condition hold iff $m_{13} < 0$. For the general definition of a star representation see Ref. 27.
[31]We use the more physical terminology. In this case it is also exact, since in a finite-dimensional space every Hermitian operator, defined everywhere in the space, is self-adjoint.
[32]T. D. Palev, J. Math. Phys. **26**, 1640 (1985); "Finite dimensional irreducible representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis," ICTP, Trieste, preprint IC/85/130, 1985.

Continuity of entropy and mutual entropy in C^*-dynamical systems

Masanori Ohya and Takashi Matsuoka
Department of Information Sciences, Science University of Tokyo, Noda City, Chiba 278, Japan

(Received 19 November 1985; accepted for publication 2 April 1986)

The lower semicontinuities of the entropy and the mutual entropy of a state for C^*-dynamical systems are proved with respect to the set of all KMS states and the set of all α-invariant states.

I. INTRODUCTION

We have several different types of entropies in quantum mechanical systems, which are fundamental tools to discuss the properties of physical systems.[1,2] The entropy of a state and the mutual entropy with respect to a state and a channel in C^*-dynamical systems were introduced and their dynamical properties were studied in Refs. 3–5. We here consider the continuity of these two C^*-entropies with respect to the set of certain states. In Sec. II, we briefly review the definitions of the C^*-entropies used throughout this paper. The lower semicontinuity of the entropy of a KMS state is studied in Sec. III, and the lower semicontinuity of the mutual entropy with respect to a KMS state and a channel is studied in Sec. IV. Finally, the lower semicontinuities of the entropy and the mutual entropy for stationary states are discussed in Sec. V under some ergodic conditions.

II. PRELIMINARIES

Let $\mathscr{A}$ be a C^*-algebra with unity I, $\mathfrak{S}$ be the set of all states on $\mathscr{A}$, and $\alpha(R)$ be a one-parameter automorphism group on $\mathscr{A}$. We call a triple $(\mathscr{A}, \mathfrak{S}, \alpha(R))$ a C^*-dynamical system and denote the GNS representation by $\{\mathscr{H}_\varphi, \pi_\varphi, x_\varphi\}$ for each state $\varphi \in \mathfrak{S}$. Further, let $\mathscr{S}$ be a weak* compact convex subset of $\mathfrak{S}$ and ex $\mathscr{S}$ be the set of all extreme points in $\mathscr{S}$.

For each state $\varphi \in \mathscr{S}$, there exists a maximal measure μ pseudosupported on ex $\mathscr{S}$ such that

$$\varphi = \int_{\mathscr{S}} \omega \, d\mu \quad \left(= \int_{(\mathrm{ex}\, \mathscr{S})} \omega \, d\mu \right). \tag{2.1}$$

This measure μ is not always unique, so we denote the set of such measures by $M_\varphi(\mathscr{S})$. Take $D_\varphi(\mathscr{S}) = \{\mu \in M_\varphi(\mathscr{S});\ \exists \{\lambda_k\} \subset R^+$ and $\{\varphi_k\} \subset \mathrm{ex}\, \mathscr{S}$ s.t. $\Sigma_k \lambda_k = 1$ and $\mu = \Sigma_k \lambda_k \delta(\varphi_k)$ with delta measure $\delta(\varphi_k)\}$, and put $H(\mu) = -\Sigma_k \lambda_k \log \lambda_k$ for $\mu \in D_\varphi(\mathscr{S})$. Then the entropy of a state φ w.r.t. $\mathscr{S}$ is defined by[3]

$$S^{\mathscr{S}}(\varphi) = \begin{cases} \inf\{H(\mu);\ \mu \in D_\varphi(\mathscr{S})\}, \\ +\infty \quad (\text{if} \ \ D_\varphi(\mathscr{S}) = \varnothing). \end{cases} \tag{2.2}$$

This entropy is an extension of von Neumann's entropy and depends on the set $\mathscr{S}$ chosen. Hence it represents the uncertainty of the state φ measured from the coordinate system $\mathscr{S}$. In particular, three cases, (1) $\mathscr{S} = \mathfrak{S}$, (2) $\mathscr{S} = I(\alpha)$ (the set of all α-invariant states), and (3) $\mathscr{S} = K(\alpha)$ (the set of all KMS states w.r.t. α), are important to analyze physical systems. Consult Ref. 3 for dynamical properties of this entropy.

The dynamical behavior of physical systems is described by the dynamical change of states. The transformation providing the state change is called a channel generally defined as follows[6,7]: Let $(\bar{\mathscr{A}}, \bar{\mathfrak{S}}, \bar{\alpha}(R))$ be another C^*-dynamical system. Then a channel Λ^* is a mapping from $\mathfrak{S}$ to $\bar{\mathfrak{S}}$ such that its dual map Λ is completely positive from $\bar{\mathscr{A}}$ to $\mathscr{A}$. If both $\mathscr{A}$ and $\bar{\mathscr{A}}$ are von Neumann algebras acting on Hilbert spaces $\mathscr{H}$ and $\bar{\mathscr{H}}$, respectively, then we require the mapping Λ to be normal. When a state $\varphi \in \mathscr{S}$ changes to another state $\bar{\varphi} = \Lambda^* \varphi \in \bar{\mathscr{S}}$ through a channel Λ^*, there exists a correlation between φ and $\bar{\varphi}$, and this correlation is expressed by the so-called compound state. The compound state w.r.t. $\mathscr{S}$ and a decomposition measure μ of (2.1) is given as

$$\Phi_\mu^{\mathscr{S}} = \int_{\mathscr{S}} \omega \otimes \Lambda^* \omega \, d\mu, \tag{2.3}$$

which is a state on the tensor product C^*-algebra $\mathscr{A} \otimes \bar{\mathscr{A}}$ (see Refs. 4 and 8). When $\mathscr{A}$ is an Abelian C^*-algebra, (2.3) is reduced to the usual compound measure for classical systems.[8]

When an initial state $\varphi \in \mathscr{S}$ goes to a final state $\bar{\varphi} = \Lambda^* \varphi$ through a channel Λ^*, it is natural for us to ask *how much information carried by φ can be transmitted to the output system*. It is the mutual entropy that represents this amount of information transmitted from φ to $\bar{\varphi}$.

In order to formulate this mutual entropy, we need the relative entropy of two states first introduced by Umegaki[9] and extended by Araki[10] and Uhlmann.[11] We here review Araki's definition of the relative entropy. Let $\mathfrak{N}$ be a σ-finite von Neumann algebra acting on a Hilbert space $\mathscr{K}$ and φ, ψ be normal states on $\mathfrak{N}$ given by $\varphi(\cdot) = \langle x, \cdot x \rangle$, $\psi(\cdot) = \langle y, \cdot y \rangle$ with $x, y \in \mathscr{K}$. Then the operator $S_{x,y}$ is defined by

$$S_{x,y}(Ay + z) = s^{\mathfrak{N}}(y) A^* x, \quad A \in \mathfrak{N}, \quad s^{\mathfrak{N}'}(y) z = 0,$$

on the domain $\mathfrak{N} y + (I - s^{\mathfrak{N}'}(y))\mathscr{K}$, where $s^{\mathfrak{N}}(y)$ is the projection from $\mathscr{K}$ to $\{\mathfrak{N}' y\}^{-1}$. For this $S_{x,y}$, the relative modular operator $\Delta_{x,y}$ is defined as $\Delta_{x,y} = (S_{x,y})^* \bar{S}_{x,y}$ and the relative entropy $S(\psi|\varphi)$ is given by

$$S(\psi|\varphi) = \begin{cases} -\langle y, (\log \Delta_{x,y}) y \rangle & (\psi \prec \varphi), \\ +\infty & (\text{otherwise}), \end{cases} \tag{2.4}$$

where $\psi \prec \varphi$ means that $\varphi(A^*A) = 0$ implies $\psi(A^*A) = 0$ for $A \in \mathfrak{N}$. If ψ and φ are states in a C^*-system, then the relative entropy $S(\psi|\varphi)$ has been defined by Uhlmann, which is, however, identical to $S(\tilde{\psi}|\tilde{\varphi})$ of Araki,[12] where $\tilde{\varphi}$ and $\tilde{\psi}$ are the canonical extensions of φ and ψ to $\pi_{\varphi+\psi}(\mathscr{A})''$.

The mutual entropy for $\varphi \in \mathscr{S}$ and Λ^* is given by[4,5]

$$I^{\mathscr{S}}(\varphi; \Lambda^*) = \lim_{\epsilon \to 0} \sup\{I_\mu^{\mathscr{S}}(\varphi; \Lambda^*);\ \mu \in F_\varphi^\epsilon(\mathscr{S})\}, \tag{2.5}$$

 0022-2488/86/082076-04$02.50

$$I_\mu^{\mathscr{S}}(\varphi;\Lambda^*) = S(\Phi_\mu^{\mathscr{S}}|\Psi), \tag{2.6}$$

where $\Psi = \varphi \otimes \Lambda^*\varphi$ and the set $F_\varphi^\epsilon(\mathscr{S})$ is

$$F_\varphi^\epsilon(\mathscr{S}) = \begin{cases} \{\mu \in D_\varphi(\mathscr{S}); \\ \quad S^{\mathscr{S}}(\varphi) \leqslant H(\mu) < S^{\mathscr{S}}(\varphi) + \epsilon < +\infty\}, \\ M_\varphi(\mathscr{S}), \text{ if } S^{\mathscr{S}}(\varphi) = +\infty. \end{cases}$$

Our quantum mechanical mutual entropy so defined is indeed an extension of the mutual information in measure theorestic formulation given by Gel'fand–Yaglom. [13]

Remarks: (1) When $S^{\mathscr{S}}(\varphi) = H(\mu)$ holds for some measure μ, $F_\varphi^\epsilon(\mathscr{S})$ is equal to the set $\{\mu \in D_\varphi(\mathscr{S}); S^{\mathscr{S}}(\varphi) = H(\mu)\}$.

(2) The mutual entropy I_μ can be expressed as

$$I_\mu^{\mathscr{S}}(\varphi:\Lambda^*) = \int_{\mathscr{S}} S(\Lambda^*\omega|\Lambda^*\varphi)d\mu, \tag{2.7}$$

for some dynamical systems.[5]

III. LOWER SEMICONTINUITY OF ENTROPY FOR KMS STATES

Let $\mathscr{S}$ be $K(\alpha)$ in this section. The next lemma is rather well known and easy to prove.

Lemma 3.1.: Let $\mathscr{A}$ be a von Neumann algebra and φ, ψ be normal states on $\mathscr{A}$. We have (1) if φ and ψ are in $\operatorname{ex} K(\alpha)$ and $\varphi \neq \psi$, then $s(\varphi) \perp s(\psi)$ [$s(\varphi)$ is the support of φ]; and (2) $\|\varphi - \psi\| = \|\,|\varphi - \psi|\,\|$.

It is known[14] that for each $\varphi \in K(\alpha)$, there exists a unique maximal measure μ of the decomposition (1.1) of φ. Therefore $D_\varphi(K(\alpha)) = \{\mu\}$ or $\varnothing$. In the sequel, let $\varphi^{(n)}$ $(n = 1,2,...)$ be a sequence of states converging to φ in norm (i.e., $\|\varphi^{(n)} - \varphi\| \to 0$) and $\mu^{(n)}, \mu$ be the decomposition measures of $\varphi^{(n)}, \varphi$, respectively.

Theorem 3.2 [Lower semicontinuity of $S^{K(\alpha)}(\varphi)$]: If $\|\varphi^{(n)} - \varphi\| \to 0$, then $S^{K(\alpha)}(\varphi) \leqslant \liminf_{n\to\infty} S^{K(\alpha)}(\varphi^{(n)})$.

Proof: The proof of this theorem consists of three cases:

(1) $\mu^{(n)} \in D_{\varphi^{(n)}}(K(\alpha))$ and $\mu \in D_\varphi(K(\alpha))$,

(2) $\mu^{(n)} \in D_{\varphi^{(n)}}(K(\alpha))$ and $\mu \notin D_\varphi(K(\alpha))$,

(3) $\mu^{(n)} \notin D_{\varphi^{(n)}}(K(\alpha))$.

Case (1): Put

$$\varphi^{(n)} = \Sigma_j \lambda_j^{(n)} \varphi_j^{(n)}$$

and

$$\varphi = \Sigma_j \lambda_j \varphi_j$$

with

$$\{\varphi_j^{(n)}\}, \{\varphi_j\} \subset \operatorname{ex} K(\alpha)$$

and

$$\lambda_1^{(n)} \geqslant \lambda_2^{(n)} \geqslant \cdots, \quad \lambda_1 \geqslant \lambda_2 \geqslant \cdots.$$

For a state

$$\omega = \sum_{n\geqslant 1} \frac{1}{2^{n+1}} \varphi^{(n)} + \frac{1}{2}\varphi,$$

Let $\{\mathscr{H}_\omega, \pi_\omega, x_\omega, U_t^\omega\}$ be the GNS representation of $\mathscr{A}$ induced by ω. Further let $\tilde{\omega}$ and $\tilde{\alpha}_t$ be the canonical extensions of ω and α_t to $\pi_\omega(\mathscr{A})''$

$$[\text{i.e., } \tilde{\omega}(Q) = \langle x_\omega, Qx_\epsilon \rangle,$$

$$\tilde{\alpha}_t^\omega = U_t^\omega Q U_t^{\omega *}, \quad Q \in \pi_\omega(\mathscr{A})''].$$

Then there exist some constants $a^{(n)}, a_i^{(n)} \geqslant 0$ for each $\varphi^{(n)} \in K(\alpha)$, $\varphi_i^{(n)} \in \operatorname{ex} K(\alpha)$ satisfying $\varphi^{(n)} \leqslant a^{(n)}\omega$, $\varphi_i^{(n)} \leqslant a_i^{(n)}\omega$. Hence there exist[14] positive operators $T^{(n)}, T_i^{(n)} \in \pi_\omega(\mathscr{A})'' \cap \pi_\omega(\mathscr{A})'$ for $\varphi^{(n)}, \varphi_i^{(n)}$ such that

$$\varphi^{(n)}(A) = \langle x_\omega, T^{(n)}\pi_\omega(A)x_\omega \rangle,$$

$\varphi_i^{(n)}(A) = \langle x_\omega, T_i^{(n)}\pi_\omega(A)x_\omega \rangle$, $A \in \mathscr{A}$, so that the canonical extensions $\tilde{\varphi}^{(n)}, \tilde{\varphi}_i^{(n)}$, to $\pi_\omega(\mathscr{A})''$ are expressed by

$$\tilde{\varphi}^{(n)}(Q) = \langle x_\omega, T^{(n)}Qx_\omega \rangle,$$

$$\tilde{\varphi}_i^{(n)}(Q) = \langle x_\omega, T_i^{(n)}Qx_\omega \rangle,$$

$$Q \in \pi_\omega(\mathscr{A})''.$$

Thus $\tilde{\varphi}^{(n)}, \tilde{\varphi}_i^{(n)}$ are normal KMS states w.r.t. $\tilde{\alpha}_t^\omega$, and $\tilde{\varphi}_i^{(n)} \in \operatorname{ex} K(\tilde{\alpha}^\omega)$ for every i. Therefore the factor decompositions of $\tilde{\varphi}^{(n)}$ and $\tilde{\varphi}$ are written as

$$\tilde{\varphi}^{(n)} = \sum_i \lambda_i^{(n)} \tilde{\varphi}_i^{(n)}, \quad \tilde{\varphi} = \sum_i \lambda_i \tilde{\varphi}_i,$$

$$\text{with } \{\tilde{\varphi}_i^{(n)}\}, \{\tilde{\varphi}_i\} \subset \operatorname{ex} K(\tilde{\alpha}^\omega),$$

which imply that the entropies $S^{K(\tilde{\alpha}^\omega)}(\tilde{\varphi}^{(n)})$ and $S^{K(\tilde{\alpha}^\omega)}(\tilde{\varphi})$ are equal to $S^{K(\alpha)}(\varphi^{(n)})$ and $S^{K(\alpha)}(\varphi)$, respectively. Furthermore we have $\|\varphi^{(n)} - \varphi\| \to 0$ iff $\|\tilde{\varphi}^{(n)} - \tilde{\varphi}\| \to 0$, which can be proved from the fact[15] that for any $Q \in \pi_\omega(\mathscr{A})$, there exists $A \in \mathscr{A}$ with $\pi_\omega(A) = Q$ and $\|A\| = \|Q\|$. Hence the proof becomes complete if we can show the inequality

$$S^{K(\tilde{\alpha}^\omega)}(\tilde{\varphi}) \leqslant \liminf_{n\to\infty} S^{K(\tilde{\alpha}^\omega)}(\tilde{\varphi}^{(n)}) \quad \text{as} \quad \|\tilde{\varphi}^{(n)} - \tilde{\varphi}\| \to 0.$$

In the sequel discussion, put $\varphi = \tilde{\varphi}$, $\varphi^{(n)} = \tilde{\varphi}^{(n)}$, and $\alpha = \tilde{\alpha}^\omega$ for notational simplicity. From the polar decomposition of states, we have

$$\varphi^{(n)} \leqslant |\varphi^{(n)} - \varphi| + \varphi, \tag{3.1}$$

$$\varphi \leqslant |\varphi^{(n)} - \varphi| + \varphi^{(n)}. \tag{3.2}$$

Put $E_n^k = \Sigma_{i=1}^k s(\varphi_i^{(n)})$ for each $k \in N$. Then we have the following inequalities by (3.1) and (1) of Lemma 3.1:

$$\sum_{i=1}^k \lambda_i^{(n)} = \varphi^{(n)}(E_n^k) \leqslant |\varphi^{(n)} - \varphi|(E_n^k) + \varphi(E_n^k)$$

$$\leqslant \|\,|\varphi^{(n)} - \varphi|\,\| + \sum_{i=1}^k \lambda_i. \tag{3.3}$$

Putting $E^k = \Sigma_{i=1}^k s(\varphi_i)$, we have the following by the same way as above:

$$\sum_{i=1}^k \lambda_i \leqslant \|\,|\varphi^{(n)} - \varphi|\,\| + \sum_{i=1}^k \lambda_i^{(n)}. \tag{3.4}$$

From (3.3), (3.4), and (2) of Lemma 3.1, it follows that

$$\left| \sum_{i=1}^k (\lambda_i^{(n)} - \lambda_i) \right| \to 0.$$

This implies $\lambda_i^{(n)} \to \lambda_i$ for every i because k is chosen arbitrarily. For a fixed k, we have

$$-\sum_{i=1}^k \lambda_i \log \lambda_i = \lim_{n\to\infty} \left(-\sum_{i=1}^k \lambda_i^{(n)} \log \lambda_i^{(n)} \right),$$

which implies

$$S^{K(\alpha)}(\varphi)=\sup_k\left\{-\sum_{i=1}^{k}\lambda_i\log\lambda_i\right\}$$
$$=\sup_k\lim_{n\to\infty}\left\{-\sum_{i=1}^{k}\lambda_i^{(n)}\log\lambda_i^{(n)}\right\}$$
$$=\sup_k\lim_{n\to\infty}\inf\left\{-\sum_{i=1}^{k}\lambda_i^{(n)}\log\lambda_i^{(n)}\right\}$$
$$\leqslant\lim_{n\to\infty}\inf\sup_k\left\{-\sum_{i=1}^{k}\lambda_i^{(n)}\log\lambda_i^{(n)}\right\}$$
$$=\lim_{n\to\infty}\inf S^{K(\alpha)}(\varphi^{(n)}).$$

Case (2): In this case the entropy $S^{K(\alpha)}(\varphi^{(n)})$ is always $+\infty$, so that the lower semicontinuity holds evidently.

Case (3): In this case the entropy $S^{K(\alpha)}(\varphi)$ is always $+\infty$. If $\lim_{n\to\infty}\inf S^{K(\alpha)}(\varphi^{(n)})$ is finite so that the lower semicontinuity of $S^{K(\alpha)}(\varphi)$ does not hold, then we shall show below that $\varphi^{(n)}$ does not converge to φ in norm. Let $\varphi^{(n)}=\Sigma_i\lambda_i\varphi_i^{(n)}$ be the discrete factor decomposition of $\varphi^{(n)}$ and $\mathscr{F}_n$ be the set $\{\varphi_i^{(n)}\}$. Since $\mu^{(n)}\in D_{\varphi^{(n)}}(K(\alpha))$ from the assumption, $\mathscr{F}_n$ is a countable set. Put $\mathscr{F}_n^{\perp}=\operatorname{ex}K(\alpha)\setminus\mathscr{F}_n$. Then $\mu^{(n)}(\mathscr{F}_n^{\perp})=0$ for each n. Suppose that $\|\varphi^{(n)}-\varphi\|\to 0$. Then $\|\mu^{(n)}-\mu\|\to 0$ because of $\|\varphi^{(n)}-\varphi\|=\|\mu^{(n)}-\mu\|$ (See Ref. 15). It follows that

$$|\mu^{(n)}(\mathscr{F}_n^{\perp})-\mu(\mathscr{F}_n^{\perp})|=|\mu(\mathscr{F}_n^{\perp})|\to 0,$$

so that for any $\epsilon>0$ there exists n_0 such that for all $n\geqslant n_0$

$$\mu(\mathscr{F}_n)>1-\epsilon. \tag{3.5}$$

Now let Q be the support of μ [i.e., $\mu(Q)=1$] and let $Q_n'=Q\setminus\mathscr{F}_n$ for $n\geqslant n_0$. Then

$$1=\mu(Q)\leqslant\mu(Q_n')+\mu(\mathscr{F}_n)\leqslant 1,$$

hence

$$\mu(Q_n')<\epsilon, \tag{3.6}$$

for $n\geqslant n_0$. Since μ is not in $D_\varphi(K(\alpha))$, there exist an uncountable subset $P\subset Q$ and $\delta>0$ such that for every $n\geqslant n_0$

$$P\subset Q_n' \quad\text{and}\quad \delta=\mu(P). \tag{3.7}$$

On the other hand, applying (3.5) and (3.6) to the above δ, we have

$$\mu(Q_n')<\delta,$$

which contradicts (3.7). Q.E.D.

IV. LOWER SEMICONTINUITY OF $I^{\mathscr{S}}(\varphi;\Lambda^*)$ FOR KMS STATES

In this section, take $\mathscr{S}=K(\alpha)$ again and assume for simplicity that both $\mathscr{A}$, $\bar{\mathscr{A}}$ are σ-finite von Neumann algebras acting on Hilbert spaces $\mathscr{H}$, $\bar{\mathscr{H}}$, respectively. Let $\varphi^{(n)}$ $(n=1,2,\ldots)$, φ be normal states on $\mathscr{A}$, where the decomposition measures are denoted by $\mu^{(n)}$, μ and let Λ^* be a (normal) channel. For the compound states $\Psi=\varphi\otimes\Lambda^*\varphi$ and

$$\Phi_\mu^{K(\alpha)}=\int_{K(\alpha)}\omega\otimes\Lambda^*\omega\,d\mu$$

introduced in Refs. 4 and 5, we have the following.

Lemma 4.1: When $\|\varphi^{(n)}-\varphi\|\to 0$, we have

(1) $\|\Psi^{(n)}-\Psi\|\to 0$

and

(2) $\|\Phi_{\mu^{(n)}}^{K(\alpha)}-\Phi_\mu^{K(\alpha)}\|\to 0.$

Proof: (1) This convergence can be observed as follows:

$$\|\Psi^{(n)}-\Psi\|$$
$$=\sup\{|(\varphi^{(n)}\otimes\Lambda^*\varphi^{(n)}-\varphi\otimes\Lambda^*\varphi)(X)|;\; X\in(\mathscr{A}\otimes\bar{\mathscr{A}})_1\},$$

where

$$(\mathscr{A}\otimes\bar{\mathscr{A}})_1\equiv\{Q\in\mathscr{A}\otimes\bar{\mathscr{A}},\;\|Q\|\leqslant 1\}.$$

Since

$$\mathscr{A}\otimes\bar{\mathscr{A}}=\{A\otimes B;\;A\in\mathscr{A},\;B\in\bar{\mathscr{A}}\}''$$

and $\|\Lambda(B)\|\leqslant\|B\|$ for any $B\in\bar{\mathscr{A}}$,

$$\|\Psi^{(n)}-\Psi\|=\sup\{|(\varphi^{(n)}\otimes\Lambda^*\varphi^{(n)}-\varphi\otimes\Lambda^*\varphi)(A\otimes B)|;\;A\in\mathscr{A}_1,\;B\in\bar{\mathscr{A}}_1\}$$
$$\leqslant\sup\{|\varphi^{(n)}(A)\varphi^{(n)}(B)-\varphi(A)\varphi(B)|;\;A,B\in\mathscr{A}_1\}$$
$$\leqslant\sup\{|\varphi^{(n)}(A)|\,|(\varphi^{(n)}-\varphi)(B)|;\;A,B\in\mathscr{A}_1\}+\sup\{|\varphi(B)|\,|(\varphi^{(n)}-\varphi)(A)|;\;A,B\in\mathscr{A}_1\}$$
$$\leqslant\|\varphi^{(n)}-\varphi\|+\|\varphi^{(n)}-\varphi\|\to 0.$$

(2) Put $\nu=\mu^{(n)}-\mu$ and let $\nu=\nu_+-\nu_-$ be the Jordan decomposition of ν. Then we obtain

$$\|\Phi_{\mu^{(n)}}^{K(\alpha)}-\Phi_\mu^{K(\alpha)}\|=\sup\left\{\left|\int_{K(\alpha)}\omega\otimes\Lambda^*\omega(A\otimes B)d\mu^{(n)}-\int_{K(\alpha)}\omega\otimes\Lambda^*\omega(A\otimes B)d\mu\right|;\;A\in\mathscr{A}_1,\;B\in\bar{\mathscr{A}}_1\right\}$$
$$\leqslant\sup\left\{\left|\int_{K(\alpha)}\omega(A)\omega(B)d\mu^{(n)}-\int_{K(\alpha)}\omega(A)\omega(B)d\mu\right|;\;A,B\in\mathscr{A}_1\right\}$$
$$=\sup\left\{\left|\int_{K(\alpha)}\omega(A)\omega(B)d\nu_+-\int_{K(\alpha)}\omega(A)\omega(B)d\nu_-\right|;\;A,B\in\mathscr{A}_1\right\}$$

$$\leqslant \sup\left\{\int_{K(\alpha)}|\omega(A)\omega(B)|d\nu_+;\ A,B,\in\mathscr{A}_1\right\}+\sup\left\{\int_{K(\alpha)}|\omega(A)\omega(B)|d\nu_-;\ A,B,\in\mathscr{A}_1\right\}$$

$$\leqslant\int_{K(\alpha)}d\nu_+ + \int_{K(\alpha)}d\nu_-$$

$$=\|\nu\|$$

$$=\|\mu^{(n)}-\mu\|$$

$$=\|\varphi^{(n)}-\varphi\|\to 0.$$

Q.E.D.

The following theorem follows from the above lemma and the lower semicontinuity of the relative entropy $S(\cdot|\cdot)$.

Theorem 4.2 [lower semicontinuity of $I^{K(\alpha)}(\varphi;\Lambda^*)$]: If $\|\varphi^{(n)}-\varphi\|\to\infty$, then

$$I^{K(\alpha)}(\varphi;\Lambda^*)\leqslant\lim_{n\to\infty}\inf I^{K(\alpha)}(\varphi^{(n)};\Lambda^*).$$

Proof: Since $\varphi^{(n)}$ and φ are KMS states, the orthogonal maximal measures $\mu^{(n)}$ for $\varphi^{(n)}$ and μ for φ are unique, respectively. Therefore

$$S^{(K(\alpha))}(\varphi^{(n)})=H(\mu^{(n)}),\quad S^{K(\alpha)}(\varphi)=H(\mu)$$

and

$$F^{\epsilon}_{\varphi^{(n)}}(K(\alpha))=\{\mu^{(n)}\},\quad F^{\epsilon}_{\varphi}(K(\alpha))=\{\mu\}.$$

Hence

$$I^{K(\alpha)}(\varphi;\Lambda^*)=\lim_{\epsilon\to 0}\sup\{I^{K(\alpha)}_{\nu}(\varphi;\Lambda^*):\ \nu\in F^{\epsilon}_{\varphi}(K(\alpha))\}$$

$$=I^{K(\alpha)}_{\mu}(\varphi;\Lambda^*)$$

$$=S(\Phi^{K(\alpha)}_{\mu}|\Psi)$$

and

$$I^{K(\alpha)}(\varphi^{(n)};\Lambda^*)=S(\Phi^{K(\alpha)}_{\mu^{(n)}}|\Psi^{(n)}).$$

From Lemma 3.1 and the lower semicontinuity of the relative entropy,[10] we have

$$I^{K(\alpha)}(\varphi;\Lambda^*)\leqslant\lim_{n\to\infty}\inf I(\varphi^{(n)};\Lambda^*).$$ Q.E.D.

V. LOWER SEMICONTINUITY OF $S^{\mathscr{S}}(\varphi)$ AND $I^{\mathscr{S}}(\varphi;\Lambda^*)$ FOR STATIONARY STATES

In this section, we consider the case $\mathscr{S}=I(\alpha)$. Let $(\mathscr{A},\alpha(R))$ be G-Abelian [i.e., $\pi_\varphi(\mathscr{A})'\cap U^{\varphi}(R)'$ is Abelian for any $\varphi\in I(\alpha)$]. Then an orthogonal maximal measure μ for each $\varphi\in I(\alpha)$ is unique,[14] and two states $\varphi,\psi,\in I(\alpha)$ are orthogonal (i.e., if ω is a positive linear functional on $\mathscr{A}$ satisfying $\omega\leqslant\varphi$ and $\omega\leqslant\psi$, then $\omega=0$) iff the supports $s(\tilde{\varphi})$ and $s(\tilde{\psi})$ of the canonical extensions $\tilde{\varphi}$ and $\tilde{\psi}$ of φ and ψ to $\pi_{\varphi+\psi}(\mathscr{A})''$ are orthogonal.[16] By using the above facts, we have the following two theorems.

Theorem 5.1: Let $(\mathscr{A},\alpha(R))$ be G-Abelian and $\varphi^{(n)}$ $(n=1,2,...)$, φ be elements of $I(\alpha)$. If $\|\varphi^{(n)}-\varphi\|\to 0$, then

$$S^{I(\alpha)}(\varphi)\leqslant\lim_{n\to\infty}\inf S^{I(\alpha)}(\varphi^{(n)}).$$

Proof: We will show that the supports $s(\varphi)$, $s(\psi)$ of φ, $\psi\in\operatorname{ex}I(\alpha)$ $(\varphi=\psi)$ are orthogonal. Take $\omega=\lambda\varphi+(1-\lambda)\psi$ for any $\varphi,\psi\in\operatorname{ex}I(\alpha)$, where $1\geqslant\lambda\geqslant 0$. Then ω is an α-invariant state and $\omega=\lambda\varphi+(1-\lambda)\psi$ is an ergodic decomposition of ω. Moreover this decomposition is unique and orthogonal because $(\mathscr{A},\alpha(R))$ is G-Abelian. Hence the supports $s(\varphi)$ and $s(\psi)$ are orthogonal, and this theorem can be proved by the same way as the case of KMS states. Q.E.D.

Let $\mathscr{A}$, $\bar{\mathscr{A}}$ be σ-finite von Neumann algebras and $\varphi^{(n)}$ $(n=1,2,...)$, φ be normal states on $\mathscr{A}$.

Theorem 4.2: Let $(\mathscr{A},\alpha(R))$ and $(\bar{A},\alpha(R))$ be G-Abelian. If $\|\varphi^{(n)}-\varphi\|\to 0$, then

$$I^{I(\alpha)}(\varphi;\Lambda^*)\geqslant\lim_{n\to\infty}\inf I^{I(\alpha)}(\varphi^{(n)};\Lambda^*).$$

Proof: As shown above, the supports $s(\varphi)$ and $s(\psi)$ of φ, $\psi\in\operatorname{ex}I(\alpha)$ $(\varphi\neq\psi)$ are orthogonal. This theorem can also be proved by the same way as the case of KMS states. Q.E.D.

The definitions (3.1) and (2.5) of our entropies might be possible to extend to arbitrary convex and compact subsets in locally convex spaces along the line of the Refs. 17 and 18.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Professor H. Umegaki for his critical reading and comments. They thank the referee for his useful suggestions to possible extensions of the present work.

[1]H. Umegaki and M. Ohya, *Quantum Mechanical Entropies* (Kyorisu Shuppan, 1983) (in Japanese).
[2]A. Wehrl, Rev. Mod. Phys. **50**, 21 (1978).
[3]M. Ohya, J. Math. Anal. Appl. **100**, 222 (1984).
[4]M. Ohya, IEEE Inform. Theory **29**, 770 (1983).
[5]M. Ohya, Springer Lect. Notes Math. **1136**, 397 (1985).
[6]H. Umegaki, J. Math. Anal. Appl. **25**, 41 (1969).
[7]M. Ohya, J. Math. Anal. Appl. **84**, 318 (1981).
[8]M. Ohya, Lett. Nuovo Cimento **38**, 402 (1983).
[9]H. Umegaki, Kodai Sem. Rep. **14**, 59 (1962).
[10]H. Araki, Publ. RIMS Kyoto Univ. **13**, 173 (1977).
[11]A. Uhlmann, Commun. Math. Phys. **54**, 21 (1977).
[12]F. Hiai, M. Ohya, and M. Tsukada, Pac. J. Math. **107**, 117 (1983).
[13]I. M. Gel'fand and A. M. Yaglom, Am. Math. Soc. Transl. **12**, 199 (1959).
[14]M. Takesaki, *Theory of Operator Algebras I* (Springer, New York, 1979).
[15]R. B. Israel, *Convexity in the Theory of Lattice Gases* (Princeton U. P., Princeton, NJ, 1979).
[16]F. Hiai, J. Oper. Theory **11**, 319 (1984).
[17]H. Uhlmann, Rep. Math. Phys. **1**, 147 (1970).
[18]W. Pusz and S. L. Woronowicz, Rep. Math. Phys. **8**, 159 (1975).

Green's function for motion in Coulomb-modified separable nonlocal potentials

B. Talukdar and U. Laha
Department of Physics, Visva-Bharati University, Santiniketan, 731 235, West Bengal, India
T. Sasakawa
Department of Physics, Tohoku University, Sendai 980, Japan

(Received 9 July 1985; accepted for publication 16 April 1986)

A closed form expression is derived for the outgoing wave radial Green's function $\mathcal{G}_l^{(+)}(r,r')$ for motion in the Coulomb plus rank one separable nonlocal potential with form factor $v_l(r) = 2^{-l} \times (l!)^{-1} r^l e^{-\beta_l r}$. Some possible applications of the result are discussed.

I. INTRODUCTION

At a center of mass energy $E = k^2 + i\epsilon$ the outgoing wave Coulomb Green's function $G_{Cl}^{(+)}(r,r')$ and the Green's function $\mathcal{G}_l^{(+)}(r,r')$ for motion in the Coulomb-distorted rank-one separable nonlocal potential satisfy the inhomogeneous differential equations

$$\left[-\frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} + \frac{2\eta k}{r} - k^2\right] G_{Cl}^{(+)}(r,r') = -\delta(r-r') \tag{1}$$

and

$$\left[-\frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} + \frac{2\eta k}{r} - k^2\right] \mathcal{G}_l^{(+)}(r,r') - \lambda_l v_l(r) \int_0^\infty v_l(r'') \mathcal{G}_l^{(+)}(r,r'') dr'' = -\delta(r-r'). \tag{2}$$

Here η is the Sommerfeld parameter. For $k\eta < 0$ we have attraction and for $k\eta > 0$, repulsion. The quantities $-\lambda_l$ and $v_l(r)$ stand for the state-dependent coupling constant and form factor of the nonlocal potential. The solution of Eq. (1) is known in the literature[1] and $G_{Cl}^{(+)}(r,r')$ is expressed in terms of regular and irregular confluent hypergeometric functions. In the present paper we are concerned with the solution of Eq. (2). We derive a closed form expression for $\mathcal{G}_l^{(+)}(r,r')$ when[2]

$$v_l(r) = 2^{-l}(l!)^{-1} r^l e^{-\beta_l r} \tag{3}$$

and examine the usefulness of the result in the study of quantum mechanical scattering by additive interactions.

In Sec. II we convert Eq. (2) into an integral equation and obtain a solution of the latter. The result for $\mathcal{G}_l^{(+)}(r,r')$ comes out in terms of certain nontrivial integrals involving $G_{Cl}^{(+)}(r,r')$. We devote Sec. III to develop a differential equation method for evaluating them. In Sec. IV we look for some applications of the expression for $\mathcal{G}_l^{(+)}(r,r')$ with particular emphasis on the outgoing wave solution, scattering phase shifts, etc. for scattering by the Coulomb-distorted separable nuclear potentials. Finally, we present some concluding remarks in Sec. V.

II. SOLUTION OF EQ. (2)

Equation (2) can be solved by converting it into an integral equation. From Eqs. (1) and (2) we write

$$\mathcal{G}_l^{(+)}(r,r') = G_{Cl}^{(+)}(r,r') + \lambda_l \int_0^\infty G_{Cl}^{(+)}(r,r'') v_l(r'') dr'' \times \int_0^\infty \mathcal{G}_l^{(+)}(r''',r') v_l(r''') dr'''. \tag{4}$$

Since the kernel in Eq. (4) is degenerate it can be solved easily to get

$$\mathcal{G}_l^{(+)}(r,r') = G_{Cl}^{(+)}(r,r') + \frac{\lambda_l \int_0^\infty dr''\, v_l(r'') G_{Cl}^{(+)}(r,r'') \int_0^\infty dr\, v_l(r) G_{Cl}^{(+)}(r,r')}{[1 - \lambda_l \int_0^\infty \int_0^\infty dr\, dr'' v_l(r) v_l(r'') G_{Cl}^{(+)}(r,r'')]}. \tag{5}$$

Clearly, the quantities to be evaluated in Eq. (5) involve the single and double transforms of $\mathcal{G}_{Cl}^{(+)}(r,r')$ by v_l's.

The Coulomb Green's function is given by

$$G_{Cl}^{(+)}(r,r'') = i(-1)^l (2k)^{2l+1} (rr'')^{l+1} e^{ik(r+r'')} \frac{\Gamma(l+1+i\eta)}{(2l+1)!} \times \Phi(l+1+i\eta, 2l+2; -2ikr_<) \times \Psi(l+1+i\eta, 2l+2; -2ikr_>), \tag{6}$$

where $r_>$ and $r_<$ are the larger and smaller values of r and r''. Here Φ and Ψ stand for the regular and irregular confluent hypergeometric functions. For our future use we shall discuss very soon some important properties of Φ and Ψ functions. Meanwhile, we note that certain indefinite integrals are implied in

$$\int_0^\infty dr''\, v_l(r'') G_{Cl}^{(+)}(r,r'')$$

and

$$\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r) v_l(r'') G_{Cl}^{(+)}(r,r'').$$

 0022-2488/86/082080-07$02.50

Buchholz[3] has considered these indefinite integrals in some detail. Use of his results and their concomitant reduction, however, do not reduce

$$\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$$

and

$$\int_0^\infty \int_0^\infty dr\, dr'' v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r'')$$

to compact analytical forms.[4] van Haeringen[5] has found that the Sturmian representation of $G_{Cl}^{(+)}(r,r')$ can be used to a good advantage to write

$$\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r'')$$

in terms of Gaussian hypergeometric functions. Although the result for the double transform is remarkably simple and elegant, a similar approach does not appear to work for

$$\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'').$$

In the following, to evaluate

$$\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$$

and

$$\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r''),$$

we take recourse to a method, different from those described in Refs. 4 and 5, and see that both results can be related to particular solutions of certain differential equations well known in the theories of mathematical physics.

III. RESULTS FOR $\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$ AND $\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r'')$

To derive closed form results for these integrals we begin by introducing two lemmas.

Lemma 1: Let the functions $F_l(r,r'')$ and $\mathscr{F}_l(r,r'')$ be related to $G_{Cl}^{(+)}(r,r'')$ by

$$F_l(r,r'') = r''^l r^{-l-1}e^{-ikr}G_{Cl}^{(+)}(r,r'') \tag{7}$$

and

$$\mathscr{F}_l(r,r'') = (rr'')^l G_{Cl}^{(+)}(r,r''). \tag{8}$$

Then the single Laplace transform $\mathscr{L}[F_l(r,r'');\beta_l] = \overline{F}_l(r,\beta_l)$ satisfies a nonhomogeneous confluent hypergeometric equation while the double Laplace transform

$$\begin{aligned}\mathscr{L}_2[\mathscr{F}_l(r,r'');\beta_l;\beta_l'] &= \{\mathscr{L}[\mathscr{L}\{\mathscr{F}_l(r,r'');r\to\beta_l\};r''\to\beta_l']\}\\ &= \overline{\mathscr{F}}_l(\beta_l,\beta_l')\end{aligned}$$

satisfies the differential equation for the Gaussian hypergeometric function.

Proof: From Eqs. (1) and (7) we have

$$\begin{aligned}&\left\{r\frac{\partial^2}{\partial r^2} + [(2l+2)+2ikr]\frac{\partial}{\partial r}\right.\\ &\quad\left. + [2ik(l+1) - 2\eta k]\right\}F_l(r,r'')\\ &\quad = (r''/r)^l e^{-ikr}\delta(r-r'').\end{aligned} \tag{9}$$

Taking the single Laplace transform $(r''\to\beta_l)$ of Eq. (9) and substituting

$$Z = -2ikr, \tag{10}$$

we get

$$\begin{aligned}&\left\{Z\frac{d^2}{dZ^2} + [(2l+2)-Z]\frac{d}{dZ} - (l+1+i\eta)\right\}\overline{F}_l(Z,\beta_l)\\ &\quad = -\frac{e^{\rho Z}}{2ik},\end{aligned} \tag{11}$$

where

$$\rho = (\beta_l + ik)/2ik. \tag{12}$$

From Eqs. (1) and (8) we obtain

$$\left[\frac{\partial^2}{\partial r^2} - \frac{2l}{r}\frac{\partial}{\partial r} - \frac{2\eta k}{r} + k^2\right]\mathscr{F}_l(r,r'') = 2(rr'')^l\delta(r-r''). \tag{13}$$

We now take the double Laplace transform of Eq. (13), substitute

$$\overline{\mathscr{F}}_l(\beta_l,\beta_l') = (2l+1)!\tilde{\mathscr{F}}_l(\beta_l,\beta_l'), \tag{14}$$

$$\begin{aligned}\tilde{\mathscr{F}}_l(\beta_l,\beta_l') &= (-1)^l[2\eta^2k(\beta_l+\beta_l')]^{-2l-2}\\ &\quad\times(\xi-1)f_l(\xi),\end{aligned} \tag{15}$$

with

$$\xi = (\beta_l' + ik)(\beta_l + ik)/(\beta_l' - ik)(\beta_l - ik) \tag{16}$$

and differentiate the resulting equation with respect to ξ to get

$$\begin{aligned}&\left[\xi(1-\xi)\frac{d^2}{d\xi^2} + \{(l+2+i\eta)\right.\\ &\quad\left. - (2-l+i\eta)\xi\}\frac{d}{d\xi} + (l-i\eta)\right]f_l(\xi) = 0.\end{aligned} \tag{17}$$

Equations (11) and (17) prove the lemma.

Lemma 2: The solutions of the differential equations for $\overline{F}_l(r,\beta_l)$ and $\overline{\mathscr{F}}_l(\beta_l,\beta_l')$ can be related to

$$\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'') \tag{18}$$

and

$$\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r''). \tag{19}$$

Proof: From Eqs. (3), (7), (8), (18), and (19) we have

$$\begin{aligned}&\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')\\ &\quad = 2^{-l}(l!)^{-1}r^{l+1}e^{ikr}\overline{F}_l(r,\beta_l)\end{aligned} \tag{20}$$

and

$$\begin{aligned}&\int_0^\infty \int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r'')\\ &\quad = 2^{-2l}(l!)^{-2}\,\overline{\mathscr{F}}_l(\beta_l,\beta_l').\end{aligned} \tag{21}$$

TABLE I. Phase shifts $\delta_Y(k)$ and $\delta_{CY}(k)$ for Yamaguchi and Coulomb plus Yamaguchi potentials with parameters given in the text.

	Phase shifts in degrees	
E_{lab} in MeV	$\delta_Y(k)$	$\delta_{CY}(k)$
5	58.8599	52.9830
10	55.2433	52.8467
15	51.6416	50.3942
20	48.4509	47.7474
25	45.6413	45.2420
30	43.1509	42.9388
35	40.9255	40.8360
40	38.9218	38.9159

Equations (20) and (21) prove the lemma.

To solve Eq. (11) we note that two independent solutions of the associated homogeneous equation are given by

$$\Phi(a,c;Z)=\frac{\Gamma(c)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(c+n)}\frac{Z^n}{n!} \tag{22}$$

and

$$\overline{\Phi}(a,c;Z)=Z^{1-c}\Phi(a-c+1,2-c;Z), \tag{23}$$

with

$$a=l+1+i\eta,\quad c=2l+2. \tag{24}$$

Note that for $c=2l+2$, Eq. (23) is not an acceptable solution. However, $\overline{\Phi}(a,c;Z)$ tends towards a solution[6] when c approaches $2l+2$. In our subsequent discussion we always mean that limit. This is no loss of generalization. See, for example, the treatment of the Coulomb field by Newton.[1] Another solution of Eq. (11), defined within the framework of the same limiting procedure, is

$$\Psi(a,c;Z)=\frac{\Gamma(1-c)}{\Gamma(a-c+1)}\Phi(a,c;Z)+\frac{\Gamma(c-1)}{\Gamma(a)}\overline{\Phi}(a,c;Z). \tag{25}$$

Given $\Phi(a,c;Z)$ and $\overline{\Phi}(a,c;Z)$, we have obtained a particular solution of Eq. (11) in the form[7]

$$[\overline{F}_l(Z,\beta_l)]_P=-\frac{1}{2ik}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta,2l+2;Z)\frac{\rho^n}{n!}, \tag{26}$$

with

$$\begin{aligned}\theta_\sigma(a,c;Z)&=\frac{1}{(c-1)}\left[\Phi(a,c;Z)\int^Z e^{-Z'}Z'^{\sigma+c-2}\overline{\Phi}(a,c;Z')dZ'-\overline{\Phi}(a,c;Z)\int^Z e^{-Z'}Z'^{\sigma+c-2}\Phi(a,c;Z')dZ'\right]\\&=[Z^\sigma/\sigma(\sigma+c-1)]\,{}_2F_2(1,c+a;\sigma+1,\sigma+c;Z).\end{aligned} \tag{27}$$

The complete primitive is

$$\begin{aligned}\overline{F}_l(r,\beta_l)&=A\Phi(l+1+i\eta,2l+2;-2ikr)+B\overline{\Phi}(l+1+i\eta,2l+2;-2ikr)\\&\quad-\frac{1}{2ik}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta,2l+2;-2ikr)\frac{\rho^n}{n!},\end{aligned} \tag{28}$$

where A and B are arbitrary constants. To determine A and B we proceed as follows.

Substituting Eq. (6) in Eq. (7) and taking the Laplace transform we get

$$\begin{aligned}\overline{F}_l(r,\beta_l)&=i(-1)^l(2k)^{2l+1}\frac{\Gamma(l+1+i\eta)}{(2l+1)!}\Big[\Psi(l+1+i\eta,2l+2;-2ikr)\\&\quad\times\int_0^r r''^{2l+1}e^{-(\beta_l-ik)}\Phi(l+1+i\eta,2l+2;-2ikr'')dr''\\&\quad+\Phi(l+1+i\eta,2l+2;-2ikr)\int_r^\infty r''^{2l+1}e^{-(\beta_l-ik)r''}\Psi(l+1+i\eta,2l+2;-2ikr'')dr''\Big].\end{aligned} \tag{29}$$

Comparing the values of $\overline{F}_l(r,\beta_l)$ from Eqs. (28) and (29) for $r=0$ and $r=\infty$ we obtain $B=0$ and

$$A=i(-1)^l(2k)^{2l+1}/[(l+1+i\eta)(\beta_l-ik)^{2l+2}]\,{}_2F_1(l+1+i\eta,2l+2;l+2+i\eta;(\beta_l+ik)/(\beta_l-ik)). \tag{30}$$

Thus from Eqs. (20), (28), and (30) we have

$$\begin{aligned}&\int_0^\infty dr''\,v_l(r'')G_{Cl}^{(+)}(r,r'')\\&\quad=[i(-1)^l2^{-l}(l!)^{-1}(2k)^{2l+1}/(l+1+i\eta)(\beta_l-ik)^{2l+2}]\\&\quad\quad\times{}_2F_1(l+1+i\eta,2l+2;l+2+i\eta;(\beta_l+ik)/(\beta_l-ik))r^{l+1}e^{ikr}\\&\quad\quad\times\Phi(l+1+i\eta,2l+2;-2ikr)-[2^{-l}(l!)^{-1}/2ik]r^{l+1}e^{ikr}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta,2l+2;-2ikr)\frac{\rho^n}{n!}.\end{aligned} \tag{31}$$

We note that Maleki and Macek[8] have derived the result for $\int_0^\infty dr''\,v_l(r'')G_{Cl}^{(+)}(r,r'')$ in a different way. In our notations their result is given by

$$\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'') = -2^{-l}(l!)^{-1}[r^{l+1}e^{ikr}/(\beta_l - ik)]\sum_{n=0}^{\infty}(l+1+i\eta+n)^{-1}\frac{(-2ikr\rho)^n}{n!}$$
$$\times\, {}_2F_1(1, i\eta - l; l+2+i\eta+n; (\beta_l+ik)/(\beta_l-ik)). \tag{32}$$

Equivalence of Eqs. (31) and (32) is not immediately clear. In Appendix A we show that by straightforward algebraic manipulations one can go from Eq. (32) to Eq. (31). Two of us have given the solution of Eq. (17) elsewhere.[9] Using this solution in Eq. (21) we get

$$\int_0^\infty\int_0^\infty dr\, dr''\, v_l(r)v_l(r'')G_{Cl}^{(+)}(r,r'')$$
$$= [2^{-2l}(l!)^{-1}(2l+1)(\beta_l+\beta_l')^{-2l-1}/(l+1+i\eta)(\beta_l'-ik)(\beta_l-ik)]$$
$$\times {}_2F_1(1, i\eta - l; l+2+i\eta; (\beta_l'+ik)(\beta_l+ik)/(\beta_l'-ik)(\beta_l-ik)). \tag{33}$$

This represents the result of van Haeringen[5] obtained by using the Sturmian representation of $G_{Cl}^{(+)}(r,r')$. Equations (5), (6), (31), and (33) taken together give our desired solution for Eq. (2).

IV. APPLICATIONS OF THE CLOSED FORM EXPRESSIONS FOR $\mathscr{G}_l^{(+)}(r,r')$

The outgoing wave Green's function for motion in a potential $V(r)$ is given by

$$\mathscr{G}_l^{(+)}(r,r') = -k^{-1}e^{-il\pi/2}\psi_l^{(+)}(k,r_<)f_l(k,r_>), \tag{34}$$

where $\psi_l^{(+)}$ and f_l stand for the physical and Jost solutions for $V(r)$. As r tends to infinity,[1]

$$\mathscr{G}_l^{(+)}(r,r') \sim -k^{-1}e^{-il\pi/2}\psi_l^{(+)}(k,r')f_l^{(\text{as})}(k,r). \tag{35}$$

The asymptotic value $f_l^{(\text{as})}(k,r)$ is prescribed both for short-range and Coulomb potentials. In particular, for the Coulomb and Coulomb-like potentials,[10]

$$f_l^{(\text{as})}(k,r) \sim e^{i(kr - \eta \ln 2kr)}. \tag{36}$$

In view of Eqs. (35) and (36), our closed form expression for $\mathscr{G}_l^{(+)}(r,r')$ derived in Sec. III yields the physical solution $\psi_l^{(+)}(k,r)$ for scattering on Coulomb plus the nonlocal potential under consideration and we have

$$\psi_l^{(+)}(k,r) = \psi_{Cl}^{(+)}(k,r) - \lambda_l[ik^l 2^{-l-1}\Gamma(l+1+i\eta)/[\Gamma(l+1)]^2(\beta_l^2+k^2)^{l+1}D_l^{(+)}(k)](\beta_l - ik/\beta_l+ik)^{i\eta}$$
$$\times e^{-\pi\eta/2}r^{l+1}e^{ikr}\Big[\{(2ik)^{2l+2}/(l+1+i\eta)\,(\beta_l-ik)^{2l+2}\}$$
$$\times {}_2F_1(l+1+i\eta, 2l+2; l+2+i\eta; \beta_l+ik/\beta_l-ik)\Phi(l+1+i\eta, 2l+2; -2ikr)$$
$$-\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta, 2l+2; -2ikr)\frac{\rho^n}{n!}\Big], \tag{37}$$

with the Coulomb physical wave function written as

$$\psi_{Cl}^{(+)}(k,r) = [(2k)^{l+1}e^{-\pi\eta/2}\Gamma(l+1+i\eta)/2(2l+1)!]r^{l+1}e^{ikr}\Phi(l+1+i\eta, 2l+2; -2ikr). \tag{38}$$

In writing Eq. (38) we had to use

$$\lim_{r\to\infty}\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$$
$$= -[(-i)^l k^l\Gamma(l+1+i\eta)/l!(\beta_l^2+k^2)^{l+1}](\beta_l-ik/\beta_l+ik)^{i\eta}e^{-\pi\eta/2}e^{i(kr-\eta\ln 2kr)}. \tag{39}$$

Derivation of Eq. (39) is rather tricky. In Appendix B we deal with this and reproduce Eq. (39) from our result in Eq. (31). Unfortunately, we were unable to obtain the asymptotic value of $\int_0^\infty dr''\, v_l(r'')\, G_{Cl}^{(+)}(r,r'')$ straightaway from the result of Maleki and Macek.[8] The quantity $D_l^{(+)}(k)$ stands for the Fredholm determinant associated with $\psi_l^{(+)}(k,r)$ and is given by

$$D_l^{(+)}(k) = 1 - \lambda_l[2^{-2l}(l!)^{-1}(2l+1)(\beta_l+\beta_l')^{-2l-1}/(l+1+i\eta)(\beta_l'-ik)(\beta_l-ik)]$$
$$\times {}_2F_1(1, i\eta-l; l+2+i\eta; (\beta_l'+ik)(\beta_l+ik)/(\beta_l'-ik)(\beta_l-ik)). \tag{40}$$

For a local potential the Fredholm determinant $D_l^{(+)}(k)$ is equal to the Jost function $f_l(k)$. For a nonlocal or combination of local and nonlocal potentials $D_l^{(+)}(k)$ and $f_l(k)$ are not identically equal; rather, they are related[11] by

$$f_l(k) = D_l^{(+)}(k)/D_l(k), \tag{41}$$

where the Fredholm determinant $D_l(k)$ associated with the regular solution is always a real[1] quantity. Naturally, the phase of $f_l(k)$ is equal to the phase of $D_l^{(+)}(k)$. Further, the phase of the Jost function is the negative of the scattering phase shift $\delta_l(k)$. Therefore, Eq. (40) provides a convenient expression for calculating $\delta_l(k)$.

For $l=0$ the form factor in Eq. (3) coincides with that of Yamaguchi.[12] For the p–p scattering in the uncoupled $^1S_0$ channel one has[13] $\lambda_0 = 2.405\ \mathrm{fm}^{-3}$ and $\beta_0 = 1.1\ \mathrm{fm}^{-1}$. We have chosen to work in units in which $\hbar^2/2m$ is equal to unity. We take $(2k\eta)^{-1} = 28.80$ fm. This is the proton Bohr radius. As $\eta \to 0$ (also $l=0$) Eq. (40) gives the well-known expression for the Yamaguchi–Fredholm determinant.[14]

Based on Eqs. (40) and (41) we have calculated the phase shifts $\delta_Y(k)$ and $\delta_{CY}(k)$ for the pure Yamaguchi as well as for Coulomb plus Yamaguchi potentials for E_{lab} between 5 to 40 MeV in steps of 5 MeV. The results are shown in Table I. We note that the values of $\delta_Y(k)$ and $\delta_{CY}(k)$ differ significantly only at low energies and the difference practically vanishes beyond $E_{\mathrm{lab}} = 40$ MeV. This is physically understandable since the Coulomb potential is expected to play a role in nuclear scattering at relatively low energies.

V. CONCLUDING REMARKS

We have derived a closed form expression for the Green's function for motion in a Coulomb-distorted separable nuclear potential and demonstrated some of its applications. Our result refers to a specific choice for the form factors of the nucleon–nucleon interaction. The method used by us can be generalized in many ways. For example, our approach can easily be extended to deal with potentials of higher rank and restriction to symmetric form factors is not compelling. We have seen that within the framework of our approach we can treat the nonsymmetric nonlocal potential of Saito,[15] which arises in the context of the orthogonality condition model. Further, the wave function obtained by us can be used to generate the half-shell-transition matrix element for Coulomb plus separable potentials.

ACKNOWLEDGMENT

This work was supported in part by the Department of Atomic Energy, the Government of India.

APPENDIX A: EQUIVALENCE OF EQS. (31) AND (32)

We write Eq. (32) in the form

$$I(r) = \int_0^\infty dr''\, v_l(r'') G_{Cl}^{(+)}(r,r'')$$

$$= -2^{-l}(l!)^{-1}[r^{l+1}e^{ikr}/(\beta_l - ik)]\Big[(l+1+i\eta)^{-1}\,{}_2F_1(1, i\eta - l; l+2+i\eta; (\beta_l+ik)/(\beta_l-ik))$$

$$+ \sum_{n=1}^{\infty}(l+1+i\eta+n)^{-1}(-2ikr\rho)^n/n!\ {}_2F_1(1,i\eta-l; l+2+i\eta+n; (\beta_l+ik)/(\beta_l-ik))\Big], \tag{A1}$$

and transform the hypergeometric functions in the first and second terms by means of the recurrence relations

$${}_2F_1(a,b;c;Z) = (1-Z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;Z) \tag{A2}$$

and

$${}_2F_1(a,b;c;Z) = (1-Z)^{-a}\,{}_2F_1(a,c-b;c;Z/(Z-1)), \tag{A3}$$

respectively. This yields

$$I(r) = -2^{-l}(l!)^{-1}[r^{l+1}e^{ikr}/(\beta_l-ik)]\Big[(l+1+i\eta)^{-1}\Big(-\frac{2ik}{\beta_l-ik}\Big)^{2l+1}$$

$$\times\ {}_2F_1(l+1+i\eta, 2l+2; l+2+i\eta; (\beta_l+ik)/(\beta_l-ik))$$

$$-\Big(\frac{\beta_l-ik}{2ik}\Big)\sum_{n=1}^{\infty}(l+1+i\eta+n)^{-1}\frac{(-2ikr\rho)^n}{n!}\,{}_2F_1\Big(1, 2l+2+n; l+2+i\eta+n; \frac{\beta_l+ik}{2ik}\Big)\Big]. \tag{A4}$$

We now write the sum in Eq. (A4) explicitly and make an appropriate iterative use of the three-term recurrence relation[6]

$$c\,{}_2F_1(a,b;c;Z) - c\,{}_2F_1(a+1,b;c;Z) + bZ\,{}_2F_1(a+1,b+1;c+1;Z) = 0, \tag{A5}$$

to get

$$I(r) = -2^{-l}(l!)^{-1}[r^{l+1}e^{ikr}/(\beta_l - ik)]\Big\{(l+1+i\eta)^{-1}\Big(-\frac{2ik}{\beta_l - ik}\Big)^{2l+1}$$
$$\times {}_2F_1(l+1+i\eta, 2l+2; l+2+i\eta; (\beta_l+ik)/(\beta_l-ik))$$
$$-\Big(\frac{\beta_l - ik}{2ik}\Big)\Big[{}_2F_1(1,2l+2; l+2+i\eta;(\beta_l+ik)/2ik)$$
$$\times\Big\{\frac{(-2ikr)}{(2l+2)} + \frac{(l+2+i\eta)}{(2l+2)(2l+3)}\frac{(-2ikr)^2}{2!} + \frac{(l+2+i\eta)(l+3+i\eta)}{(2l+2)(2l+3)(2l+4)}\frac{(-2ikr)^3}{3!} + \cdots\Big\}$$
$$-\frac{(-2ikr)}{(2l+2)}\Big\{1 + \frac{(l+2+i\eta)}{(2l+3)}\frac{(-2ikr)}{2} + \frac{(l+2+i\eta)\,(l+3+i\eta)}{(2l+3)\,(2l+4)}\frac{(-2ikr)^2}{6} + \cdots\Big\}$$
$$-\frac{(-2ikr)^2}{2(2l+3)}\Big\{1 + \frac{(l+3+i\eta)}{(2l+4)}\frac{(-2ikr)}{3} + \frac{(l+3+i\eta)(l+4+i\eta)}{(2l+4)(2l+5)}\frac{(-2ikr)^2}{12} + \cdots\Big\}\frac{\rho}{1!}$$
$$-\frac{(-2ikr)^3}{3(2l+4)}\Big\{1 + \frac{(l+4+i\eta)}{(2l+5)}\frac{(-2ikr)}{4} + \frac{(l+4+i\eta)\,(l+5+i\eta)}{(2l+5)\,(2l+6)}\frac{(-2ikr)^2}{20} + \cdots\Big\}\frac{\rho^2}{2!} - \cdots\Big]\Big\}. \quad \text{(A6)}$$

From Eqs. (27), (A3), (A6), and

$$\Phi(a,c;Z) = \frac{\Gamma(c)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(c+n)}\frac{Z^n}{n!}, \quad \text{(A7)}$$

we obtain

$$I(r) = 2^{-l}(l!)^{-1}r^{l+1}e^{ikr}\Big\{[i(-1)^l(2k)^{2l+1}/(\beta_l - ik)^{2l+2}(l+1+i\eta)]$$
$$\times {}_2F_1(l+1+i\eta, 2l+2; l+2+i\eta; (\beta_l+ik)/(\beta_l-ik))\Phi(l+1+i\eta, 2l+2; -2ikr)$$
$$-\frac{1}{2ik}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta, 2l+2; -2ikr)\frac{\rho^n}{n!}\Big\}, \quad \text{(A8)}$$

which is the desired expression Eq. (31).

APPENDIX B: DERIVATION OF EQ. (39)

To calculate the asymptotic limit of Eq. (31) we rewrite it as

$$I(r) = \int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$$
$$= 2^{-l}(l!)^{-1}r^{l+1}e^{ikr}\Big[\{i(-1)^l(2k)^{2l+1}/(\beta_l - ik)^{2l+2}(l+1+i\eta)\}$$
$$\times {}_2F_1(l+1+i\eta, 2l+2; l+2+i\eta; (\beta_l+ik)/(\beta_l-ik))\Phi(l+1+i\eta, 2l+2; -2ikr)$$
$$-\frac{1}{2ik}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta, 2l+2; -2ikr)\frac{\rho^n}{n!}\Big]$$
$$= 2^{-l}(l!)^{-1}r^{l+1}e^{ikr}\Big[\{i(-1)^l(2k)^{2l+1}\Gamma(l+1+i\eta)/(2l+1)!\}\Phi(l+1+i\eta, 2l+2; -2ikr)$$
$$\times\int_0^\infty r'^{2l+1}e^{-(\beta_l - ik)r'}\Psi(l+1+i\eta, 2l+2; -2ikr')dr' - \frac{1}{2ik}\sum_{n=0}^{\infty}\theta_{n+1}(l+1+i\eta, 2l+2; -2ikr)\frac{\rho^n}{n!}\Big]. \quad \text{(B1)}$$

In Eq. (B1) we have used[16]

$${}_2F_1(b,S; 1+S+b-d; 1-\mu/a) = [a^S\Gamma(1+b+S-d)/\Gamma(1+S-d)\Gamma(S)]\int_0^\infty e^{-ax}x^{S-1}\Psi(b,d;\mu x)dx,$$
$$\operatorname{Re} S > 0, \quad 1+\operatorname{Re} S > \operatorname{Re} d. \quad \text{(B2)}$$

With the help of Eqs. (25) and (27) we express Eq. (31) in the form

$$I(r) = 2^{-l}(l!)^{-1}\Gamma(l+1+i\eta)r^{l+1}e^{ikr}/(2l+1)!\Big[i(-1)^l(2k)^{2l+1}$$
$$\times\Phi(l+1+i\eta, 2l+2; -2ikr)\int_0^\infty r'^{2l+1}e^{-(\beta_l-ik)r'}\Psi(l+1+i\eta, 2l+2; -2ikr')dr'$$
$$+\sum_{n=0}^{\infty}\frac{\rho^n}{n!}\Big\{\Phi(l+1+i\eta, 2l+2; -2ikr)\int^r dr'(-2ikr')^{n+2l+1}e^{2ikr'}\Psi(l+1+i\eta, 2l+2; -2ikr')$$
$$-\Psi(l+1+i\eta, 2l+2; -2ikr)\int^r dr'(-2ikr')^{n+2l+1}e^{2ikr'}\Phi(l+1+i\eta, 2l+2; -2ikr')\Big\}\Big]. \tag{B3}$$

Carrying out the sum first we get

$$I(r) = 2^{-l}(l!)^{-1}(2ik)^{2l+1}\Gamma(l+1+i\eta)r^{l+1}e^{ikr}/(2l+1)!$$
$$\times\Big[\Phi(l+1+i\eta, 2l+2; -2ikr)\int_0^\infty r'^{2l+1}e^{-(\beta_l-ik)r'}\Psi(l+1+i\eta, 2l+2; -2ikr')dr'$$
$$-\Big\{\Phi(l+1+i\eta, 2l+2; -2ikr)\int^r r'^{2l+1}e^{-(\beta_l-ik)r'}\Psi(l+1+i\eta, 2l+2; -2ikr')dr'$$
$$-\Psi(l+1+i\eta, 2l+2; -2ikr)\int^r r'^{2l+1}e^{-(\beta_l-ik)r'}\Phi(l+1+i\eta, 2l+2; -2ikr')dr'\Big\}\Big]. \tag{B4}$$

As $r\to\infty$ the first term cancels the second term in the square brackets of Eqs. (B4) to give the desired asymptotic value

$$\lim_{r\to\infty} I(r) = \lim_{r\to\infty}\int_0^\infty dr''\, v_l(r'')G_{Cl}^{(+)}(r,r'')$$
$$= -[(-ik)^l\Gamma(l+1+i\eta)/l!(\beta_l^2+k^2)^{l+1}](\beta_l - ik/\beta_l+ik)^{i\eta}e^{-\pi\eta/2}e^{i(kr-\eta\ln 2kr)}. \tag{B5}$$

In deriving Eq. (B5) we have used[1]

$$\lim_{r\to\infty} i(2kr)^{l+1}e^{ikr}\Psi(l+1+i\eta, 2l+2; -2ikr) = -(i)^l e^{-\pi\eta/2}e^{i(kr-\eta\ln 2kr)} \tag{B6}$$

and the well-known integral

$$\int_0^\infty e^{-\lambda Z}Z^\nu\Phi(a,c;pZ)dZ = [\Gamma(\nu+1)/\lambda^{\nu+1}]_2F_1(a,\nu+1;c;p/\lambda). \tag{B7}$$

[1]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1966).
[2]L. Crepinsek, C. B. Lang, H. Oberhummer, W. Plessas, and H. Zingl, Acta. Phys. Austr. **42**, 139 (1975); L. Crepinsek, H. Oberhummer, W. Plessas, and H. Zingl, *ibid.* **39**, 345 (1974).
[3]H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969).
[4]B. R. Johnson and J. O. Hirschfelder, J. Math. Phys. **20**, 2484 (1979); M. K. F. Wong and E. H. Y. Yeh, J. Math. Phys. **26**, 1701 (1985).
[5]H. van Haeringen, J. Math. Phys. **24**, 1274 (1983).
[6]A. Erdelyi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 1.
[7]B. Talukdar, D. K. Ghosh, and T. Sasakawa, J. Math. Phys. **23**, 1700 (1982).
[8]N. Maleki and J. Macek, Phys. Rev. A **21**, 1403 (1980).
[9]B. Talukdar, U. Laha, and S. R. Bhattaru, J. Phys. A: Math. Gen. **18**, L359 (1985).
[10]H. van Haeringen, J. Math. Phys. **20**, 1109 (1979).
[11]C. S. Warke and R. K. Bhaduri, Nucl. Phys. A **162**, 289 (1971).
[12]Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[13]H. van Haeringen, Nucl. Phys. A **253**, 355 (1975).
[14]B. Talukdar, U. Das, and S. Mukhopadhyaya, J. Math. Phys. **20**, 372 (1979).
[15]S. Saito, Progr. Theor. Phys. **40**, 893 (1968).
[16]B. Talukdar, D. K. Ghosh, and T. Sasakawa, J. Math. Phys. **25**, 323 (1984).

Conformally symmetric radiating spheres in general relativity

M. Esculpi
Departamento de Física Aplicada, Facultad de Ingeniería, Universidad Central de Venezuela, Caracas 1051, Venezuela

L. Herrera[a)]
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela

(Received 23 September 1985; accepted for publication 6 March 1986)

A method used to study the evolution of radiating anisotropic (principal stresses unequal) spheres is applied to the case in which the space-time (within the sphere) admits a one-parameter group of conformal motions. Two different kind of models are obtained, depending on the equation of state for the stresses. In one case the energy flux density at the boundary of the sphere (the luminosity) should be given as a function of the timelike coordinate in order to integrate the system of equations. In the other case the luminosity is inferred from the equation of state for the stresses. Both models are integrated numerically and their eventual applications to some astrophysical problems are discussed.

I. INTRODUCTION

In recent papers,[1-4] the assumption of the existence of a one-parameter group of conformal motions has been exploited, with particular emphasis in spherically symmetric distributions of matter (as an additional hypothesis, we assume the orbits of the group to be orthogonal to the four-velocity of the matter). The main results emerging from these papers may be resumed as follows.

(1) There is a link between the "stiff" equation of state ($P=\rho$) and the existence of a class of conformal motions (special conformal motions).[1]

(2) The spherically symmetric static solutions (admitting a one-parameter group of conformal motions) may be matched with the Schwarzschild (vacuum) metric, only for a restricted class of conformal motions (which excludes the homothetic motions).

(3) For the nonstatic (but nonradiating) spherically symmetric solutions, the matching with the Schwarzschild metric at the boundary of the matter may be accomplished only for anisotropic fluids.[2,3] In this case we were able to find some analytical solutions representing expanding, contracting, and oscillating distributions of matter, respectively. Both the expanding and the contracting solutions tend asymptotically ($t\rightarrow\infty$) to spheres with a surface potential equal to $\frac{1}{3}$ (see Ref. 3).

(4) It was possible to find distributions of matter with vanishing gravitational mass (solutions that may be matched with the Minkowski space-time).[4]

In view of all these results, we feel it is worth exploring further the models and the consequences arising from the assumption of the existence of a one-parameter group of conformal motions.

In the present work we shall consider radiating spherically symmetric distributions of matter such that the space-time within the sphere admits a one-parameter group of conformal motions. The solutions will be constructed using a method introduced some years ago[5] and which have been successfully employed to describe spherically symmetric radiating systems.[6-8] A brief résumé of this method as well as the conventions and the field equations are given in Sec. II.

In Sec. III, we include the specific symmetry of the problem (conformal motions). Two examples are worked out explicitly in Secs. IV and V. Finally the results are discussed and speculations about the possible applications of the models to some astrophysical problems are presented.

II. THE FIELD EQUATIONS AND CONVENTIONS

Our starting point is Bondi's approach to study the evolution of gravitating spheres[9]—the difference is that we shall consider anisotropic matter instead of perfect fluids.

Thus let us consider a nonstatic distribution of matter that is spherically symmetric: In radiation coordinates,[10]

$$dS^2 = e^{2\beta}((V/r)du^2 + 2\,du\,dr) - r^2(d\theta^2 + \sin^2\theta\,d\phi^2)\,, \tag{1}$$

where β and V are functions of u and r. Here $u\equiv x^0$ is the timelike coordinate, $r\equiv x^1$ is a null coordinate, and $\theta, \phi\equiv x^{2,3}$ are the usual angle coordinates. In these coordinates the components of the energy momentum tensor are distinguished by a bar and differentiation with respect to u and r is denoted by suffixes 0 and 1, respectively.

Thus it can be shown[5,6,9] that the Einstein equations may be written as

$$\frac{\rho + P\omega^2}{1-\omega^2} + \epsilon = \frac{r}{V} e^{-2\beta}\overline{T}_{00} = \frac{1}{4\pi r(r-2\widetilde{m})}\left(-\widetilde{m}_0 e^{-2\beta} + \frac{r-2\widetilde{m}}{r}\widetilde{m}_1\right), \tag{2}$$

$$\frac{\rho - P\omega}{1+\omega} = e^{-2\beta}\overline{T}_{01} = \frac{\widetilde{m}_1}{4\pi r^2}, \tag{3}$$

$$\frac{1-\omega}{1+\omega}(\rho + P) = \frac{V}{r}e^{-2\beta}\overline{T}_{11} = \frac{r-2\widetilde{m}}{2\pi r^2}\beta_1, \tag{4}$$

[a)] Postal address: Apartado 80793, Caracas 1080 A, Venezuela.

 0022-2488/86/082087-10$02.50

$$P_\perp = \widetilde{T}^2_2$$

$$= -\frac{\beta_{01}e^{-2\beta}}{4\pi} + \frac{1}{8\pi}\left(1 - \frac{2\widetilde{m}}{r}\right)\left(2\beta_{11} + 4\beta_1^2 - \frac{\beta_1}{r}\right)$$

$$+ \frac{3\beta_1(1 - 2\widetilde{m}_1) - \widetilde{m}_{11}}{8\pi r}, \tag{5}$$

where ρ, P,$P_\perp$, and ω are, respectively, the energy density, the radial pressure, the tangential pressure, and the radial velocity of matter, as measured by a locally Minkowski observer. Also the flux of radiation $\hat{\epsilon}$, as measured by the same observer, is related to ϵ by

$$\epsilon = \hat{\epsilon}\cdot[(1+\omega)/(1-\omega)] .$$

The field equations may be integrated outside the matter to obtain

$$\beta = 0, \quad V = r - 2\widetilde{m}(u), \quad \epsilon = -\widetilde{m}_0/4\pi r(4 - 2\widetilde{m}), \tag{6}$$

where $\widetilde{m}$ is a function of integration depending on u. Inside the matter the function $\widetilde{m}(u)$ is generalized to $\widetilde{m}(u,r)$ by putting everywhere

$$V = e^{2\beta}[r - 2\widetilde{m}(u,r)] . \tag{7}$$

Also note that the velocity of matter in the radiative coordinate is given by

$$\frac{dr}{du} = \frac{V}{r}\frac{\omega}{1-\omega}. \tag{8}$$

Next, let us define the two auxiliary functions

$$\tilde{\rho} \equiv (\rho - \omega P)/(1+\omega) , \tag{9}$$

$$\widetilde{P} \equiv (P - \omega\rho)/(1+\omega) , \tag{10}$$

hereafter referred to as the effective density and the effective pressure, respectively. Observe that from the field equations (3) and (4), we have

$$\widetilde{m} = \int_0^r 4\pi r^2 \tilde{\rho}\, dr , \tag{11}$$

$$\beta = \int_a^r \frac{2\pi r^2}{r - 2\widetilde{m}}(\tilde{\rho} + \widetilde{P})\, dr , \tag{12}$$

where $r = a(u)$ defines the boundary of the anisotropic fluid.

Now the algorithm to construct the solution goes as follows (for details see Refs. 5 and 6).

(1) Take a static interior solution of the Einstein equations for anisotropic matter with spherical symmetry and with given

$$\rho_{st} = \rho(r) , \quad P_{st} = P(r) .$$

(2) Assume that the r dependence of $\widetilde{P}$ and $\tilde{\rho}$ is the same as of P_{st} and ρ_{st}, but being careful with the boundary condition, which now reads, because of (10),

$$\widetilde{P}_a = -\omega_a \tilde{\rho}_a .$$

(From now on the suffix a indicates that the quantity is evaluated at the surface.)

(3) With the r dependence of $\tilde{\rho}$ and $\widetilde{P}$ and using (11) and (12), one gets $\widetilde{m}$ and β up to three functions of u, which will be specified below.

(4) For these three functions one has two differential equations, one of which is (8) evaluated at $r = a$ and the other is

$$[T^\mu{}_{1;\mu}]_a = 0 .$$

Another u-dependent equation can be obtained evaluating (6) at $r = a + 0$. Thus,

$$E(u) = (4\pi r^2\epsilon)|_{r=a+0} = [\widetilde{m}_0/(1 - 2\widetilde{m}/r)]_{r=a+0} .$$

Thus one has three differential equations for five unknown functions of u, which are the radius a, the velocity of the surface, and the functions $\widetilde{m}$, E and $P_\perp$ evaluated at the surface [see Eqs. (15), (17), and (21)].

(5) Given one of the functions, and specifying the equation of state relating the tangential pressure with the other dynamical variables, the system may be integrated for any particular initial data.

(6) Feeding back the result of integration in the expressions for β and $\widetilde{m}$, these two functions are completely determined.

(7) Using (2)–(5) and the equation of state for the tangential pressure, ρ, $P_\perp$, P, ω, and ϵ may be found.

In this paper we shall assume the existence of a one-parameter group of conformal motions plus the orthogonality between the four-velocity and the orbits of the group. We shall see in the next section that as a consequence of this additional symmetry it is possible, in some cases, to integrate the system of equations at the surface, without prescribing *a priori* the luminosity E.

In other cases, both the luminosity and the equations of state for the stresses should be given to integrate the system.

We would like to close this section with the following remark: In order to make the method outlined above completely consistent it is necessary to match the interior solution with the Vaidya metric at the boundary of the fluid distribution (Darmois or Lichnerowicz conditions). It is easy to check that these conditions are equivalent to the continuity of the functions β and $\widetilde{m}$ across the boundary, and to the equation[11]

$$-\beta_{0a} + (1 - 2\widetilde{m}_a/a(u))\beta_{1a} - \widetilde{m}_{1a}/2a(u) = 0 , \tag{13}$$

where, as before, we have defined the boundary by the equation

$$r = a(u) .$$

Thus we shall demand the continuity of β and $\widetilde{m}$ across the boundary [see Eqs. (11) and (12)]. On the other hand, condition (13) is completely equivalent to one of the equations at the surface [see Eq. (13′) below].

Finally, the vanishing of the radial pressure at the boundary (which is also assumed explicitly in this work) may be shown to be a direct consequence of (13), (8), (4), and the continuity of β across the boundary [Eq. (12)].

III. THE SURFACE EQUATIONS AND THE CONFORMAL MOTIONS

A. The equations at the surface

As should be clear from the previous section, the crucial point in the algorithm is the system of equations for the quantities evaluated at the surface (surface equations). Two of them are the same for any model with a spherically symmetric distribution of matter. Thus, from Eqs. (6) and (8) we get

$$\dot{a}=(1-2m/a)\omega_a/(1-\omega_a)\,, \tag{13'}$$

with $\dot{a}\equiv da/du$, and where $m\equiv\tilde{m}_a$ is the total mass.

Scaling the radius a, the total mass m, and the timelike coordinate u by the initial mass $\tilde{m}(u=0)\equiv m(0)$,

$$A\equiv a/m(0)\,,\quad M\equiv m/m(0)\,,\quad u/m(0)\to u\,,$$

and defining

$$F\equiv 1-2M/A\,,\quad \Omega\equiv 1/(1-\omega_a)\,, \tag{14}$$

we can write Eq. (13′) as

$$\dot{A}=F(\Omega-1)\,, \tag{15}$$

which is the first surface equation, and which, as we mentioned in the previous section, is implied by the junction conditions.

The second surface equation relates the total mass loss rate with the energy flux through the surface. This can be obtained by evaluating Eq. (6) for $r=a+0$ and takes the form

$$\dot{M}=-FE\,. \tag{16}$$

Or, using (14) and (15),

$$\dot{F}/F=[2E+(1-F)(\Omega-1)]/A\,. \tag{17}$$

The third equation at the surface will be obtained from the conservation equation $T^{\mu}_{1;\mu}=0$ evaluated at the surface.

Thus from

$$(T^{\mu}_{1;\mu})_a=0\,,$$

we get

$$-\left[\frac{(\dot{\tilde{\rho}}+\dot{\tilde{P}})}{(1-2\tilde{m}/r)}\right]_{0a}+\left(\frac{\partial\tilde{P}}{\partial r}\right)_a+\left\{\frac{(\tilde{\rho}+\tilde{P})}{(1-2\tilde{m}/r)}\left[4\pi r\tilde{P}+\frac{\tilde{m}}{r^2}\right]\right\}_a=\left[\frac{2}{r}(P_\perp-\tilde{P})\right]_a\,. \tag{18}$$

Or, after a lengthy and tedious calculation,

$$-\frac{\dot{\tilde{\rho}}_a}{\tilde{\rho}_a}+\frac{\dot{F}}{F}+\frac{\dot{\Omega}}{\Omega}+(\Omega-1)\left[4\pi a\tilde{\rho}_a\frac{(3\Omega-1)}{\Omega}+\Omega F\frac{\tilde{p}_{1a}}{\tilde{\rho}_a}-\frac{(3+F)}{2a}+\frac{2}{a}\frac{\Omega F}{\tilde{\rho}_a}(P_\perp-P)_a\right]+\frac{\Omega^2F\tilde{R}_a}{\tilde{\rho}_a}=0\,, \tag{19}$$

where

$$\tilde{R}_a(u)=\left\{\frac{(\tilde{\rho}+\tilde{P})}{(1-2\tilde{m}/r)}\left(4\pi r\tilde{P}+\frac{\tilde{m}}{r^2}\right)+\frac{\partial\tilde{P}}{\partial r}-\frac{2}{r}(P_\perp-P)\right\}_a\,. \tag{20}$$

If the effective density $\tilde{\rho}$ is separable, i.e., $\tilde{\rho}=f(u)h(r)$, then Eq. (19) becomes

$$\dot{F}/F+(\dot{\Omega}/\Omega)(1-F)=G(F,\Omega,A)\,, \tag{21}$$

with

$$G\equiv(F-1)(\Omega-1)\left[4\pi a\tilde{\rho}_a\frac{(3\Omega-1)}{\Omega}+\frac{\Omega F\tilde{p}_{1a}}{\tilde{\rho}_a}-\frac{(3+F)}{2a}+K(a)F+\frac{2\Omega F}{a\tilde{\rho}_a}P_{\perp a}\right]-\frac{(1-F)\Omega^2F\tilde{R}_a(u)}{\tilde{\rho}_a}\,, \tag{22}$$

where

$$K(a)\equiv\frac{d}{du}\ln\left[\frac{1}{a}\int_0^a dr\,r^2\frac{h(r)}{h(a)}\right].$$

In the next part of this section we shall complement the system of equations (15), (17), and (21) with another surface equation obtained from the additional symmetry of the problem (conformal motion).

B. The conformal motion

As was already mentioned, we shall assume that the space-time within the sphere admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\,g_{\mu\nu}=\psi(u,r)g_{\mu\nu}\,, \tag{23}$$

where the left-hand side defines the Lie derivative of the metric tensor, $\psi(u,r)$ is an arbitrary function of their arguments, and the vector field ξ^α has the general form

$$\xi^\alpha=\sigma(u,r)\delta^\alpha_u+\lambda(u,r)\delta^\alpha_r\,. \tag{24}$$

We shall further assume that

$$\xi^\alpha U_\alpha=0\,, \tag{25}$$

where U^α defines the four-velocity of matter.

Now, from condition (25) we get at once

$$\sigma=(-r/V)(1-\omega)\lambda\,, \tag{26}$$

where we have used (24) and the fact that the components of the four-velocity read

$$U^\mu\left[e^{-2\beta}\left(\frac{1-\omega}{1+\omega}\right)^{1/2}\left(1-\frac{2\tilde{m}}{r}\right)^{-1/2},\ \frac{\omega(1-2\tilde{m}/r)^{1/2}}{(1-\omega^2)^{1/2}},0,0\right] \tag{27}$$

or, in convariant components,

$$U_\mu\left[e^{2\beta}\left(\frac{1-2\tilde{m}/r}{1-\omega^2}\right)^{1/2},\left(\frac{1-\omega}{1+\omega}\right)^{1/2}\left(1-\frac{2\tilde{m}}{r}\right)^{-1/2},0,0\right]. \tag{28}$$

Next, we obtain from Eq. (23), after simple manipulations,

$$2\beta_0\sigma+(\beta_1r-\tfrac{1}{2})\psi+r\psi_1/2+\sigma_0=0\,, \tag{29}$$

$$\sigma\left(2\beta_0+\frac{V_0}{V}\right)+\psi\left(\beta_1r+\frac{V_1r}{2V}-\frac{3}{2}\right)+2\sigma_0+\frac{r^2}{V}\psi_0=0\,, \tag{30}$$

$$\sigma_1=0\,, \tag{31}$$

$$\lambda=\psi r/2\,. \tag{32}$$

Feeding back (32) into (26), we can integrate (31), to obtain

$$\psi=g(u)V/r^2(1-\omega)\quad\text{and}\quad\sigma=-g(u)/2\,, \tag{33}$$

where $g(u)$ is an arbitrary function of integration depending on u. Next, taking derivatives of (33),

$$\psi_1 = \frac{g(u)V_1}{r^2(1-\omega)} - \frac{2g(u)V}{r^2(1-\omega)} + \frac{g(u)V\omega_1}{r^2(1-\omega)^2}, \quad (34)$$

$$\psi_0 = \frac{\dot g(u)V}{r^2(1-\omega)} + \frac{g(u)V_0}{r^2(1-\omega)} + \frac{g(u)V\omega_0}{r^2(1-\omega)^2}, \quad (35)$$

and feeding back (33)–(35) into (29)–(30), we get

$$-\beta_0 g + \frac{Vg}{r^2(1-\omega)}\left(\beta_1 r - \frac{3}{2}\right) + \frac{V_1 g}{2r(1-\omega)} + \frac{gV\omega_1}{2r(1-\omega)^2} = \frac{\dot g}{2}, \quad (36)$$

$$-\beta_0 g + \frac{V_0 g}{2V}\left(\frac{1+\omega}{1-\omega}\right) + \frac{Vg}{(1-\omega)r^2}\left(\beta_1 r - \frac{3}{2}\right) + \frac{V_1 g}{2(1-\omega)r} + \frac{\omega_0 g}{(1-\omega)^2} = -\frac{\dot g\omega}{(1-\omega)}. \quad (37)$$

Finally, multiplying (36) by $2\omega/(1-\omega)$ and adding it to (37) we obtain

$$-\left(\frac{1+\omega}{1-\omega}\right)\beta_0 + \left[\frac{1+\omega}{(1-\omega)^2}\right]\frac{V}{r^2}\left(\beta_1 r - \frac{3}{2}\right) + \frac{V_1}{2r}\frac{(1+\omega)}{(1-\omega)^2} + \frac{V_0}{2V}\frac{(1+\omega)}{(1-\omega)} + \frac{V}{r}\frac{\omega_1\omega}{(1-\omega)^3} + \frac{\omega_0}{(1-\omega)^2} = 0. \quad (38)$$

Parenthetically, this last equation is equivalent to the condition

$$\mathop{L}_{\xi} U_\mu = \psi U_\mu/2, \quad (39)$$

which follows from the symmetry of the problem.

We shall now evaluate Eq. (38) at the surface $r=a(u)$. Recalling that $\beta(a,u)=0$ and using (7) we get

$$\left[\frac{(1+\omega)}{(1-\omega)^2}\right]_a \left\{2\beta_1\left(1-\frac{2\tilde m}{r}\right) - \frac{3}{2r}\left(1-\frac{2\tilde m}{r}\right) + \frac{1}{2r}(1-2\tilde m_1)\right\}_a - \left\{\frac{\tilde m_0}{(r-2\tilde m)}\left(\frac{1+\omega}{1-\omega}\right)\right\}_a + \left\{\left(1-\frac{2\tilde m}{r}\right)\frac{\omega\omega_1}{(1-\omega)^3} + \frac{\omega_0}{(1-\omega)^2}\right\}_a = 0 \quad (40)$$

or, using Eq. (13),

$$\frac{(1+\omega_a)}{(1-\omega_a)^2}\left[2\beta_{0a} - \frac{3}{2a}\left(1-\frac{2m}{a}\right) + \frac{1}{2a}\right] - \frac{1+\omega_a}{1-\omega_a}\frac{\tilde m_{0a}}{(a-2m)} + \left(1-\frac{2m}{a}\right)\frac{\omega_a\omega_{1a}}{(1-\omega_a)^3} + \frac{\omega_{0a}}{(1-\omega_a)^2} = 0. \quad (41)$$

Near the surface we may write

$$\tilde m(r,u) \simeq m + \tilde m_{1a}(r-a) + \cdots,$$

$$\omega(u,r) \simeq \omega(a) + \omega_{1a}(r-a) + \cdots,$$

then, taking derivatives and using (13′),

$$\tilde m_{0a} = \dot m - \dot a\tilde m_{1a} = \dot m - (1-2m/a)[\omega_a/(1-\omega_a)]\tilde m_{1a}, \quad (42)$$

$$\omega_{0a} = \dot\omega_a - \dot a\omega_{1a} = \dot\omega_a - (1-2m/a)[\omega_a/(1-\omega_a)]\omega_{1a}. \quad (43)$$

Feeding back (42) and (43) in to (41) and evaluating β_{1a} and $\tilde m_{1a}$ from the field equations (3) and (4), we obtain

$$(1+\omega_a)\left\{\frac{3m}{a^2} - \frac{1}{a}\right\} - \frac{(1-\omega_a^2)\dot m}{a(1-2m/a)} + \dot\omega_a = 0. \quad (44)$$

Or, using the dimensionless variables introduced above, we obtain finally

$$\tfrac{3}{2}(1-F) + A\dot\Omega/\Omega(2\Omega-1) - 1 = -E/\Omega. \quad (45)$$

Now the complete set of equations at the surface is formed by Eqs. (15), (17), (21), and (45).

Two different models (at least) may be derived from these surface equations, depending on the equation of state for the stresses. We shall develop these models in the next two sections.

IV. MODEL I

In this model we start with the static anisotropic solution found in Ref. 1. Namely

$$8\pi\rho = -\frac{3}{4}C + \left(\frac{1}{2} - \frac{H}{4}\right)\frac{1}{r^2}, \quad (46)$$

$$8\pi P = \frac{3}{4}C + \left(\frac{3H}{4} + \frac{1}{2}\right)\frac{1}{r^2}, \quad (47)$$

$$8\pi P_\perp = \frac{3C}{4} + \left(\frac{H}{4} + \frac{1}{2}\right)\frac{1}{r^2}, \quad (48)$$

with the constants C and H related to the radius a by

$$a^2 = -(2/3C)(1+3H/2),$$

with $C<0$ and $-\frac{2}{3}\leqslant H\leqslant 0$.

Next, following the method outlined above we choose

$$\tilde\rho = \frac{f(u)}{2}\left[\frac{(1+3H/2)}{a^2} + \frac{(1-H/2)}{r^2}\right], \quad (49)$$

where $f(u)$ is a function of u and the radius a is now a function of u too. And for the effective pressure, we take

$$\tilde P = \frac{f(u)}{r^2}\left[(1-\omega_a)\left(1+\frac{H}{2}\right) - \frac{(1-H/2)}{2} - \frac{(1+3H/2)}{2a^2}r^2\right], \quad (50)$$

which satisfies the boundary condition

$$\tilde P_a = -\omega_a\tilde\rho_a. \quad (51)$$

Now we can use (49) and (50) to integrate (11) and (12). We obtain

$$\tilde m = \int_0^r 4\pi r^2\tilde\rho\, dr = 2\pi f(u)\left[\left(1+\frac{3H}{2}\right)\frac{r^3}{3a^2} + \left(1-\frac{H}{2}\right)r\right]. \quad (52)$$

Evaluating $\tilde m$ at $r=a$, we obtain

$$f(u) = (3/16\pi)(1-F). \quad (53)$$

For β, we obtain

$$\beta = \frac{3(1+H/2)}{4\Omega[4-3(1-F)(1-H/2)]} \ln \left\{ \frac{4Fr^2/a^2}{[4-(1-F)(3(1-H/2)+(1+3H/2)r^2/a^2)]} \right\}, \tag{54}$$

and, taking derivatives of (54) and (52),

$$m(0)\beta_1 \equiv \bar{\beta}_1 = \frac{3(1+H/2)(1-F)}{2\Omega A\,[4-(1-F)(3(1-H/2)+(1+3H/2)r^2/a^2)](r/a)}, \tag{55}$$

$$m(0)^2\beta_{11} \equiv \bar{\beta}_{11} = \frac{3(1+H/2)(1-F)}{2\Omega A^2[4-(1-F)(3(1-H/2)+(1+3H/2)(r^2/a^2))]^2(r/a)^2} \times \left\{ 3(1-F)\left[\left(1-\frac{H}{2}\right)+\left(1+\frac{3H}{2}\right)\frac{r^2}{a^2}\right]-4 \right\}, \tag{56}$$

$$m(0)^2\beta_{01} \equiv \bar{\beta}_{01} = -\bar{\beta}_1\left[\frac{\dot{\Omega}}{\Omega}+\frac{4\dot{F}}{(1-F)[4-(1-F)(3(1-H/2)+(1+3H/2)(r^2/a^2))]}\right] - \dot{A}\left(\frac{\bar{\beta}_1}{A}+\left(\frac{r}{a}\right)\bar{\beta}_{11}\right), \tag{57}$$

$$\tilde{m}_1 = \tfrac{3}{8}(1-F)[(1+3H/2)(r/a)^2+(1-H/2)], \tag{58}$$

$$\tilde{m}_{11}m(0) \equiv \bar{\tilde{m}}_{11} = [3(1-F)(1+3H/2)/4A](r/a). \tag{59}$$

Next, using (49) and (50) it is a simple matter to obtain

$$K(a) = 2/a \tag{60}$$

and

$$m(0)G(F,\Omega,A) = \frac{-3(1-F)^2(\Omega-1)(2\Omega-1)(1+H/2)}{4\Omega A} + \frac{(1-F)F\Omega(1+3H/2)}{A(1+H/2)} - \frac{(1-F)^2(3-2\Omega)}{2A} + \frac{\Omega F\,32\pi\bar{P}_{\perp a}A}{3(1+H/2)}, \tag{61}$$

with $\bar{P}_{\perp a} \equiv P_{\perp a}m(0)^2$, and where we have assumed the equation of state

$$P_\perp - P = -\gamma(u)H/16\pi r^2 \tag{62}$$

for the stresses, which generalizes in a simple way the corresponding equation of state of the static solution (46)–(48).

From the boundary conditions it results that

$$\gamma(u) = -P_{\perp a}16\pi a^2/H, \tag{63}$$

so that

$$m(0)^2P \equiv \bar{P} = \bar{P}_\perp - P_{\perp a}a^2/r^2. \tag{64}$$

Now, using (61) in Eq. (21), we may express $P_{\perp a}$ in terms of F, E, A, and Ω.

Next for a given E we can integrate the system

$$\dot{A} = F(\Omega-1), \tag{65}$$

$$\frac{\dot{F}}{F} = \frac{2E+(1-F)(\Omega-1)}{A}, \tag{66}$$

$$\frac{\dot{\Omega}}{\Omega} = \frac{(2\Omega-1)}{A}\left(1-\frac{3}{2}(1-F)-\frac{E}{\Omega}\right). \tag{67}$$

Then the physical variables P, $P_\perp$, ρ, ω, and ϵ can be calculated for any piece of material, from

$$\bar{P}_\perp \equiv m(0)^2P_\perp = -\frac{\bar{\beta}_{01}e^{-2\beta}}{4\pi}+\frac{1}{8\pi}\left(1-\frac{2\tilde{m}}{r}\right)\left(2\bar{\beta}_{11}+4\bar{\beta}_1^2-\frac{\bar{\beta}_1}{A(r/a)}\right)+\frac{3\bar{\beta}_1(1-2\tilde{m}_1)}{8\pi(r/a)A}-\frac{\bar{\tilde{m}}_{11}}{8\pi(r/a)A}, \tag{68}$$

$$\bar{P} \equiv \bar{P}_\perp - \bar{P}_{\perp a}/(r/a)^2, \tag{69}$$

$$\bar{\rho} \equiv m(0)^2\rho = \frac{\tilde{m}_1(1+\omega)}{4\pi(r^2/a^2)A^2}+\bar{P}\omega, \tag{70}$$

$$\omega = 1-\frac{(1-2\tilde{m}/r)\bar{\beta}_1}{2\pi(r/a)A\,[\tilde{m}_1/4\pi(r^2/a^2)A^2+\bar{P}]}, \tag{71}$$

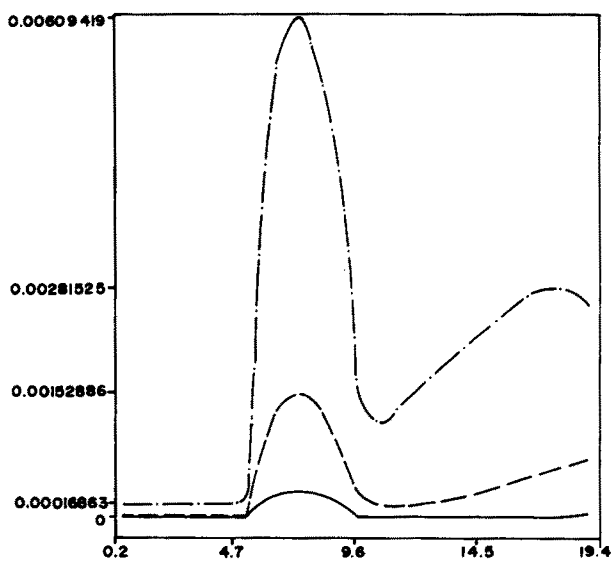


FIG. 1. The radiation flux as a function of the timelike coordinate for model I, with the initial data $A(0)=5$, $F(0)=0.4$, $\Omega(0)=0.9$, and $H=-0.55$. The solid line represents $r/a=1$, the dashed line $r/a=0.5$, and the dot–dashed line $r/a=0.25$.

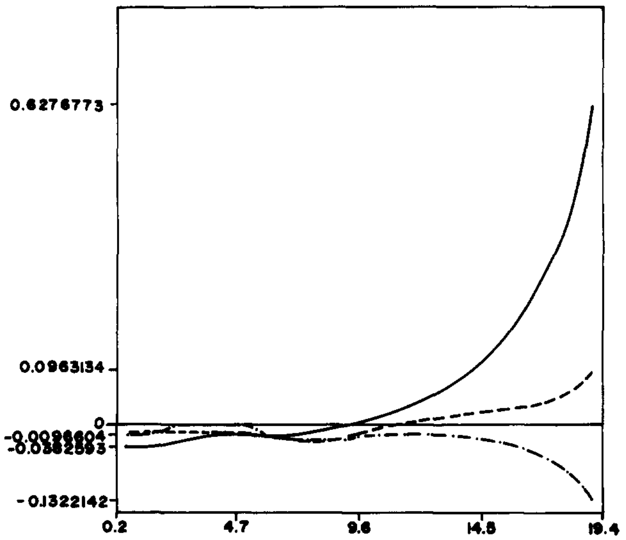


FIG. 2. The velocity (dr/du) as a function of the timelike coordinate for model I, with the same initial data and same value of H as in Fig. 1. The solid line, the dashed line, and the dot–dashed line represent, respectively, the regions $r/a = 1$, 0.5, and 0.25.

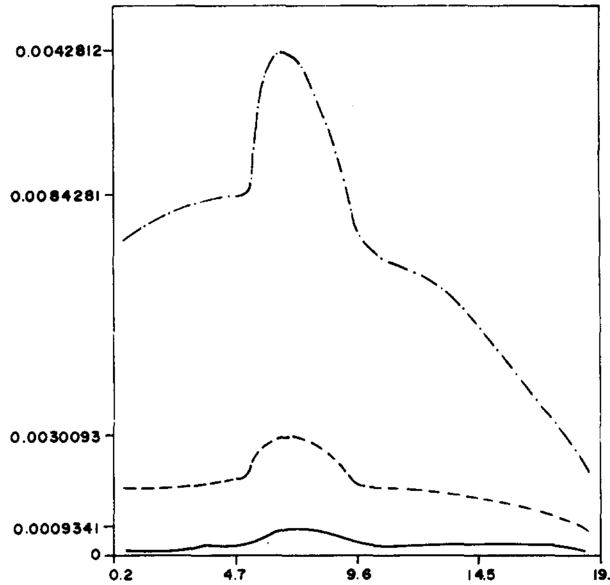


FIG. 4. $\overline{P}_\perp \equiv P_\perp m(0)^2$ (model I), as a function of the timelike coordinate for the same initial data and H as Fig. 1. The solid line, the dashed line, and the dot–dashed line represent, respectively, the regions $r/a = 0.9$, 0.5, and 0.25.

$$
\begin{aligned}
\overline{\epsilon} \equiv \epsilon m(0)^2 = {} & \frac{1}{4\pi A^2 (r^2/a^2)} \left\{ \frac{e^{-2\beta}}{8(1-2\tilde{m}/r)} \right. \\
& \times [A\dot{F}[(1+3H/2)(r/a)^3 + 3(1-H/2)(r/a)] \\
& + 2F(1-F)(\Omega-1)(1+3H/2)(r/a)^3] \\
& \left. - \frac{3}{8}\frac{(1-F)}{1-\omega}\omega\left[\left(1+\frac{3H}{2}\right)\left(\frac{r^2}{a^2}\right) + \left(1-\frac{H}{2}\right)\right]\right\} \\
& - \frac{\overline{P}\omega}{1-\omega} .
\end{aligned} \tag{72}
$$

In this model we have chosen FE to be a finite pulse (a section of a parabola), such that the total radiated mass is one tenth of the initial mass. Thus we take

$$
E = \frac{27}{20} \frac{A(u_1)(1-F(u_1))}{2F(u_2-u_1)^3} [(u_2-u_1)^2 - 4(u-\bar{u})^2] , \tag{73}
$$

with $u_2 > u > u_1$, $\bar{u} = (u_1+u_2)/2$, and $E(u) = 0$, for $u \leqslant u_1$ and $u \geqslant u_2$.

As for the initial data, we have chosen

$$
F(u=0) = 0.4 , \quad A(u=0) = 5 , \quad \Omega(u=0) = 0.9 .
$$

Figure 1 shows the flux of radiation for different regions

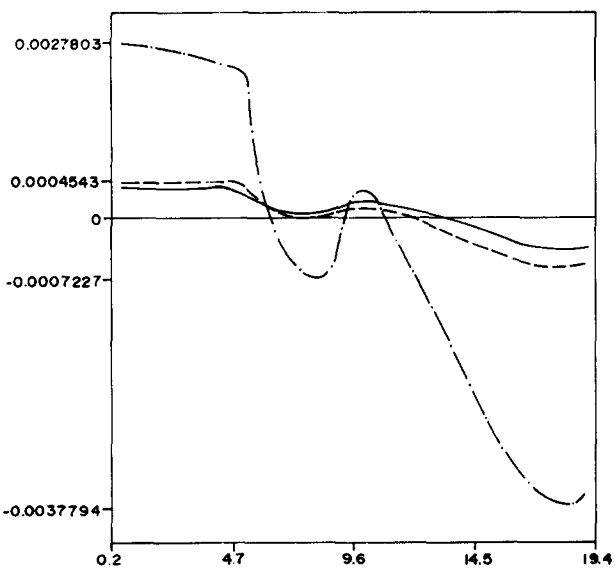


FIG. 3. $\overline{P} \equiv m(0)^2 P$ (model I) as a function of the timelike coordinate for the same initial data and H as in Fig. 1. The solid line represents the function (multiplied by 10) in the region $r/a = 0.9$. The dashed line represents the function at $r/a = 0.5$. The dot–dashed line gives the function for $r/a = 0.25$.

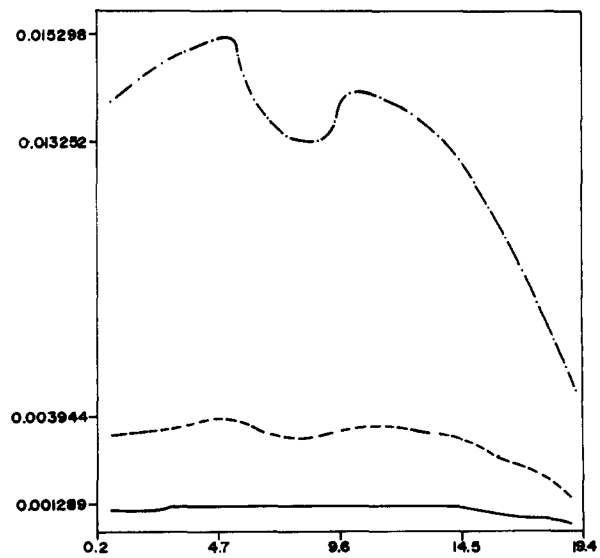


FIG. 5. $\overline{\rho} \equiv m(0)^2 \rho$ (model I) as a function of the timelike coordinate for the same initial data and H as in Fig. 1. As in Fig. 4, the regions $r/a \equiv 0.9$, 0.5, and 0.25 are represented, respectively, by the solid line, the dashed line, and the dot–dashed line.

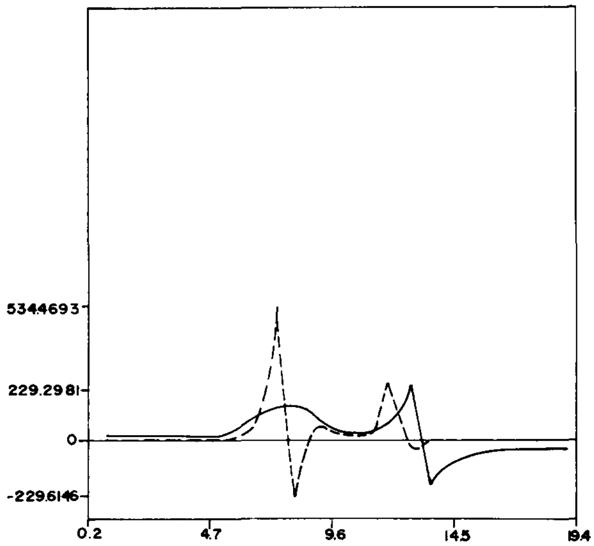


FIG. 6. $(P_\perp - P)/P$ (model I) as a function of the timelike coordinate for the same initial data and H as in Fig. 1 for the regions $r/a = 0.9$ and 0.5, represented, respectively, by the solid and the dashed line.

of the sphere (included the surface). It is worth observing the appearance of a second pulse of radiation for the inner regions and the strong absorption in the outer layers.

Figure 2 shows the profile of the velocity (here we mean dr/du, the velocity in radiative coordinates) for different pieces of the material. All regions except the inner ones ($r/a \leqslant 0.25$) expand after some time.

The profiles of the radial and tangential pressure are displayed in Figs. 3 and 4.

In Fig. 5 the evolution of the density is shown. Observe that as the second pulse starts in the region $r/a = \frac{1}{4}$ ($u \gtrsim 10{,}6$) the density of this region begins to decrease, even though this piece of material is contracting (see Fig. 2).

Finally the ratio $(P_\perp - P)/P$ is displayed in Fig. 6. It shows clearly that all regions tend to configuration with, essentially, the same degree of anisotropy as the initial ones.

We shall further discuss all these results in the last section.

V. MODEL II

In this section we shall assume an equation of state more restrictive than (62). Doing so we shall be able to integrate the surface equations without prescribing *a priori* the luminosity. In other words the energy flux at the surface will be obtained from the surface equations.

We start with the same anisotropic solution as in the preceding section [Eqs. (46)–(48)]. Thus, as in model I we choose

$$\tilde{\rho} = \frac{f(u)}{2}\left[\frac{(1+3H/2)}{a^2} + \frac{(1-H/2)}{r^2}\right]. \tag{74}$$

Also, we assume the equation of state

$$\tilde{P} + \tilde{\rho} = [\alpha(u)/r](1+H/2), \tag{75}$$

with $\alpha(u) \equiv f(u)(1-\omega_a)$. Thus

$$\tilde{P} = \frac{f(u)}{r^2}\left[(1-\omega_a)\left(1+\frac{H}{2}\right) - \frac{(1-H/2)}{2} - \frac{(1-3H/2)r^2}{2a^2}\right] \tag{76}$$

and

$$\tilde{m} = \frac{2\pi f}{a^2}\left[\frac{(1+3H/2)r^3}{3} + \left(1-\frac{H}{2}\right)a^2 r\right] \tag{77}$$

with $f(u) = (3/16\pi)(1-F)$. Now, instead of (62) the following equation of state

$$P_\perp - P = \alpha(u)/r^2 - (\tilde{\rho} + \tilde{P}) \tag{78}$$

will be assumed, which generalizes the equation

$$8\pi(P_\perp - P) = 1/r^2 - 8\pi(\rho + P), \tag{79}$$

valid for the static case.

Evaluating (78) at the surface we get

$$P_{\perp a} = -\frac{Hf(u)(1-\omega_a)}{2a^2} = -\frac{3H(1-F)}{32\pi a^2\Omega} \tag{80}$$

or

$$\bar{P}_{\perp a} \equiv m(0)^2 P_{\perp a} = \frac{-3H(1-F)}{32\pi A^2\Omega}. \tag{81}$$

We can now calculate G from Eq. (22).

Using (74), (76), (77), and (81) we obtain

$$\bar{G} \equiv m(0)G = -\frac{3(1-F)^2(\Omega-1)(2\Omega-1)(1+H/2)}{4\Omega A} + \frac{F(1-F)\Omega(1+3H/2)}{A(1+H/2)} - \frac{(1-F)^2(3-2\Omega)}{2A} - \frac{F(1-F)H}{(1+H/2)A}, \tag{82}$$

where we have also used the fact that

$$K(a) = 2/a,$$

as in model I.

Obviously, since the effective variables are the same as in model I, the expressions for $\beta, \beta_1, \beta_{11}, \beta_{01}, \tilde{m}_1$, and $\tilde{m}_{11}$ will also be the same as in the previous model.

Let us now turn to the surface equations (15), (17), (21), and (45).

We may solve the system for E to obtain

$$E = \frac{\bar{G}A\Omega}{F(2\Omega-1)+1} - \frac{(1-F)\Omega[3F(2\Omega-1)-1]}{2[F(2\Omega-1)+1]}. \tag{83}$$

Then the surface equations reduce to

$$\frac{\dot{F}}{F} = \frac{2\bar{G}A\Omega - (1-F)\{(2\Omega-1)[F(2\Omega+1)-1]\}}{A\,[F(2\Omega-1)+1]}, \tag{84}$$

$$\frac{\dot{\Omega}A}{(2\Omega-1)\Omega} = \frac{F(2\Omega+1)-1-A\bar{G}}{F(2\Omega-1)+1}, \tag{85}$$

$$\dot{A} = F(\Omega-1). \tag{86}$$

The integration of the system (84)–(86) for given initial data allows one to obtain $\beta, \beta_1, \beta_{11}, \beta_{01}, \tilde{m}_1, \tilde{m}$, and $\tilde{m}_{11}$ for any piece of the material. Then the physical variables ρ, ω, $P_\perp$, and ϵ are calculated from the expressions (68), (70),

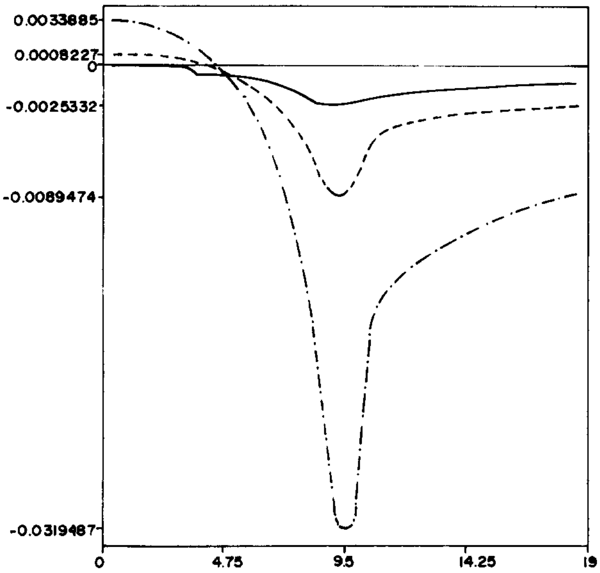


FIG. 7. The radiation flux as a function of the timelike coordinate for model II, with the initial data $F(0) = 0.6$, $A(0) = 0.5$, $\Omega(0) = 0.83$, and $H = -0.55$. The three regions $r/a = 1$, 0.5, and 0.25 are represented, respectively, by the solid line, the dashed line, and the dot–dashed line.

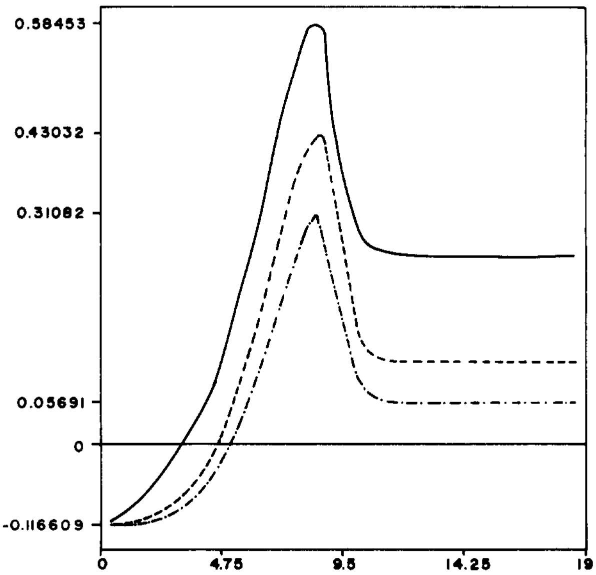


FIG. 9. dr/du as a function of the timelike coordinate (model II) for the same initial data and $H =$ as in Fig. 7, for the three regions $r/a = 1$, 0.5, and 0.25 represented, respectively, by the solid line, the dashed line, and the dot–dashed line.

(71), and (72). To calculate P we use (78), thus

$$m(0)^2P = \overline{P} = \overline{P}_\perp + 3H(1-F)/32\pi\Omega(r^2/a^2)A^2. \quad (87)$$

Finally the luminosity is obtained from (83).

As an specific example we have integrated numerically the system (84)–(86) for the initial data

$$F(u=0) = 0.6\,, \quad A(u=0) = 5\,, \quad \Omega(u=0) = 0.83\,,$$

for different values of the anisotropic constant H.

The most striking feature of this model concerns the luminosity, which some time after the initial moment becomes negative (the sphere is not emitting but absorbing radiation) (see Fig. 7). It also is interesting to note that the outer regions not only absorb but radiate inward, so that the incoming pulse across the inner regions (say $r/a = 0.25$) is bigger than the incoming pulse at the surface. After attaining a peak, the pulse slowly tends to zero (see Fig. 7).

As shown in Figs. 8 and 9, all regions of the sphere will expand after some time, but the velocity of the expansion will depend on the location of the region and the value of H.

Figures 10 and 11 display the profile of the density and

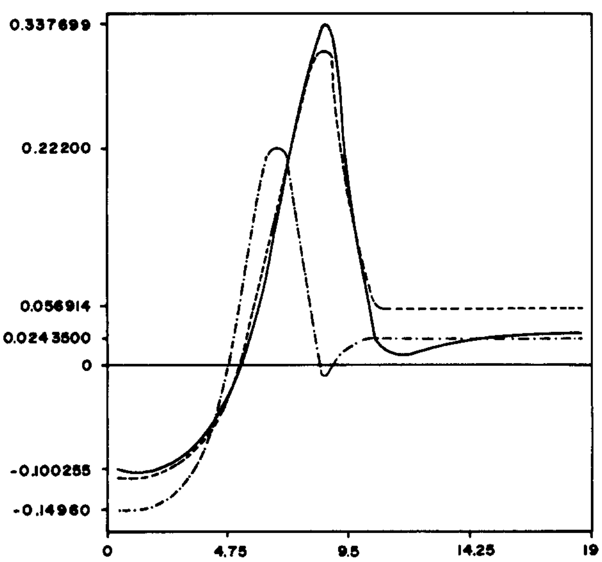


FIG. 8. dr/du as a function of the timelike coordinate (model II), for the same initial data as in Fig. 7 and $r/a = 0.25$. The solid line represents $H = -0.45$, the dashed line $H = -0.55$, and the dot–dashed line $H = -1$.

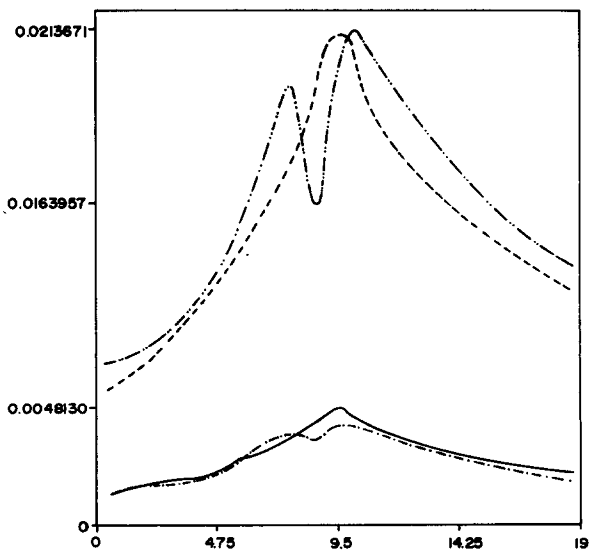


FIG. 10. $\bar{\rho} \equiv m(0)^2\rho$ as a function of the timelike coordinate (model II) for the same initial data as in Fig. 7. The solid line and the dashed line represent the function for $H = -0.55$ and $r/a = 0.5$ and 0.25, respectively. The dot–dashed line and the double dot–dashed line represent the function for $H = -1$ and $r/a = 0.5$ and 0.25, respectively.

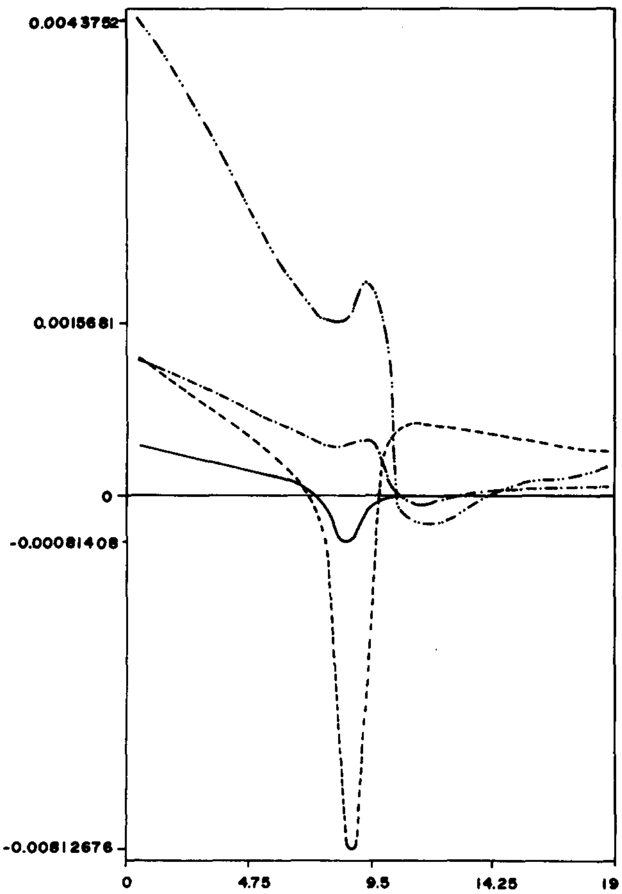


FIG. 11. $\overline{P}_\perp \equiv P_\perp m(0)^2$ as a function of the timelike coordinate (model II) for the same initial data as in Fig. 7. The solid line and the dashed line represent the function (multiplied by 0.4) for $H = -1$ and $r/a = 0.5$ and 0.25, respectively. For $H = -0.55$ and $r/a = 0.5$ and 0.25 the function is represented by a dot–dashed line and double dot–dashed line.

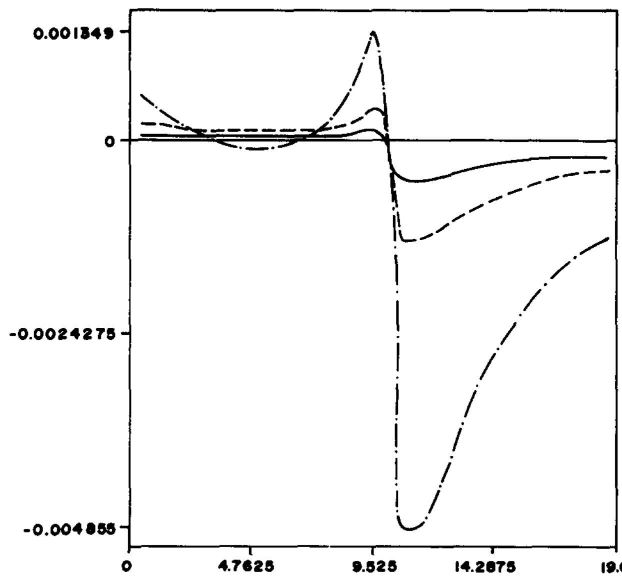


FIG. 12. $\overline{P} \equiv Pm(0)^2$ as a function of the timelike coordinate (model II) for the same initial data as in Fig. 7 and $H = -0.45$. The three regions $r/a = 0.75$, 0.5, and 0.25 are represented by the solid line, the dashed line, and the dot–dashed line, respectively.

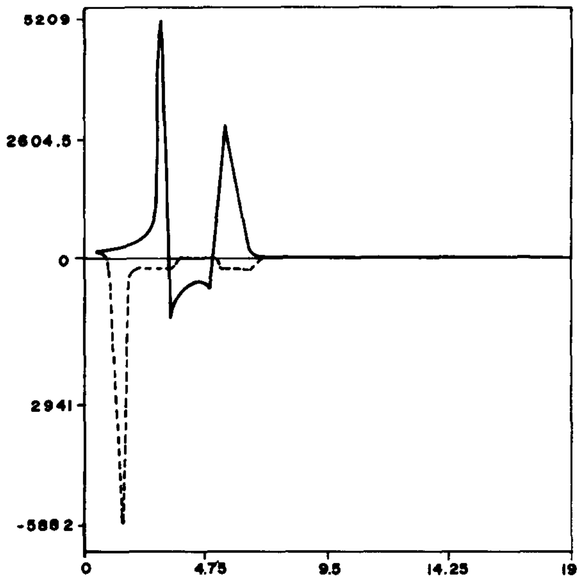


FIG. 13. $(P_\perp - P)/P$ (model II) as a function of the timelike coordinate, for the same initial data and the same H as in Fig. 7. The two regions $r/a = 0.9$ and 0.45 are represented by the solid line and the dot–dashed line, respectively.

the tangential pressure for different values of H and different pieces of the material.

Figure 12 shows the evolution of the radial pressure for $H = -0.45$ and different regions of the sphere. The behavior of the radial pressure near $u \approx 10$ suggests the formation of a shock. Though it is not clear from the figure, a more detailed analysis shows that the shock moves inward.

Finally Fig. 13 shows that the degree of anisotropy decreases with time after some wild swings when the peak of the incoming pulse of radiation passes by.

In the next section we shall try to relate the results of the precedent section with some astrophysical processes.

VI. DISCUSSION

The model I, considered in Sec. IV, shares some features that strongly remind us of the general pattern of a supernova: An inner core which contracts $(r/a) \approx \frac{1}{4}$, and the expansion, after a bounce, of the more external regions. It is worth stressing that even during a certain period of contraction, the density of the inner region is decreasing, due to the appearance of the second pulse of radiation, which is characteristic for that region (see Fig. 1).

A model for the collapse of a massive star, whose central core loses an important portion of it mass by emission of radiation (neutrino), was considered many years ago by Michel.[12] The main advantage of that kind of model consists in the fact that the expansion of the envelope does not require extremely efficient transport mechanisms.

The second model (Sec. V) presents the strange property that some time after the initial moment the sphere is no longer emitting but absorbing radiation, a situation that one does not expect to deal with in the realm of astrophysical processes. However, a few years ago a model was proposed[13] to explain the origin of gas in quasars. According to this model, red dwarf stars close to a quasar absorb the quasar radiation and reverse their evolution by expanding.

Doing so, the stars would have lower surface gravities and could be ablated by radiation pressure, giving rise to the so-called "broad line clouds" in quasars.

Our sphere in model II expands during the absorption period, with the peculiarity that the incoming flux grows up as we move inward into the sphere. Thus, the central regions are attained not only by the absorbed radiation at the surface but the radiation emitted (inward) by the outer regions during the expansion.

We find encouraging the fact that relatively simple models, such as the one presented here, could reproduce the general features of some astrophysical phenomena. It remains to be seen if the models could be specialized so as to describe those processes in finer detail.

Finally we would like to mention that the perfect-fluid case ($P_\perp = P$) has been integrated also. However, in this case, the luminosity, which is inferred from the surface equations, is negative, and its absolute value is a monotonically increasing function of the timelike coordinate.

Thus the matching of perfect-fluid solutions with the Vaidya metric on the boundary of the fluid implies a rather unphysical flux of radiation across that boundary. This result, in some sense, generalizes a previous result[3] [see point (3) in the Introduction] concerning the impossibility of matching a conformally symmetric perfect-fluid solution (with the additional restriction $\xi^\alpha U_\alpha = 0$) with the vacuum Schwarzschild metric.

ACKNOWLEDGMENT

We would like to thank A. H. Taub for interesting comments and discussions.

[1] L. Herrera, J. Jiménez, L. Leal, M. Esculpi, J. Ponce de León, and V. Galina, J. Math. Phys. **25**, 3274 (1984).
[2] L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 778 (1985).
[3] L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 2018 (1985).
[4] L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 2847 (1985).
[5] L. Herrera and J. Jiménez, and G. Ruggeri, Phys. Rev. D **22**, 2305 (1980).
[6] M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, Phys. Rev. D **25**, 2527 (1982).
[7] A. Patiño and H. Rago, Lett. Nuovo Cimento **38**, 321 (1983).
[8] K. D. Krori, P. Bergehain, and R. Sarma, Phys. Rev. D **31**, 734 (1985).
[9] H. Bondi, Proc. R. Soc. London Ser. A **281**, 39 (1964).
[10] H. Bondi, M. G. W. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[11] L. Herrera and J. Jiménez, Phys. Rev. D **28**, 2987 (1983).
[12] F. C. Michel, Astrophys. J. **138**, 1097 (1963).
[13] W. G. Mathews, Astrophys. J. **272**, 390 (1983).

Axially symmetric gravitational radiation from isolated sources and the deviation from spherical symmetry

L. Herrera[a)] and J. Jimenez
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela

(Received 4 December 1985; accepted for publication 6 March 1986)

Axially (and reflection) symmetric space-times (Bondi metric) are studied in terms of scalar functions that measure the deviation of the system from spherical symmetry. The case of small departures from sphericity is considered, and the corresponding solution is completely specified up to the first order. Finally in discussing some aspects of the initial (characteristic) value problem, a very concise proof of a theorem by Papapetrou [A. Papapetrou, J. Math. Phys. 6, 1405 (1965)] is given.

I. INTRODUCTION

The problem of gravitational radiation emitted from an isolated source has been studied from many different points of view and by means of different approaches and techniques. One of the most interesting approaches to this problem is, probably, the method first introduced by Bondi *et al.*[1] and later generalized by Sachs.[2]

The purpose of this paper is to present a partial restatement of the Bondi formalism in terms of scalar functions, which, we believe, may help us to gain a deeper insight into the nature of radiating systems. More specifically, these functions will measure the departure of a given system (asymptotically flat, axially and reflection symmetric) from sphericity and staticity.

The motivation to undertake such a task arises from the simple fact that radiation (as it follows from the Birkhoff's theorem) is a process associated with deviations of spherical symmetry. Thus, functions measuring departures from sphericity and staticity are expected to contain all the essential information of a radiating system. This is, by the way, the very idea underlying the multipole expansion approach.

We shall display the expressions that give the metric explicitly in terms of those functions.

Next, we shall consider small departures from sphericity. Then the solution (up to first order) will be found. It will be seen up to this (first) order of approximation that the Bondi mass is constant although the metric is, in general, time dependent. Finally we shall consider systems that, on a given null hypersurface ($u = \text{const}$), coincide with some static solution. As a "by-product" of this analysis, a theorem by Papapetrou is proved in a very simple way.

II. THE METHOD

Let us consider a nonstatic, axially and reflection symmetric metric[1] that, in radiation coordinates, takes the form

$$ds^2 = (Vr^{-1}e^{2b} - U^2r^2e^{2g})du^2 + 2e^{2b}\,du\,dr + 2Ur^2e^{2g}\,du\,d\theta - r^2(e^{2g}\,d\theta^2 + e^{-2g}\sin^2\theta\,d\phi^2), \tag{1}$$

where U, V, g, and b are functions of u, θ, and r. Here $u \equiv x^0$ is the timelike coordinate, $r = x^1$ is a null coordinate, and θ and ϕ are two angular coordinates. The condition that the solution be truly isolated requires that the metric functions be regular everywhere; in particular on the polar axis ($\theta = 0,\pi$). This means that V, b, $(U/\sin\theta)$, and $(g/\sin^2\theta)$ are regular functions of $\cos\theta$ as $\cos\theta = \pm 1$. We would like to stress that this condition will be satisfied all through this paper, and that its violation would lead to a completely different set of results.

It is well known that the field equations are split into two groups: the main equations and the supplementary conditions (actually there is also a trivial equation). The former read

$$b_1 = \tfrac{1}{2}rg_1^2, \tag{2}$$

$$[r^4e^{2(g-b)}U_1]_1 - 2r^2[b_{12} - g_{12} + 2g_1g_2 - 2b_2r^{-1} - 2g_1\cot\theta] = 0, \tag{3}$$

$$2V_1 + \tfrac{1}{2}r^4e^{2(g-b)}U_1^2 - r^2U_{12} - 4rU_2 - r^2U_1\cot\theta - 4r\cot\theta U + 2e^{2(b-g)} \times[-1 - (3g_2 - b_2)\cot\theta - g_{22} + b_{22} + b_2^2 + 2g_2(g_2 - b_2)] = 0, \tag{4}$$

$$2r(rg)_{01} + (1 - rg_1)V_1 - (rg_{11} + g_1)V - r(1 - rg_1)U_2 - r^2(\cot\theta - g_2)U_1 + r(2g_{12}r + 2g_2 + rg_1\cot\theta - 3\cot\theta)U + e^{2(b-g)}[-1 - (3g_2 - 2b_2)\cot\theta - g_{22} + 2g_2(g_2 - b_2)] = 0. \tag{5}$$

Differentiation with respect to u, θ, and r are denoted by subscripts 0, 1, and 2, respectively.

Next, if one assumes that the metric functions may be expanded in terms of series in powers of r^{-1}, then using (2)–(5) one obtains

$$g = c(u,\theta)r^{-1} + [C(u,\theta) - \tfrac{1}{6}c^3]r^{-3} + O(r^{-4}), \tag{6}$$

$$b = -(c^2/4)r^{-2} + O(r^{-4}) \tag{7}$$

$$U = -(c_2 + 2c\cot\theta)r^{-2} + [2N(u,\theta) + 3cc_2 + 4c^2\cot\theta]r^{-3} + O(r^{-4}), \tag{8}$$

$$V = r - 2M(u,\theta) - [N_2 + N\cot\theta - c_2^2 - 4c_2c\cot\theta - \tfrac{1}{2}c^2(1 + 8\cot\theta)^2]r^{-1} + O(r^{-3}), \tag{9}$$

with

a) Postal address: L. Herrera, Apartado 80793, Caracas 1080A, Venezuela.

 0022-2488/86/082097-08$02.50

$$4C_0 = 2c^2c_0 + 2cM + N\cot\theta - N_2. \tag{10}$$

The three arbitrary functions of integration, M, N, and c, are related by the two supplementary conditions

$$M_0 = -c_0^2 + \tfrac{1}{2}(c_{22} + 3c_2\cot\theta - 2c)_0, \tag{11}$$

$$-3N_0 = M_2 + 3cc_{02} + 4cc_0\cot\theta + c_0c_2. \tag{12}$$

Next, the associated tetrad may be written as

$$l^\mu = e^{-2b}\delta_1^\mu,\quad n^\mu = \delta_0^\mu - (V/2r)\delta_1^\mu + U\delta_2^\mu,$$
$$m^\mu = (1/r\sqrt{2})(e^{-g}\delta_2^\mu + ie^g\csc\theta\delta_3^\mu), \tag{13}$$

or, in covariant components,

$$l_\mu = \delta_\mu^0,\quad n_\mu = (Ve^{2b}/2r)\delta_\mu^0 + e^{2b}\delta_\mu^1,$$
$$m_\mu = (r/\sqrt{2})(Ue^g\delta_\mu^0 - e^g\delta_\mu^2 - ie^{-g}\sin\theta\,\delta_\mu^3). \tag{14}$$

For the spin coefficients we get[3]

$$\kappa = \epsilon = 0,\quad \rho = -e^{-2b}/r,\quad \sigma = -e^{-2b}g_1,$$
$$\gamma = \frac{e^{-2b}}{2}\left[\left[\frac{Ve^{2b}}{2r}\right]_1 - (e^{2b})_0 - U(e^{2b})_2\right],$$
$$\alpha = \frac{1}{4}\left[\frac{re^g e^{-2b}}{\sqrt{2}}U_1 - \frac{\sqrt{2}e^{-g}}{r}b_2 - \frac{\sqrt{2}e^{-g}}{r}(\cot\theta - g_2)\right],$$
$$\beta = \frac{1}{4}\left[\frac{re^g e^{-2b}}{\sqrt{2}}U_1 - \frac{\sqrt{2}e^{-g}}{r}(b_2 + g_2 - \cot\theta)\right],$$
$$\tau = \frac{e^{-2b}}{2\sqrt{2}}\left[re^gU_1 - \frac{2e^{-g}e^{2b}}{r}b_2\right], \tag{15}$$
$$\pi = \frac{e^{-2b}}{2\sqrt{2}}\left[re^gU_1 + \frac{2e^{-g}e^{2b}}{r}b_2\right],$$
$$\lambda = U\left[g_2 - \frac{\cot\theta}{2}\right] + g_0 + \frac{1}{2}\left[U_2 - \frac{V}{r}g_1\right],$$
$$\mu = \frac{U}{2}\cot\theta - \frac{V}{2r^2} + \frac{U_2}{2},\quad \nu = \frac{e^{2b}V_2e^{-g}}{2\sqrt{2}r^2},$$

and the Ricci scalars are given, in terms of the spin coefficients, by[3]

$$\Phi_{00} = D\rho - \bar\delta\kappa - (\rho^2 + \sigma\bar\sigma) - (\epsilon + \bar\epsilon)\rho + \kappa\tau + \kappa(3\alpha + \bar\beta - \bar\pi), \tag{16}$$

$$\Phi_{10} = D\alpha - \bar\delta\epsilon - (\rho + \bar\epsilon - 2\epsilon)\alpha - \beta\bar\sigma + \beta\epsilon + \kappa\lambda + \bar\kappa\lambda - (\epsilon + \rho)\pi, \tag{17}$$

$$\Phi_{20} = D\lambda - \bar\delta\pi - (\rho\lambda + \bar\sigma\mu) - \pi^2 - (\alpha - \bar\beta)\pi - \nu\bar\kappa + (3\epsilon - \bar\epsilon)\lambda, \tag{18}$$

$$\Phi_{22} = \delta\nu - \Delta\mu - \mu^2 + 3\beta\nu - \bar\lambda\lambda + \bar\alpha\nu - \tau\nu + \bar\nu\pi - \bar\gamma\mu - \gamma\mu, \tag{19}$$

$$\Phi_{12} = \delta\gamma - \Delta\beta + (\bar\alpha + \beta - \tau)\gamma - \mu\tau + \sigma\nu + \epsilon\bar\nu + \beta(\gamma - \bar\gamma - \mu) - \alpha\bar\lambda, \tag{20}$$

$$\Phi_{11} = \tfrac{1}{2}\{D\gamma - \Delta\epsilon + \delta\alpha - \bar\delta\beta - (\tau + \bar\pi)\alpha - (\bar\tau + \pi)\beta + (\epsilon + \bar\epsilon)\gamma + \epsilon(\gamma + \bar\gamma) - \tau\pi - \nu\kappa - (\mu\rho - \lambda\sigma) - \alpha\bar\alpha - \beta\bar\beta + 2\alpha\beta - \gamma(\rho - \bar\rho) - \epsilon(\mu - \bar\mu)\}, \tag{21}$$

$$\Lambda = \tfrac{1}{3}\{D\mu - \delta\pi - \delta\alpha - \bar\delta\beta - \mu(\rho + \bar\rho) - \gamma(\rho - \bar\rho) + \nu\kappa + \bar\epsilon\mu + \epsilon\bar\mu - \alpha\bar\alpha - \beta\bar\beta - \pi\bar\pi + \pi(\bar\alpha - \beta) - 2\alpha\beta - \Phi_{11}\}, \tag{22}$$

where

$$D \equiv e^{-2b}\frac{\partial}{\partial r},\quad \Delta \equiv \frac{\partial}{\partial u} - \frac{V}{2r}\frac{\partial}{\partial r} + U\frac{\partial}{\partial\theta},$$
$$\delta \equiv \frac{1}{\sqrt{2}r}\left(e^{-g}\frac{\partial}{\partial\theta} + \frac{ie^g}{\sin\theta}\frac{\partial}{\partial\phi}\right). \tag{23}$$

As is well known, the Bondi metric (in its most general form) admits only one Killing vector, associated with the axial symmetry. On the other hand, in the spherically symmetric case (Schwarzschild) there are three additional Killing vectors associated with the sphericity and staticity. These Killing vectors may be expressed as

$$\underset{(1)}{\xi}{}^\alpha = \tfrac{1}{2}(1 - 2m/r)l_s^\alpha + n_s^\alpha, \tag{24}$$

$$\underset{(2)}{\xi}{}^\alpha = -(r/\sqrt{2})(\cos\phi + i\cos\theta\sin\phi)m_s^\alpha - (r/\sqrt{2})(\cos\phi - i\cos\theta\sin\phi)\bar m_s^\alpha, \tag{25}$$

$$\underset{(3)}{\xi}{}^\alpha = -(r/\sqrt{2})(-\sin\phi + i\cos\theta\cos\phi)m_s^\alpha - (r/\sqrt{2})(-\sin\phi - i\cos\theta\cos\phi)\bar m_s^\alpha, \tag{26}$$

where $(l_s^\alpha, n_s^\alpha, m_s^\alpha, \bar m_s^\alpha)$ form the tetrad associated with the Schwarzschild metric (i.e., $b = g = U = 0$, $V = r - 2m$) and m is a constant representing Schwarzschild mass. As a first step we shall calculate the tetrad projections of the Lie derivative of the Bondi metric with respect to vector fields such that, in the spherically symmetric case, they coincide with (24)–(26) [as a matter of fact we only need to consider (24) and either (25) or (26)]. Clearly, these quantities (tetrad projections of the Lie derivatives), while nonvanishing for a general Bondi metric, will measure how far the system is from spherical symmetry.

Now, the tetrad components of the Lie derivatives of the metric tensor with respect to a general vector field ξ^α are[4]

$$2A\,\mathrm{Re}(\gamma) + 2\,\mathrm{Re}(C\bar\nu) + \Delta A, \tag{27}$$

$$2A\,\mathrm{Re}(\epsilon) - 2B\,\mathrm{Re}(\gamma) + 2\,\mathrm{Re}(C\bar\pi) - 2\,\mathrm{Re}(C\tau) + DA + \Delta B, \tag{28}$$

$$B\bar\nu - A(\bar\alpha + \beta + \tau) - C\bar\lambda - \bar C\mu - 2i\bar C\,\mathrm{Im}(\gamma) - \delta A + \Delta\bar C, \tag{29}$$

$$-2B\,\mathrm{Re}(\epsilon) - C\kappa - \bar C\bar\kappa + DB, \tag{30}$$

$$-A\kappa + B(\bar\alpha + \beta + \bar\pi) + C\sigma + \bar C[\bar\rho - 2i\,\mathrm{Im}(\epsilon)] - \delta B + D\bar C, \tag{31}$$

$$2A\,\mathrm{Re}(\rho) - 2B\,\mathrm{Re}(\mu) + 2\,\mathrm{Re}[C(\bar\alpha - \beta)] - \delta C - \bar\delta\bar C, \tag{32}$$

$$A\sigma - B\bar\lambda - \delta\bar C - \bar C(\bar\alpha - \beta), \tag{33}$$

with

$$\xi^\alpha = Al^\alpha + Bn^\alpha + Cm^\alpha + \bar C\bar m^\alpha \tag{34}$$

[this C is not to be confounded with the function $C(u,\theta)$ in

Eqs. (6)–(10)]. Let us start with the vector field (24), generating the time independence of the Schwarzschild metric, we obtain, using (27)–(33),

$$\Delta A + 2A\gamma \equiv \overset{(1)}{f}, \tag{35}$$

$$DA - 2\gamma \equiv \overset{(2)}{f}, \tag{36}$$

$$\nu - A(\alpha + \beta + \tau) \equiv \overset{(3)}{f}, \tag{37}$$

$$DB \equiv 0, \tag{38}$$

$$\alpha + \beta + \pi \equiv \overset{(4)}{f}, \tag{39}$$

$$2A\rho - 2\mu \equiv \overset{(5)}{f}, \tag{40}$$

$$a\sigma - \lambda \equiv \overset{(6)}{f}, \tag{41}$$

with $A = \frac{1}{2}(1 - 2m/r)$, $B = 1$, $C = 0$, and the functions f are defined by (35)–(41). At this point it is important to make the following remark: the vector field ξ^α with the values of the coefficients A, B, C, as given above, is a Killing vector generating the time independence for the Schwarzschild metric, but not for a static axially symmetric metric (Weyl). We make this choice because we are specially interested in deviations from spherical symmetry. It should be clear, however, that this choice is not unique. Equations (35)–(41) may be rewritten as

$$\gamma = \tfrac{1}{2}\left[DA - \overset{(2)}{f}\right] = (1/2A)\left[\overset{(1)}{f} - \Delta A\right], \tag{42}$$

$$\nu = A(\alpha + \beta + \tau) + \overset{(3)}{f}, \tag{43}$$

$$\alpha + \beta + \pi = \overset{(4)}{f}, \tag{44}$$

$$\mu = A\rho - \tfrac{1}{2}\overset{(5)}{f}, \tag{45}$$

$$\lambda = A\sigma - \overset{(6)}{f}. \tag{46}$$

Next, let us consider the vector field (25), associated with the spherical symmetry of the Schwarzschild metric. We obtain, from (27)–(33),

$$2\nu \operatorname{Re} C \equiv \overset{(1)}{F} \cos\phi, \tag{47}$$

$$2(\pi - \tau)\operatorname{Re} C \equiv \overset{(2)}{F} \cos\phi, \tag{48}$$

$$\Delta\overline{C} - C\lambda - \overline{C}\mu \equiv \overset{(3)}{F} \cos\phi + i\,\overset{(4)}{F} \sin\phi, \tag{49}$$

$$D\overline{C} + C\sigma + \overline{C}\rho \equiv \overset{(5)}{F} \cos\phi + i\,\overset{(6)}{F} \sin\phi, \tag{50}$$

$$2(\alpha - \beta)\operatorname{Re} C - 2\operatorname{Re}(\delta C) \equiv \overset{(7)}{F} \cos\phi, \tag{51}$$

$$\overline{C}(\beta - \alpha) - \delta\overline{C} = \overset{(8)}{F} \cos\phi + i\,\overset{(9)}{F} \sin\phi, \tag{52}$$

where, again, the functions F are defined by the tetrad component of the Lie derivative of the metric tensor with respect to a vector field with tetrad components

$$A = B = 0, \quad C = (-r/\sqrt{2})(\cos\phi + i\cos\theta\sin\phi).$$

Such a vector field will coincide with (25) for the spherically symmetric case. Next it follows, from (47)–(52) and (11), that

$$\nu = -\overset{(1)}{F}/\sqrt{2}r, \tag{53}$$

$$\tau - \pi = \overset{(2)}{F}/\sqrt{2}r, \tag{54}$$

$$\lambda = \frac{1}{r\sqrt{2}}\left[\overset{(3)}{F} + \frac{\overset{(4)}{F}}{\cos\theta}\right] + \frac{U}{2}\tan\theta, \tag{55}$$

$$\mu = \frac{1}{r\sqrt{2}}\left[\overset{(3)}{F} - \frac{\overset{(4)}{F}}{\cos\theta}\right] - \frac{V}{2r^2} - \frac{U}{2}\tan\theta, \tag{56}$$

$$\rho = -\frac{e^{-2b}}{r} + \frac{1}{r\sqrt{2}}\left(\frac{\overset{(6)}{F}}{\cos\theta} - \overset{(5)}{F}\right), \tag{57}$$

$$\sigma = -\frac{1}{r\sqrt{2}}\left[\overset{(5)}{F} + \frac{\overset{(6)}{F}}{\cos\theta}\right], \tag{58}$$

$$\overset{(7)}{F}/2 + \overset{(8)}{F} = 0, \tag{59}$$

$$(\beta - \alpha) - \frac{e^g}{\sqrt{2}r}\cot\theta = \frac{\overset{(7)}{F}}{\sqrt{2}r}, \tag{60}$$

$$\overset{(9)}{F} - \frac{\overset{(7)}{F}}{2}\cos\theta = \frac{e^{-g}\sin\theta - e^g\sin\theta}{2} = -\sin\theta\sinh g. \tag{61}$$

Also, it follows from (11) that

$$\tau = \alpha + \beta, \tag{62}$$

thus

$$\tau + \pi = \overset{(4)}{f}, \tag{63}$$

$$\tau - \pi = \overset{(2)}{F}/\sqrt{2}r, \tag{64}$$

or

$$\tau = \tfrac{1}{2}\left[\overset{(4)}{f} + \overset{(2)}{F}/\sqrt{2}r\right], \tag{65}$$

$$\pi = \tfrac{1}{2}\left[\overset{(4)}{f} - \overset{(2)}{F}/\sqrt{2}r\right], \tag{66}$$

and

$$\beta - \alpha = \overset{(7)}{F}/\sqrt{2}r + (e^g\cot\theta)/\sqrt{2}r. \tag{67}$$

Then using (42)–(67), we can express the spin coefficients in terms of the functions f and F, as follows:

$$\kappa = \epsilon = 0, \tag{68}$$

$$\gamma = \tfrac{1}{2}\left[DA - \overset{(2)}{f}\right], \tag{69}$$

$$\rho = -\frac{e^{-2b}}{r} + \frac{1}{\sqrt{2}r}\left[\frac{\overset{(6)}{F}}{\cos\theta} - \overset{(5)}{F}\right], \tag{70}$$

$$\sigma = -\frac{1}{\sqrt{2}r}\left[\overset{(5)}{F} + \frac{\overset{(6)}{F}}{\cos\theta}\right], \tag{71}$$

$$\alpha = \frac{1}{4}\left[\overset{(4)}{f} + \frac{\overset{(2)}{F}}{\sqrt{2}r} - \frac{\sqrt{2}}{r}\left(\overset{(7)}{F} + e^{g}\cot\theta\right)\right], \tag{72}$$

$$\beta = \frac{1}{4}\left[\overset{(4)}{f} + \frac{\overset{(2)}{F}}{\sqrt{2}r} + \frac{\sqrt{2}}{r}\left(\overset{(7)}{F} + e^{g}\cot\theta\right)\right], \tag{73}$$

$$\tau = \tfrac{1}{2}\left[\overset{(4)}{f} + \overset{(2)}{F}/\sqrt{2}r\right], \tag{74}$$

$$\lambda = \frac{1}{\sqrt{2}r}\left[\overset{(3)}{F} + \frac{\overset{(4)}{F}}{\cos\theta}\right] + \frac{U}{2}\tan\theta, \tag{75}$$

$$\pi = \tfrac{1}{2}\left[\overset{(4)}{f} - \overset{(2)}{F}/\sqrt{2}r\right], \tag{76}$$

$$\mu = \frac{1}{r\sqrt{2}}\left[\overset{(3)}{F} - \frac{\overset{(4)}{F}}{\cos\theta}\right] - \frac{V}{2r^2} - \frac{U}{2}\tan\theta, \tag{77}$$

$$\nu = -\overset{(1)}{F}/\sqrt{2}r. \tag{78}$$

These have the following constraints:

$$\overset{(1)}{f} = A\left[DA - \overset{(2)}{f}\right] + \Delta A, \tag{79}$$

$$\overset{(1)}{F}/\sqrt{2}r + A\left[\overset{(4)}{f} + \overset{(2)}{F}/\sqrt{2}r\right] + \overset{(3)}{f} = 0, \tag{80}$$

$$\overset{(5)}{f} = 2A\rho - 2\mu, \tag{81}$$

$$\overset{(6)}{f} = A\sigma - \lambda, \tag{82}$$

$$\overset{(7)}{F}/2 + \overset{(8)}{F} = 0, \tag{83}$$

$$\left(\overset{(7)}{F}/2\right)\cos\theta - \overset{(9)}{F} = \sin\theta\sinh g. \tag{84}$$

Next, we shall express the metric functions b, g, U, and V in terms of the functions f and F.

Thus, from

$$\rho = e^{-2b}/r,$$

and (70) and (71), it follows at once that

$$\overset{(5)}{F} = \overset{(6)}{F}/\cos\theta, \tag{85}$$

and

$$\sigma = -(\sqrt{2}/r)\,\overset{(5)}{F}. \tag{86}$$

Then from the field equation

$$\Phi_{00} = 0 = D\rho - \rho^2 - \sigma^2 = 0, \tag{87}$$

we obtain, using (70), (71), and (85),

$$e^{-4b} = 1 + 4\int_r^\infty \frac{\left[\overset{(5)}{F}(u,r',\theta)\right]^2}{r'}\,dr', \tag{88}$$

and from

$$\sigma = -g_1 e^{-2b}, \tag{89}$$

we get, using (86),

$$g = -\sqrt{2}\int_r^\infty \frac{e^{2b}\,\overset{(5)}{F}(u,r',\theta)}{r'}\,dr'. \tag{90}$$

Next, using (63) and (15)

$$U = -\int_r^\infty \frac{\sqrt{2}}{r'}\,e^{(2b-g)}\,\overset{(4)}{f}(u,\theta,r')\,dr'. \tag{91}$$

Finally, from (42) it follows that

$$V = e^{-2b}(r-2m) - 2e^{-2b}r \times \int_r^\infty \left[e^{2b}\,\overset{(2)}{f}(u,r',\theta) - (e^{2b})_0 - U(e^{2b})_2\right]dr'. \tag{92}$$

We would like to close this section with the following remark: The vanishing of the functions F implies the vanishing of the f's, as required by Birkhoff's theorem. Furthermore, the vanishing of $\overset{(5)}{F}$ alone leads to the spherically symmetric case. In fact, if $\overset{(5)}{F} = 0$, then it follows at once from (88) and (90) that

$$b = g = 0, \tag{93}$$

and from (3) we get

$$U = 2N(u,\theta)/r^3, \tag{94}$$

where we have used (6), (8), and (93). Next, from (4) and (5) we obtain

$$2V_1 + 18N^2/r^4 - 2 = 0, \tag{95}$$

$$V_1 - 2N_2/r^2 - 1 = 0, \tag{96}$$

which imply

$$N = 0, \tag{97}$$

and, by virtue of (94),

$$U = 0. \tag{98}$$

Then using (95) [or (96)] and the supplementary conditions (11) and (12), we have

$$V = r - 2M, \tag{99}$$

$$M = \text{const.}$$

Equations (93), (98), and (99) define the Schwarzschild metric in Bondi coordinates. The fact that the vanishing of $\overset{(5)}{F}$ implies the spherical symmetry will be exploited in the next section to study small perturbations off Schwarzschild.

III. SMALL PERTURBATIONS OFF SCHWARZSCHILD

Let us now consider small departures from the spherical symmetry. In the language of the preceding section that means

$$\overset{(5)}{F} = \epsilon h(u,r,\theta), \tag{100}$$

where $|\epsilon| < 1$ and $h(u,r,\theta)$ is an arbitrary function of its arguments. Neglecting terms of the order $O(\epsilon^n)$ with $n \geq 2$, we obtain, from (88) and (90),

$$b = 0, \tag{101}$$

$$g = -\sqrt{2}\epsilon \int_r^\infty \frac{h\,dr'}{r'}. \tag{102}$$

Next, from (3), we get, up to the first order in ϵ,

$$[r^4 e^{2(g-b)} U_1]_1 = [r^4 e^{2g} U_1]_1 = 2r^2(-g_{12} - 2g_1 \cot\theta), \tag{103}$$

or, using (102),

$$U = -\frac{L(u,\theta)}{3r^3} - 2\sqrt{2}\epsilon \int_r^\infty \frac{1}{r'^4} \times \left\{ L(u,\theta) \int_{r'}^\infty \frac{h(u,r'',\theta)}{r''} dr'' - \left[\int r'(h_2 + 2h\cot\theta)\,dr' \right] \right\} dr'. \tag{104}$$

Similarly, we can find an expression for V. Using the field equation (4) we obtain

$$\begin{aligned}
&+ 2r\left\{ -\frac{L_2}{3r^3} - 2\sqrt{2}\,\epsilon \int_r^\infty \left[L\int_{r'}^\infty \frac{h\,dr''}{r''} - \int r'(h_2 + 2h\cot\theta)\,dr' \right]_2 \frac{dr'}{r'^4} \right\} \\
&+ \frac{r^2}{2}\cot\theta \left\{ \frac{L}{r^4} + \frac{2\sqrt{2}\epsilon}{r^4}\left[L\int_r^\infty \frac{h\,dr'}{r'} - \int r(h_2 + 2h\cot\theta)\,dr \right] \right\} \\
&+ 2r\cot\theta \left\{ -\frac{L}{3r^3} - 2\sqrt{2}\epsilon \int_r^\infty \left[L\int_{r'}^\infty \frac{h\,dr''}{r''} - \int r'(h_2 + 2h\cot\theta)\,dr' \right] \frac{dr'}{r'^4} \right\} \\
&- \frac{1}{4} r^4 \left(1 - 2\sqrt{2}\epsilon \int_r^\infty \frac{h\,dr'}{r'} \right) \left\{ \frac{L^2}{r^8} + \frac{4\sqrt{2}L\epsilon}{r^8}\left[L\int_r^\infty \frac{h\,dr'}{r'} - \int r(h_2 + 2h\cot\theta)dr \right] \right\} \\
&+ \left(1 + 2\sqrt{2}\epsilon \int_r^\infty \frac{h}{r'}dr' \right)\left[1 - 3\cot\theta\sqrt{2}\epsilon \int_r^\infty \frac{h_2}{r'}dr' - \sqrt{2}\epsilon \int_r^\infty \frac{h_{22}}{r'}dr' \right],
\end{aligned} \tag{105}$$

with $L(u,\theta) = \tilde{L}(u,\theta) + l(u,\theta)\epsilon$, which may be written as

$$V_1 = 1 - \frac{L_2}{6r^2} - \frac{L\cot\theta}{6r^2} - \frac{L^2}{4r^4} + \sqrt{2}\epsilon R(u,r,\theta), \tag{106}$$

where $R(u,r,\theta)$ represent the terms multiplied by ϵ in (105). Or, integrating,

$$V = r - 2M(u,\theta) + \frac{L_2 + L\cot\theta}{6r} + \frac{L^2}{12r^3} + \sqrt{2}\epsilon \int R\,dr, \tag{107}$$

where the function of integration $-2M(u,\theta)$ has been taken from (9). Finally, feeding back these results in (5), we obtain

$$-\frac{(\tilde{L}\cot\theta + \tilde{L}_2)}{6r^2} - \frac{\tilde{L}^2}{4r^4} + \frac{\tilde{L}_2}{3r^2} + O(\epsilon) = 0, \tag{108}$$

which implies

$$\tilde{L} = 0. \tag{109}$$

Thus

$$U = 2\sqrt{2}\epsilon \int_r^\infty \frac{[\int r'(h_2 + 2h\cot\theta)\,dr']}{r'^4} dr' - \frac{\epsilon l(u,\theta)}{3r^3}, \tag{110}$$

and

$$\begin{aligned}
V_1 = 1 + \sqrt{2}\epsilon \Bigg\{ &-\frac{1}{r^2}\int r(h_2 + 2h\cot\theta)_2\,dr - 4r\int_r^\infty \frac{[\int (h_2 + 2h\cot\theta)_2 r'\,dr']}{r'^4} dr' - \frac{1}{r^2}\cot\theta \int r(h_2 + 2h\cot\theta)\,dr \\
&+ 4r\cot\theta \int_r^\infty \frac{[\int (h_2 + 2h\cot\theta) r'\,dr']}{r'^4} dr' - 3\cot\theta \int_r^\infty \frac{h_2}{r'}dr' - \int_r^\infty \frac{h_{22}}{r'}dr' + 2\int_r^\infty \frac{h}{r'}dr' \Bigg\} \\
&+ [\text{plus terms with } \epsilon l(u,\theta)].
\end{aligned} \tag{111}$$

Also, from (5) it follows that

$$(r - 2M)(rg_1) + \frac{r^2}{2}(U\cot\theta - U_2) = 2\int r(g_0 r)_1\,dr + \epsilon\Theta(u,\theta), \tag{112}$$

where $\Theta(u,\theta)$ is an arbitrary function of its arguments.

Next, we shall assume that the function $h(u,\theta,r)$ may be expanded in a series of inverse powers of r. Thus

$$h(u,r,\theta) = -\sum_{n=1}^{\infty} \frac{\overset{(n)}{H}(u,\theta)}{\sqrt{2}r^n}, \tag{113}$$

then

$$g = \epsilon \sum_{n=1}^{\infty} \frac{\overset{(n)}{H}(u,\theta)}{nr^n}, \tag{114}$$

$$U = -\epsilon\left\{\frac{\left[\overset{(1)}{H} + 2\overset{(1)}{H}\cot\theta\right]}{r^2} + \frac{2}{3}\frac{\left[\overset{(2)}{H}_2 + 2\overset{(2)}{H}\cot\theta\right]}{r^3}\left(\frac{1}{3} + \ln r\right) - 2\sum_{n=3}^{\infty}\frac{\overset{(n)}{H}_2 + 2\overset{(n)}{H}\cot\theta}{(n+1)(n-2)r^{n+1}}\right\} - \frac{\epsilon l(u,\theta)}{3r^3}, \tag{115}$$

and, using (112),

$$\begin{aligned}
&\sum_{n=1}^{\infty}\frac{2M\overset{(n)}{H}}{r^n} - \sum_{n=1}^{\infty}\frac{\overset{(n)}{H}}{r^{n-1}} - \left\{\frac{\left[\overset{(1)}{H}_2 + 2\overset{(1)}{H}\cot\theta\right]\cot\theta}{2} - \frac{\left[\overset{(1)}{H}_2 + 2\overset{(1)}{H}\cot\theta\right]_2}{2}\right. \\
&\quad + \frac{1}{3r}\left(\frac{1}{3} + \ln r\right)\left[\left(\overset{(2)}{H}_2 + 2\overset{(2)}{H}\cot\theta\right)\cot\theta - \left(\overset{(2)}{H}_2 + 2\overset{(2)}{H}\cot\theta\right)_2\right] \\
&\quad \left. - \sum_{n=3}^{\infty}\frac{\left[\overset{(n)}{H}_2 + 2\overset{(n)}{H}\cot\theta\right]\cot\theta - \left[\overset{(n)}{H}_2 + 2\overset{(n)}{H}\cot\theta\right]_2}{(n+1)(n+2)r^{n-1}}\right\} \\
&\quad = \Theta(u,\theta) - \overset{(2)}{H}_0\ln r + 2\sum_{n=3}^{\infty}\frac{\overset{(n)}{H}_0(n-1)}{n(n-2)r^{n-2}} + \frac{1}{6r}[l(u,\theta)\cot\theta - l_2(u,\theta)].
\end{aligned} \tag{116}$$

From (6) it follows that $\overset{(2)}{H} \equiv 0$, then

$$\overset{(1)}{H} + \frac{1}{2}\left\{\left[\overset{(1)}{H}_2 + 2\overset{(1)}{H}\cot\theta\right]\cot\theta - \left[\overset{(1)}{H}_2 + 2\overset{(1)}{H}\cot\theta\right]_2\right\} = -\Theta(u,\theta), \tag{117}$$

$$M\overset{(1)}{H} = \frac{2}{3}\overset{(3)}{H}_0 - \frac{1}{6}(l\cot\theta - l_2), \tag{118}$$

$$-\overset{(3)}{H} - \frac{\left\{\left[\overset{(3)}{H}_2 + 2\overset{(3)}{H}\cot\theta\right]\cot\theta - \left[\overset{(3)}{H}_2 + 2\overset{(3)}{H}\cot\theta\right]_2\right\}}{20} = \frac{3}{4}\overset{(4)}{H}_0, \tag{119}$$

$$2M\overset{(n)}{H} - \overset{(n+1)}{H} - \frac{\left[\overset{(n+1)}{H}_2 + 2\overset{(n+1)}{H}\cot\theta\right]\cot\theta - \left[\overset{(n+1)}{H}_2 + 2\overset{(n+1)}{H}\cot\theta\right]_2}{(n+2)(n+3)} = \frac{2\overset{(n+2)}{H}_0(n+1)}{n(n+2)} \quad (n\geqslant 3). \tag{120}$$

Comparing the preceding results with (6)–(9), we get

$$\epsilon\overset{(1)}{H}(u,\theta) = c(u,\theta), \tag{121}$$

$$\epsilon\overset{(3)}{H}(u,\theta) = 3\,[C(u,\theta) - c^3/6] \tag{122}$$

or, keeping terms of order $O(\epsilon)$,

$$\epsilon\overset{(3)}{H} = 3C(u,\theta). \tag{123}$$

Also, from (118) and (119) one gets

$$Mc = 2C_0 + \tfrac{1}{6}(l\cot\theta - l_2)\epsilon \tag{124}$$

and

$$-4C + 5[C_2\cot\theta + 2C - C_{22}] = \epsilon\overset{(4)}{H}_0/3. \tag{125}$$

Next, comparing (115) with (8) we see that

$$N = -\epsilon l/6. \tag{126}$$

Thus (124) becomes

$$4C_0 = 2cM + N\cot\theta - N_2, \tag{127}$$

which is just Eq. (10) up to the order ϵ. The recurrence relation (120) gives the constraints for the time derivatives of the rest of the coefficients in (114). We have still the supplementatry conditions (11) and (12), which read (up to the order ϵ)

$$M_0 = \tfrac{1}{2}(c_{22} + 3c_2 \cot \theta - 2c)_0, \tag{128}$$

$$-3N_0 = M_2. \tag{129}$$

Observe that in general the function M may be written as

$$M = m_s + \epsilon \widetilde{m}, \tag{130}$$

where m_s is the Schwarzschild mass and N is of order ϵ as it follows from (126), then

$$\epsilon \widetilde{m}_0 = \tfrac{1}{2}(c_{22} + 3c_2 \cot \theta - 2c)_0, \tag{131}$$

$$-3N_0 = \epsilon \widetilde{m}_2. \tag{132}$$

The Bondi mass aspect

$$m = -\frac{1}{2} \int_0^\pi c_0^2 \sin \theta \, d\theta$$

is constant at order ϵ. This of course is the same situation as in the linear approximation.[1]

We may further introduce the quadrupole moment as[7,8]

$$Q(u) = 3n \int_0^\pi C P_2^2 (\cos \theta) \sin \theta \, d\theta, \tag{133}$$

where is a numerical factor and P_2^2 is the associated Legendre polynomial

$$P_2^2 = 3 \sin^2 \theta. \tag{134}$$

Then, using (127),

$$Q_{00} = 9n \int_0^\pi \left[\frac{1}{2} c_{00} m_s + N_{00} \cot \theta - N_{20} \right] \sin^3 \theta \, d\theta \tag{135}$$

or, using (132),

$$Q_{00} = 9n \int_0^\pi \left[\frac{1}{2} c_{00} \widetilde{m}_s - \frac{\widetilde{m}_{20} \cot \theta}{3} + \frac{\widetilde{m}_{22}}{3} \right] \sin^3 \theta \, d\theta. \tag{136}$$

We may now assume that c_0 may be expanded in Legendre polymonials in $\cos \theta$ with u dependent coefficients. Then, following the same line of argument as in Ref. 8, a relationship between the radiated energy and the change of quadrupole moments may be obtained (see Eq. V17 in Ref. 8).

IV. INITIALLY STATIC SYSTEMS

Let us now assume that, on a given null hypersurface $u = u_i = \text{const}$, a Bondi metric (general) coincides with some static solution (g_e, b_e, U_e, V_e). Then

$$g_i \equiv g(u_i, r, \theta) = g_e(r, \theta), \tag{137}$$

$$b_i \equiv b(u_i, r, \theta) = b_e(r, \theta), \tag{138}$$

$$U_i \equiv U(u_i, r, \theta) = U_e(r, \theta), \tag{139}$$

$$V_i \equiv V(u_i, r, \theta) = V_e(r, \theta), \tag{140}$$

and

$$g_{1i} \equiv g_1(u_i, r, \theta) = g_{e1}, \quad g_2(u_i, r, \theta) = g_{e2}, \tag{141}$$

$$b_{1i} \equiv b_1(u_i, r, \theta) = b_{e1}, \quad b_2(u_i, r, \theta) = b_{e2}, \tag{142}$$

$$U_{1i} \equiv U_1(u_i, r, \theta) = U_{e1}, \quad U_2(u_i, r, \theta) = U_{e2}, \tag{143}$$

$$V_{1i} \equiv V_1(u_i, r, \theta) = V_{e1}, \quad V_2(u_i, r, \theta) = V_{e2}. \tag{144}$$

Where the subscript e refers to the static solution and the subscript i refers to the general (radiative) solution evaluated on the hypersurface $u = u_i$. It follows from (137)–(144) that

$$\kappa_e = \epsilon_e = 0, \quad \rho_i = \rho_e, \quad \sigma_i = \sigma_e,$$
$$\alpha_i = \alpha_e, \quad \beta_e = \beta_i, \quad \tau_i = \tau_e,$$
$$\pi_i = \pi_e, \quad \mu_e = \mu_i, \quad \nu_i = \nu_e,$$

and

$$\gamma_i = \gamma_e - b_{0i}, \tag{145}$$

$$\lambda_i = \lambda_e + g_{0i}. \tag{146}$$

Since both the static (V_e, b_e, g_e, U_e) and the general (dynamic) metric (U, b, g, V) are solutions of the Einstein vacuum equations, then we get, from the vanishing of the Ricci scalars (16)–(22) at the hypersurface $u = u_i$,

$$\Phi_{20} = 0 \Rightarrow D g_{0i} = \rho_i \, g_{0i} \Rightarrow g_{0i} = c_{0i}/r, \tag{147}$$

$$\Phi_{11} = 0 \Rightarrow -D b_{0i} = -g_{0i} \sigma_i \Rightarrow \frac{\partial b_{0i}}{\partial r} = -\frac{c_{0i}\sqrt{2}}{r^2} e^{2bi} \overset{(5)}{F}_i, \tag{148}$$

$$\Phi_{12} = 0 \Rightarrow -\delta b_{0i} - \beta_{0i} = g_{0i} \alpha_i, \tag{149}$$

$$\phi_{22} = 0 \Rightarrow -\mu_{0i} = 2\lambda_e \, g_{0i} + g_{oi}^2 - 2b_{0i} \, \mu_i. \tag{150}$$

The vanishing of the other Ricci scalars is trivially satisfied.

Next, using (90) and (147), we get

$$2b_{0i} \overset{(5)}{F}_i + \overset{(5)}{F}_{0i} = -(c_{0i}/\sqrt{2}r) \, e^{-2bi}, \tag{151}$$

and we recall that $\overset{(5)}{F}_i = \overset{(5)}{F}_e$, $b_i = b_e$. Now, it is clear that if there are not "news" at $u = u_i$ (i.e., $c_{0i} = 0$), then $g_{0i} = 0$, $b_{0i} = 0$, $\beta_{0i} = \mu_{0i} = 0$. And it follows from (15) that $V_{0i} = U_{0i} = 0$. Let us now assume that c_0 vanishes not only at the hypersurface $u = u_i$, but in a finite interval $u_i \leqslant u \leqslant u_f$. Then, it is obvious from the results above that in this interval the solution will be static. Thus, if the metric coincides with a static metric at $u = u_i$, and is nonradiative $(c_0 = 0)$ for $u > u_i$, then it is also static for $u > u_i$.

We would like to mention that a similar result was found by Papapetrou[5,6] some years ago. There are, however, some differences in the conditions imposed, namely the following.

(a) The absence of radiation is defined by Papapetrou as $c_{00} = 0$, instead of $c_0 = 0$, as in our case.

(b) We do not require the space-time to be static below $u = u_i$, as is the case in Refs. 5 and 6, but just to coincide at $u = u_i$ with a static metric.

Finally we would like to remark that the conditions $c_{0i} = 0$ and $\overset{(5)}{F}_{0i} = 0$ are completely equivalent. In fact, if $c_{0i} = 0$ then $b_{0i} = 0$ and the vanishing of $\overset{(5)}{F}_{0i}$ follows from

(151). On the other hand if $\overset{(5)}{F}_{0i} = 0$, then it follows from (88), (90), and (147) that $c_{0i} = 0$.

V. CONCLUSIONS

We have seen so far that it is possible to reformulate the Bondi approach in terms of the functions measuring the deviation from spherical symmetry. The method seems to be specially suitable in the case of small perturbations off Schwarzschild. It is worth stressing the possibility of different alternatives in the choice of the vector field with respect to which the Lie derivatives of the metric tensor are calculated.

[1]H. Bondi, M. G. J. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[2]R. K. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962).
[3]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[4]L. Herrera and J. Jimenez, Phys. Rev. D **28**, 2987 (1983).
[5]A. Papapetrou, J. Math. Phys. **6**, 1405 (1965).
[6]I. Moret-Bailly and A. Papapetrou, Ann. Inst. H. Poincaré **6**, 205 (1967).
[7]E. T. Newman and T. W. Unti, J. Math. Phys. **6**, 1806 (1965).
[8]R. Berezdivin and L. Herrera, J. Math. Phys. **18**, 418 (1977).

Dual mass, $\mathscr{H}$-spaces, self-dual gauge connections, and nonlinear gravitons with topological origin

Anne Magnon[a)]

Physics Department, University of Syracuse, Syracuse, New York 13244-1130 and Département de Mathématiques, Université de Clermont-Fd. 63170 Aubière, France

(Received 12 December 1985; accepted for publication 6 March 1986)

An analogy between source-free, asymptotically Taub–NUT magnetic monopole solutions to Einstein's equation and self-(anti-self-) dual gauge connections is displayed, which finds its origin in the first Chern class of these space-times. A definition of asymptotic graviton modes is proposed that suggests that a subclass of Penrose's nonlinear gravitons or Newman's $\mathscr{H}$-spaces could emerge from nontrivial space-time topologies.

I. INTRODUCTION

Most approaches, presently available in the quantum theory of interacting fields, are the perturbative ones. Although the importance of nonperturbative effects has gradually been acknowledged, difficulties paving the way to an exact theory seem to be so formidable that other methods had to be substituted. Among those, semiclassical methods have been most helpful. In particular, ideas connected with Feynman path integrals have led to techniques of Euclideanization and a lot of effort has been devoted to the investigation of Yang–Mills fields in Euclidean space and solutions such as instantons, the viewpoint here being that such solutions signal physically interesting quantum processes, which escape perturbative descriptions.

In the case of quantum gravity, the issue is even less settled. Due to the absence of a preferred time, what is meant by Euclideanization is not *a priori* clear. For instance, one can decide to complexify Lorentzian solutions of Einstein's equation and investigate their Euclidean sections. However, such sections are relatively scarce. Another possibility is to search for Euclidean solutions in their own right, expecting that some of them will carry an interesting physical interpretation. This has led to the notion of gravitational instanton: black-hole instantons describing states of thermal equilibrium of the "quantized gravitational field," locally asymptotically flat instantons signaling tunneling processes, and compact instantons contributing to the space-time foam picture. This in turn has developed Wheeler's[1] viewpoint according to which a very large fluctuation of the metric and even of the short scale space-time topology should be expected in quantum gravity. The reason is that the action for the gravitational field is not scale invariant, unlike that of the Yang–Mills or electromagnetic fields. Consequently, a large fluctuation of the metric over a short length scale does not have a very large action and therefore exhibits a small damping in the path integral.

Since the path integral approach is a convenient way to handle nontrivial topologies, much attention has been devoted to solutions that are expected to have a dominant contribution to the path integral, the hope being that instanton solutions with complicated topologies would be such solutions, i.e., metrics near stationary phase points of the action, and much attention has been devoted to the investigation and classification of gravitational instantons, based on their topological invariants such as the Euler number χ and signature τ.

On another hand, since stationary points of the action are provided, in the case of Yang–Mills fields by self- (anti-self-) dual (bundle) connections over Euclidean four-space, one expects that self- (anti-self-) dual (complex) solutions to Einstein's equation should play an important role in the quantum gravity program. Penrose's nonlinear graviton construction via twistorial methods and Newman's $\mathscr{H}$-spaces theory have aimed at developing such considerations.

In this paper we would like to strengthen this viewpoint. We shall take advantage of the existence of a particular class of Lorentzian source-free solutions to Einstein's equation, the asymptotically NUT gravitational magnetic monopoles, to propose that (a subclass of) Penrose's nonlinear gravitons could be viewed (at least asymptotically) as self- (anti-self-) dual gauge connections on a suitable bundle, provided nontrivial topological features of the space-time manifold are incorporated in their description. This result will be obtained as follows. In a previous series of papers[2–4] we have presented a general framework for the description of the geometry and asymptotic behavior of (real Lorentzian) space-times in the presence of magnetic mass. An important feature lies in the fact that these space-times exhibit the structure of a nontrivial U(1) bundle over a base space with nontrivial second cohomology group. This in turn, implies the existence of a bundle connection one-form at infinity, analogous to the Maxwell connection, the flux of the corresponding curvature two-form being a measurement of the enclosed magnetic charge (i.e., the first Chern class of the bundle). This charge (an integral over a two-sphere surrounding the nontrivial topological features) is purely topological. Hence the situation for gravitational magnetic monopoles is analogous to that of the electromagnetic monopoles. Restricting ourselves to (Lorentzian) asymptotically NUT magnetic monopoles that are real sections of a complex (right- or left-flat) asymptotically Taub–NUT solution, one can show the existence (in the neighborhood of their complexified conformal null boundary) of asymptotically self- (anti-self-) dual bundle curvature two-forms and connections, implying the existence of integrable propaga-

[a)] Détachée du Ministère des Relations Extérieures, Paris, France.

 0022-2488/86/082105-06$02.50

tion laws along specific complex two-planes. Thus, under such conditions, the right- or left-flat Taub–NUT complex solutions proposed by Penrose as an illustration of his nonlinear graviton construction provide an example of space-times that are in many respects similar to the self-dual Yang–Mills gauge fields (obtained, i.e., via Ward's generating procedure). Since the existence of our connections relies crucially on the nonvanishing of the first Chern class of the Lorentzian slices, we propose to associate them with graviton modes originating within the space-time nontrivial topology. Since nontrivial topologies provide new quantum numbers, via superselected photonic sectors, which are analogous to spin and mass, in the quantization of Maxwell fields, these results could be interpreted as a generalization in the case of gravity: existence of "superselected self- (anti-self-) dual bundle connections." A unification with the already available[5] scheme of quantization at null infinity (based on the description of asymptotic gravitational degrees of freedom via equivalence classes of metric connections) is possible. We hope that these considerations will bring support to the viewpoint according to which nontrivial topologies might have a crucial role to play in the unification of quantization methods available for Yang–Mills fields and for gravity.

II. PRELIMINARIES

In this section we would like to briefly review the notion of self-dual connection in the context of principal bundles or vector bundles over Riemannian manifolds. The description of Yang–Mills fields is based on such a framework. Since the following sections aim at displaying an analogy between gravity (in the complex left- or right-flat regime) and Yang–Mills gauge fields, this section will serve as a mathematical introduction.

Let Λ^p denote the bundle of exterior p-forms over $\mathscr{M}$, an oriented Riemannian manifold of even dimension 21, and $\Gamma(\Lambda^p)$ its space of smooth sections.

On a principal G-bundle $\mathscr{B}$ over $\mathscr{M}$, a connection is a one-form ω, valued in the Lie algebra g of G, and its curvature Ω is the g-valued two-form

$$d\omega + \tfrac{1}{2}[\omega,\omega], \tag{1}$$

a section of $\tilde{g} \otimes \Lambda^2$, $\tilde{g}$ denoting the vector bundle associated to $\mathscr{B}$ by the adjoint representation. On a vector bundle $\mathscr{V}$ over $\mathscr{M}$, a connection is defined by its covariant derivative ∇, a first-order linear differential operator

$$\nabla\colon\ \Gamma(\mathscr{V}\otimes\Lambda^0)\to\Gamma(\mathscr{V}\otimes\Lambda^1). \tag{2}$$

This covariant derivative ∇ has a natural extension D to $\Gamma(\mathscr{V}\otimes\Lambda^1)$ defined by

$$D(\alpha_0\otimes\alpha_1) = \nabla\alpha_0\wedge\alpha_1 + \alpha_0\otimes d\alpha_1, \tag{3}$$

where $\alpha_0\in\Gamma(\mathscr{V}\otimes\Lambda^0)$ and $\alpha_1\in\Gamma(\mathscr{V}\otimes\Lambda^1)$. The curvature Ω is defined as the composition $D\nabla\in\Gamma(\mathrm{End}\ \mathscr{V}\otimes\Lambda^2)$. The relation between ω and ∇ can be described as follows. Any representation $\mathscr{V}_G$ of G on a vector space $\mathscr{V}$ induces a particular local basis $\{e_i\}$ of $\mathscr{V}$, a local section of $\mathscr{B}\times\mathscr{V}_G$. The pullback of ω via this section induces a matrix of one-forms ω_{ij}, the resulting action of ∇ being defined via

$$\nabla e_i = \sum \omega_{ij}\otimes e_j. \tag{4}$$

Conversely, if $\mathscr{V}$ is a ∇-preserving representation of G, the above formula defines ω on the principal bundle of G-frames. This ω is also called the gauge potential, and its curvature Ω the gauge field.

Definition 1: A gauge transformation on a principal G-bundle $\mathscr{B}$ is a diffeomorphism f: $\mathscr{B}\to\mathscr{B}$ such that (i) $f(g\cdot p) = gf(p)$, $p\in\mathscr{B}$, $g\in G$, and (ii) f preserves each fiber.

Under such a mapping, the connection one-form ω transforms according to $f^{-1}\omega = f^{-1}\,df + (\mathrm{Ad}\,f^{-1})\omega$, and the covariant derivative according to $f^{-1}\nabla = f^{-1}\nabla f$.

Definition 2: A connection is said to be self-dual if its curvature Ω satisfies $\Omega = {}^*\Omega$, and anti-self-dual if $\Omega = -{}^*\Omega$.

Self- (anti-self-) duality is invariant under the action of conformal transformations on the base space. In the particular case of a U(1) bundle, Ω is a closed two-form defining the first Chern class in $H^2(\mathscr{M},R)$.

It can be shown[6] that a bundle with self-dual connection must satisfy some topological properties. In particular the first Chern class and Pontrjagin class (evaluated on the fundamental cycle on a compact base manifold) must be positive.

Recall also that a form

$$\Omega = \omega^{(2,0)}{}_{ab}\,dz^a\wedge dz^b + \omega^{(1,1)}{}_{ab}\,dz^a\wedge d\bar{z}^b + \omega^{(0,2)}{}_{ab}\,d\bar{z}^a\wedge d\bar{z}^b$$

is of type (1,1) on C^2 if $\omega^{(2,0)} = \omega^{(0,2)} = 0$. If Ω is of type (1,1) for all possible such complex structures on R^4, then Ω is anti-self-dual; one has[6] the following theorem.

Theorem: If $\mathscr{M}$ is a complex manifold modeled on a four-dimensional Riemannian manifold, and $\mathscr{B}$ is a C^∞ principal bundle with connection ∇ whose curvature Ω is of type (1,1), $\mathscr{B}$ admits a natural holomorphic structure, ∇ being the (unique) corresponding Hermitian connection.

Remarks:

(i) Using the complex structure on $\mathscr{M}$ and complexified fiber G^C in the case of a compact structure group, one can get an almost complex structure on the principal bundle $\mathscr{B}$. Integrability of this almost complex structure follows from self-duality of the conformal curvature tensor on the base manifold $\mathscr{M}$.

(ii) The above theorem will be useful in the forthcoming sections. More precisely we shall focus on a particular class of Lorentzian solutions to Einstein's equation which admits a conformal null boundary $\mathscr{I}$, a nontrivial U(1) bundle over S^2. Assuming that such solutions are real sections of self-dual complex solutions (left-flat complex, holomorphic, nonsingular, invertible metrics g_{ab} on a four-dimensional complex manifold M, in the sense of Refs. 7 and 8), it will be useful to assume that $\mathbb{C}\mathscr{I}$ (a complex thickening of $\mathscr{I}$) inherits a natural, almost complex structure from its bundle structure.

III. ASYMPTOTICALLY NUT LORENTZIAN MAGNETIC MONOPOLES AND NULL REGIME

In this section we shall adopt the notations of Ref. 9. In view of our investigation (Sec. IV) of the complex right-

(left-) flat regimes, we summarize from Refs. 3 and 4 the situation at null infinity for Lorentzian asymptotically NUT magnetic monopole solutions. Let $(M, g_{ab} = \Omega^2 \hat{g}_{ab})$ denote the (unphysical) space-time obtained from such a space-time $(\hat{M}, \hat{g}_{ab})$ after conformal completion. Since the unphysical Weyl tensor $C_{abcd} = \Omega K_{abcd}$ vanishes at $\mathscr{I}$, we have

$$R_{abcd} = g_{a[c} S_{d]b} - g_{b[c} S_{d]a} \tag{5}$$

with

$$S_{ab} = R_{ab} - \tfrac{1}{6} R g_{ab}. \tag{6}$$

Let i^* denote the pullback operation at $\mathscr{I}$. The two basic fields on $\mathscr{I}$ are $\mathbf{n}^b = i^*(g^{ab} \nabla_a \Omega)$ and $\mathbf{g}_{ab} = i^* g_{ab}$ (the degenerate metric). Introducing a Newman–Penrose null tetrad field in the neighborhood of $\mathscr{I}$ $(n^a, m^a, \overline{m}^a, l^a)$, and its pullback $i^* n^a = \mathbf{n}^a$, $i^* m^a = \mathbf{m}^a$, $i^* \overline{m}^a = \overline{\mathbf{m}}^a$, $i^* l_a = \mathbf{l}_a$, $i^* m_a = \mathbf{m}_a$, $i^* \overline{m}_a = \overline{\mathbf{m}}_a$, $i^* n_a = 0$, one can define $\gamma_{ab} = D_a \mathbf{l}_b$, and obtain the following expressions:

$$\gamma_{ab} = \sigma^0{}_{ab} + \tfrac{1}{2} \mathbf{g}_{ab} \mathbf{g}^{mn} \gamma_{mn}, \tag{7}$$

where $\sigma^0{}_{ab} = \bar{\sigma}^0 \mathbf{m}_a \mathbf{m}_b + \sigma^0 \overline{\mathbf{m}}_a \overline{\mathbf{m}}_b$ (σ^0 is known as the asymptotic shear),

$$S_{ab} = -2\dot{\sigma}^0{}_{ab} + \tfrac{1}{2} \mathbf{g}_{ab} \mathscr{R} \tag{8}$$

(where $\mathscr{R}$ is the scalar curvature of the manifold of orbits of $\mathbf{n}^a$), and

$$\mathbf{N}_{ab} = \mathbf{S}_{ab} - \boldsymbol{\rho}_{ab}, \tag{9}$$

the symmetric traceless News tensor, invariant under conformal rescalings. One can further introduce the "electric" and "magnetic" components of the (rescaled) Weyl tensor,

$$K^{ab} = -K^{ambn} n_m n_n, \quad {}^*K^{ab} = -{}^*K^{ambn} n_m n_n, \tag{10}$$

and their pullbacks $\mathbf{K}^{ab}, {}^*\mathbf{K}^{ab}$, with the following properties:

$$D_a {}^*\mathbf{K}^{ab} = 0, \quad D_a \mathbf{K}^{ab} = 0, \tag{11}$$

$$\mathbf{g}_{ma} \mathbf{K}^{mb} = -\epsilon_{amp} \mathbf{n}^p {}^*\mathbf{K}^{mb}, \tag{12}$$

$$\mathbf{g}_{ma} {}^*\mathbf{K}^{mb} = \epsilon_{amp} \mathbf{n}^p \mathbf{K}^{mb}, \tag{13}$$

where $\epsilon_{abc} = i^*(\epsilon_{abcd} l^d)$ and $\epsilon^{abc} = i^*(\epsilon^{abcd} n_d)$. The "magnetic" component ${}^*\mathbf{K}^{ab}$ is going to play a crucial role in the forthcoming section. Note first that if a stationary Killing vector field $\hat{\xi}^a$ is available on $(\hat{M}, \hat{g}_{ab}, \hat{\nabla}_a)$, we have shown[3] that

$$i^* \hat{\nabla}_{[a} \hat{\lambda}^{-1} \hat{\xi}_{b]} = D_{[a} \mathbf{w}_{b]} \tag{14}$$

(where $\hat{g}_{ab} \hat{\xi}^a \hat{\xi}^b = -\hat{\lambda}$), can be identified with

$$D_{[a} \mathbf{S}_{b]}{}^c \mathbf{l}_c \equiv D_{[a} \mathbf{v}_{b]}, \tag{15}$$

a constant multiple of $\frac{1}{4} \epsilon_{abm} {}^*\mathbf{K}^{cm} \mathbf{l}_c$. This leads us to introduce $\Omega_{ab} = \epsilon_{abm} {}^*\mathbf{K}^{cm} \mathbf{l}_c$, a closed two-form on $\mathscr{I}$. The related connection one-form $\mathbf{v}_b$ can be viewed as a "Maxwell" connection induced at $\mathscr{I}$ by $\hat{\lambda}^{-1} \hat{\xi}_b$ [recall that in the presence of nonvanishing magnetic mass, i.e.,

$$\int_{S^2} \hat{\nabla}_{[a} \hat{\lambda}^{-1} \hat{\xi}_{b]} \, dS^{ab} \neq 0,$$

$\mathscr{I}$ is a nontrivial U(1) bundle over S, the two-sphere of its null generators]. The presence of a magnetic monopole is related to that of a wire singularity on $(\hat{M}, \hat{g}_{ab})$, which originates in this monopole and registers at infinity as follows: $D_{[a} \mathbf{v}_{b]}$ is the lift of a closed (not globally exact) two-form $D_{[a} \nu_{b]}$ on the base space S; the discontinuity in ν_b is the imprint, at infinity, of the wire singularity.

In absence of isometry, on another hand, we have shown (3) that the total magnetic mass $N_{\mathscr{I}}$ (an integral over a cross section C of $\mathscr{I}$) is such that

$$N_{\mathscr{I}} = \int_C \Omega_{ab} \, dS^{ab} \equiv \int_S D_{[a} \nu_{b]} \, dS^{ab}$$

and is a measurement of the number of times the (nontrivial) $\mathscr{I}$ bundle winds around its S^1 fiber. It is thus clear that, in the absence of isometry $\hat{\xi}^a$, the presence of a nontrivial U(1) bundle structure at $\mathscr{I}$ is a characteristic feature of Lorentzian asymptotically NUT dual mass gravitational monopoles.

Theorem: Source-free, asymptotically NUT Lorentzian solutions to Einstein's equation are nontrivial S^1 bundles over a base space with nontrivial second cohomology group. Their magnetic mass is a purely topological charge, a measurement of the flux of the curvature two-form on the nontrivial compact $\mathscr{I}$ bundle (or equivalently the first Chern class of the space-time).

Corollary: The presence of nonvanishing magnetic mass induces transition functions at null infinity reflecting the nontriviality of the $\mathscr{I}$ bundle (lens space) over its S^2 base space, or equivalently the degree of mappings

$$g: S^2 \to \Pi_2(S^2).$$

IV. COMPLEX REGIME AND ASYMPTOTICALLY (ANTI-) SELF-DUAL BUNDLE CONNECTIONS WITH TOPOLOGICAL ORIGIN

Recall that the self-dual Taub–NUT instanton can be written in the form

$$\begin{aligned} ds^2 = {} & (r-n)(r+n)^{-1}(d\tau + 2n \cos\theta \, d\phi)^2 \\ & + (r^2 - n^2)(d\theta^2 + \sin^2\theta \, d\phi^2) \\ & + (r+n)(r-n)^{-1} dr^2. \end{aligned} \tag{16}$$

The Dirac string singularity at the north pole ($\theta = 0$) can be removed by introducing a new coordinate

$$\tau' = \tau + 2n\phi. \tag{17}$$

Similarly the Dirac string singularity at the south pole ($\theta = \pi$) can be removed by introducing a new coordinate

$$\tau'' = \tau - 2n\phi. \tag{18}$$

Because ϕ is defined as modulo 2π, τ', and τ'' must be identified as modulo $8\pi n$. These identifications give the surfaces $r > n$ the topology of three-spheres with $(\tau(2n)^{-1}, \phi)$ as Euler coordinates. Due to presence of the Killing vector field $K^a = \partial/\partial\tau$, one can relate the self-duality of the Riemann curvature $(R_{abcd} = \frac{1}{2} e_{abef} R^{ef}{}_{cd})$ to that of the two-form $K_{ab} = \nabla_{[a} K_{b]}$. The self-dual (anti-self-dual) parts of K_{ab} are defined via $K^{\pm}{}_{ab} = \frac{1}{2}(K_{ab} \pm \frac{1}{2} e_{abef} K^{ef})$, and (provided the curvature is itself self-dual) one can show that K_{ab} is self-dual everywhere if it is self-dual at one point. Appearance and disappearance of pairs of self- and anti-self-dual instantons have been proposed to describe quantum fluctuations of the metric. As we shall see now, the right- (left-) flat complex regime, though probably more realistic, is not so straightforward.

Let us now focus on source-free asymptotically NUT Lorentzian gravitational magnetic monopoles. We assume that such solutions are (at least asymptotically) real sections of asymptotically left- (right-) flat complex space-times in the sense of Refs. 7, 8, and 10.

On one hand, such Lorentzian slices exhibit an interesting bundle structure [a nontrivial U(1) bundle over a base space with a nonvanishing second homology group]. From the previous section we know that this bundle structure arises from the nonvanishing of the first Chern class, or equivalently the presence of nonvanishing magnetic mass (i.e., gravitational magnetic monopole). On another hand, left- (right-) flat complex solutions to Einstein's equation have been introduced by Newman[10] ($\mathscr{H}$-spaces) and Penrose[7,8] (definition of the nonlinear graviton within the framework of deformed twistor space) as possible candidates for the description of one particle states in a future theory of quantum gravity. One expects that (in the Euclidean regime) such solutions would provide extrema for the path integral functional, and therefore play a role similar to that of self-dual Yang–Mills gauge fields. In agreement with the results presented in Ref. 10, we shall assume that a complex solution can be defined, at least asymptotically on a complex manifold modeled on the real manifold associated to the Lorentzian (magnetic monopole) solutions under consideration. In a suitable complex neighborhood of infinity we shall combine the bundle structure (arising from the Lorentzian regime) with the self-dual features of the complexified solution. The result will be summarized in our ability to select a family of (anti-) self-dual bundle gauge connections that find their origin in the first Chern class of these space-times. An analogy between gravity and Yang–Mills fields (Sec. III) will consequently be underlined that does not require the existence of Euclidean (instantonic) sections or the introduction of bundles over four-dimensional Riemannian spaces. A definition of graviton–antigraviton pairs that find their origin in the space-time topology will be proposed.

Let us briefly recall the definition of (right-) left-flat space-times. Denote by $(\widetilde{M},\tilde{g}_{ab})$ a four-dimensional complex manifold with a complex, holomorphic, nonsingular invertible metric $\tilde{g}_{ab}$. Vacuum equation $\widetilde{R}_{ab}=0$ will be assumed, thus implying the following decomposition of the Weyl curvature tensor into its irreducible (spinor) components:

$$C_{abcd}=\psi_{ABCD}\epsilon_{A'B'}\epsilon_{C'D'}+\tilde{\psi}_{A'B'C'D'}\epsilon_{AB}\epsilon_{CD}, \tag{19}$$

with

$$\psi_{ABCD}=\psi_{(ABCD)},\quad \tilde{\psi}_{A'B'C'D'}=\tilde{\psi}_{(A'B'C'D')}; \tag{20}$$

ψ_{ABCD} ($\tilde{\psi}_{A'B'C'D'}$) corresponds to the anti-self-dual (resp. self-dual) part of the conformal curvature:

$$C^-{}_{abcd}=\psi_{ABCD}\epsilon_{A'B'}\epsilon_{C'D'}, \tag{21}$$

$$C^+{}_{abcd}=\tilde{\psi}_{A'B'C'D'}\epsilon_{AB}\epsilon_{CD}, \tag{22}$$

$$(C^\pm{}_{abcd})^*=e_{ab}{}^{ef}C^\pm{}_{efcd}=\pm iC^\pm{}_{abcd}, \tag{23}$$

$$\overline{C}^+{}_{abcd}=C^-{}_{abcd}. \tag{24}$$

A space-time is right (resp. left) conformally flat if $\tilde{\psi}_{A'B'C'D'}=0$ (resp. $\psi_{ABCD}=0$), and is thus essentially complex. We shall further assume (in the asymptotic region of the Lorentzian section) the existence of a rescaled (real) curvature tensor K_{abcd} subject to

$$C^-{}_{abcd}+C^+{}_{abcd} \equiv C_{abcd}=\Omega K_{abcd}=\Omega(K^-{}_{abcd}+K^+{}_{abcd}). \tag{25}$$

Recall also[10,11] the existence of an adapted complex asymptotic chart (z^a, $a=0,1,2,3$) suitable for the introduction of a complexified conformal null boundary $\mathbb{C}\mathscr{I}$ (in the sense of Penrose) and of a complex function Z (the good-cut function) enabling the definition of a field of null tetrads adapted to the asymptotic behavior of the complex solution[9–11]:

$$L_a=Z_a=\partial_a Z=O_AO_{A'}, \tag{26}$$

$$M_a=M_a(Z)=O_AI_{A'}, \tag{27}$$

$$\widetilde{M}_a=\widetilde{M}_a(Z)=I_AO_{A'}, \tag{28}$$

$$N_a=I_AI_{A'}, \tag{29}$$

with the normalizations $O_AI^A=O_{A'}I^{A'}=1$. Such right- (left-) flat complex solutions have emerged from the theory of $\mathscr{H}$-spaces[10] and can be viewed as a thickened region (not necessarily unique) of their Lorentzian section(s). The following expression of the Levi-Cività tensor will be useful:

$$e_{abcd}=(i/2)(\epsilon_{AC}\epsilon_{BD}\epsilon_{A'D'}\epsilon_{B'C'}-\epsilon_{AD}\epsilon_{BC}\epsilon_{A'C'}\epsilon_{B'D'}). \tag{30}$$

The resulting three-tensors admitting pullbacks at $\mathbb{C}\mathscr{I}$ are

$$e_{abc}=e_{abcd}L^d=(i/2)(O_BO_{A'}\epsilon_{AC}\epsilon_{B'C'}-O_{B'}O_A\epsilon_{A'C'}\epsilon_{BC}), \tag{31}$$

$$e^{abc}=e^{abcd}N_d=(i/2)(I^BI^{A'}\epsilon^{AC}\epsilon^{B'C'}-I^{B'}I^A\epsilon^{A'C'}\epsilon^{BC}). \tag{32}$$

Definition: The curvature two-form (on the $\mathscr{I}$ bundle), which induces (via its flux) the total magnetic mass of the gravitational monopole, defines a two-flat

$$\Omega^*{}_{ab}=i^*(e_{abm}{}^*K^{mc}L_c). \tag{33}$$

We shall refer to $\Omega^*{}_{ab}$ as being the magnetic two-flat. The corresponding complex two-form in the left-flat (complex) solution is

$$\begin{aligned}\Omega^{*-}{}_{ab}&=-ie_{abm}K^{-imjc}N_iN_jL_c\\&=\tfrac{1}{2}(O_BO_{A'}I_A\epsilon_{B'M'}-O_{B'}O_AI_B\epsilon_{A'M'})\tilde{\psi}^{I'M'J'C'}I_{I'}I_{J'}O_{C'}\\&=\Omega^-{}_{AA'BB'}.\end{aligned} \tag{34}$$

Similarly, the two-form associated to the total mass of the solution is given by

$$\begin{aligned}\Omega^-{}_{ab}&=e_{abm}K^{-imjc}N_iN_jL_c\\&=e_{abm}K^{-mc}L_c=-i\Omega^-{}_{AA'BB'}.\end{aligned} \tag{35}$$

Hence the following theorem.

Theorem 1: The mass M and magnetic mass M^* of a left- (right-) flat complex solution are related via

$$M^*=-iM\quad (M^*=iM).$$

Remark: The two-form $\Omega^{*-}{}_{ab}$ is not self- (or anti-self-) dual. This follows immediately from

$$e_{ab}{}^{cd}\Omega^{*-}{}_{cd} = i(O_B O_{B'} I_A \epsilon_{A'M'} - O_A O_{A'} I_B \epsilon_{B'M'}) \times \tilde{\psi}^{I'M'J'L'} I_{I'} I_{J'} O_{L'}, \tag{36}$$

a two-form perpendicular to the two-flat $M_{[a}\widetilde{M}_{b]}$, while $\Omega^{*-}{}_{ab}$ is perpendicular to $N_{[a}L_{b]}$.

We shall now introduce the self- (anti-self-) dual components of $\Omega^{*-}{}_{ab}$ via projection on suitable self- (anti-self-) dual two-flats.

Theorem 2: The complex two-flats $F^{n+}{}_{AA'BB'}$ ($n = 1,2,3$), respectively, defined via

$$O_{A'}O_{B'}\epsilon_{AB}, \quad I_{A'}I_{B'}\epsilon_{AB}, \quad O_{(A'}I_{B')}\epsilon_{AB},$$

are self-dual

$[e_{CC'DD'}{}^{AA'BB'}F^{n+}{}_{AA'BB'} = iF^{n+}{}_{CC'DD'} \quad (n = 1,2,3,)]$

while their primed analogs, $\widetilde{F}^{n-}{}_{AA'BB'}$ ($n = 1,2,3$), defined via

$$O_A O_B \epsilon_{A'B'}, \quad I_A I_B \epsilon_{A'B'}, \quad O_{(A}I_{B)}\epsilon_{A'B'},$$

are anti-self-dual

$(\epsilon_{CC'DD'}{}^{AA'BB'}\widetilde{F}^{n-}{}_{AA'BB'} = -i\widetilde{F}^{n-}{}_{CC'DD'})$.

The proof follows immediately from the expression of $e_{abcd} \equiv e_{AA'BB'CC'DD'}$ [formula (30)]; antisymmetry is obvious.

Theorem 3: The two-flats

$\omega^{n+}{}_{AA'BB'} \equiv \Omega^{*-}{}_{[AA'MM'}F^{n+MM'}{}_{BB']}$

(resp. $\widetilde{\omega}^{n-}{}_{AA'BB'} \equiv \Omega^{*-}{}_{[AA'MM'}\widetilde{F}^{n-MM'}{}_{BB']}$)

are self-dual (anti-self-dual) ($n = 1,2$). Their respective expressions are

$$\omega^{1+}{}_{AA'BB'} = \tilde{\psi}^{M'T'N'K'} I_{M'} I_{N'} O_{T'} O_{K'} L_{[b}\widetilde{M}_{a]}, \tag{37}$$

$$\omega^{2+}{}_{AA'BB'} = -\tfrac{1}{2}\tilde{\psi}^{M'T'N'K'} I_{M'} I_{N'} O_{K'} I_B I_{B'} O_A \epsilon_{A'T'} \tag{38}$$

$$= \tilde{\psi}^{M'T'N'K'} I_{M'} I_{N'} O_{T'} O_{K'} N_{[a}M_{b]}, \tag{39}$$

$$\widetilde{\omega}^{1-}{}_{AA'BB'} = \tilde{\psi}^{M'T'N'K'} I_{M'} I_{N'} O_{T'} O_{K'} L_{[b}M_{a]}, \tag{40}$$

$$\widetilde{\omega}^{2-}{}_{AA'BB'} = \tilde{\psi}^{M'T'N'K'} I_{M'} I_{N'} O_{T'} O_{K'} N_{[a}\widetilde{M}_{b]}. \tag{41}$$

The proof is straightforward.

Corollary: The self- (anti-self-) dual components $\omega^{\pm}{}_{AA'BB'}$ of the magnetic two-flat vanish when contracted with an anti-self- (self-) dual two-flat F^{n-} (F^{n+}).

Theorem 4: The two-forms $\omega^{n\pm}{}_{AA'BB'}$ ($n = 1,2$) are self- (anti-self-) dual restrictions of the magnetic two-flat $\Omega^{*-}{}_{ab}$ to totally null two-planes.

Proof: It suffices to notice that the null vectors L_b and $\widetilde{M}_b$ can be spanned by the two-plane-forming family of null vectors $O_{B'}\lambda_B$ (O_B fixed, λ_B varying from O_B to I_B), L_b and M_b by the family of null vectors $O_B\lambda_{B'}$ (O_B fixed, $\lambda_{B'}$ varying from $O_{B'}$ to $I_{B'}$), N_b and M_b by the family of null vectors $I_{B'}\lambda_B$ ($I_{B'}$ fixed, λ_B varying from I_B to O_B), and N_b, $\widetilde{M}_b$ by the family of null vectors $I_B\lambda_{B'}$, ($\lambda_{B'}$ varying from $I_{B'}$ to $O_{B'}$).

Theorem 5: The forms $\omega^{1\pm}{}_{AA'BB'}$ admit a complex gauge potential

$\omega^{1\pm}{}_{AA'BB'} = \nabla_{[AA'}A^{\pm}{}_{BB']}$.

This follows from the fact that $\Omega^{*}{}_{ab}$ admits a gauge potential ν_b on $\mathscr{I}$, ($\Omega^{*}{}_{ab} = \nabla_{[a}\nu_{b]}$), and from its spinor expression

$\phi_{(AB)}\epsilon_{A'B'} + \tilde{\phi}_{A'B'}\epsilon_{AB}$.

From $\omega^{+}{}_{AA'BB'} = \phi_{(AB)}\epsilon_{A'B'}$ with $\nabla^{AB'}\phi_{AB} = 0$ (resp. $\widetilde{\omega}^{-}{}_{AA'BB'} = \tilde{\phi}_{(A'B')}\epsilon_{AB}$ with $\nabla^{A'B}\tilde{\phi}_{A'B'} = 0$) one concludes the existence of a gauge potential $A^{+}{}_{BB'}$ (resp. $A^{-}{}_{BB'}$) (since $i^*N_a = 0$ we shall not, from now onwards, be concerned by the forms ω^2).

Corollary: In a suitable neighborhood of $\mathbb{C}\mathscr{I}$, $A^{+}{}_{BB'}$ (resp. $A^{-}{}_{BB'}$) defines an integrable propagation on the complex nontrivial space-time bundle provided one chooses a closed path γ that lies within an anti-self-dual (resp. self-dual) complex totally null two-plane and which does not cross singularities of this potential. (The proof is straightforward since along such paths, one has

$$\int_\gamma A_b \, dS^b = 0.)$$

Remarks:

(i) Recall that for Yang–Mills fields (over a base space M) with n internal degrees of freedom (a_i, $i = 1,...,n$) the curvature two-form F is derived from the formula

$$\nabla_{[\alpha}\nabla_{\beta]}t^a{}_\gamma = \tfrac{1}{2}(R_{\alpha\beta\gamma}{}^{\delta}\delta^a{}_b + F_{\alpha\beta}{}^a{}_b\delta_\gamma{}^\delta)t^b{}_\delta. \tag{42}$$

The curvature two-forms $\omega^{\pm}{}_{AA'BB'}$ can be viewed as gravitational analogs of F_{ab}, the internal gauge degrees of freedom being provided by the totally null directions displayed in the proof of Theorem 4. As will be mentioned in our concluding remarks, such internal degrees can be viewed as "transverse" to the radiative degrees of freedom usually related to the presence of the BMS group at $\mathscr{I}$.

(ii) In the absence of magnetic mass, the (complex) space-time bundle is trivial, and the analogy with non-Abelian Yang–Mills self-dual gauge connections should be revised; also, if $C_1(\Omega^{*}{}_{ab}) = 0$ (no magnetic monopole) an analysis of the origin of singularities in the gauge potential is required.

(iii) We propose that pairs of self- (anti-self-) dual gauge connections ($A^{+}{}_{BB'}$, $A^{-}{}_{BB'}$) could describe graviton–antigraviton modes emerging from the nonvanishing of the first Chern class of the Lorentzian (gravitational) magnetic monopole. The onset of nontrivial topologies in the Lorentzian regime could thus be described, within this framework, via creation of graviton–antigraviton pairs; in strong analogy with the instantonic model presented in the beginning of this section, clouds of such pairs could be associated with the "space-time foam."

V. CONCLUDING REMARKS

(i) Recall that the radiative degrees of freedom can be described at $\mathscr{I}$ via equivalence classes of metric connections D's. It is easy to show that connections D, $\widetilde{D}$ corresponding to a given physical space-time ($\widehat{M}, \hat{g}_{ab}$) must be related via

$$(D_a - \widetilde{D}_a)K_b = f(K_c n^c)g_{ab}, \tag{43}$$

for some function f on $\mathscr{I}$. Two such connections are said to be equivalent. Furthermore the transformation $D_a \to \tilde{D}_a$ leaves the two fields N_{ab}, $*K^{ab}$ invariant provided D and $\tilde{D}$ belong to the same equivalence class. If $N_{ab} = 0$ (compact $\mathscr{I}$) one can thus focus on the degrees of freedom incorporated in the symmetric and traceless tensor $*K^{ab}$. In the presence of radiation, on the other hand, one can consider that the basic variables are the equivalence classes $\{D\}$'s. Recall that two connections D^1 and D^2 on $\mathscr{I}$ are related via

$$(D^1{}_a - D^2{}_a)K_b = \Sigma_{ab} K_m \mathbf{n}^m, \tag{44}$$

for some tensor field Σ_{ab} satisfying $\Sigma_{ab}\mathbf{n}^b = 0$. It thus follows that a pair of equivalence classes can be characterized by γ_{ab}, the trace-free part of Σ_{ab}: there are precisely two radiative degrees of freedom at conformal infinity. A scheme of asymptotic quantization can be derived consequently. First one introduces the (phase) space Γ of equivalence classes, an infinite dimensional affine space, the symplectic tensor $\mathbf{\Omega}$ on Γ being defined via

$$\mathbf{\Omega}_{(D)}(\gamma,\tilde{\gamma}) = \int_{\mathscr{I}} (\gamma_{ab}\mathscr{L}_{\mathbf{n}}\tilde{\gamma}_{cd} - \tilde{\gamma}_{ab}\mathscr{L}_{\mathbf{n}}\gamma_{cd})\mathbf{g}^{ac}\mathbf{g}^{bd}\, d\mathscr{I}. \tag{45}$$

The restriction to right- (left-) flat complex solutions induces a negative- (positive-) frequency decomposition of γ_{ab}: $\gamma_{ab} = \gamma^+{}_{ab} + \gamma^-{}_{ab}$. The procedure presented in Sec. IV, which, starting from $*K^{-ab}$, induces self- (anti-self-) dual curvature two-forms [formulas (37)–(41)], can be reproduced. Since

$$\epsilon_{abm}\gamma^{-\,mp}M_p = fL_{[a}M_{b\,]}, \tag{46}$$

the (right-) left-handed asymptotic graviton modes, defined[5] via

$$\epsilon^{mnp}L_p\mathbf{g}_{nb}\gamma^-{}_{ma} = \pm i\gamma^-{}_{ab}, \tag{47}$$

appear as transverse to those presented in Sec. IV; this confirms the viewpoint according to which gauge fields $A^{\pm}{}_{BB'}$ could describe (solitonic) graviton modes originating within the space-time nontrivial topology.

(ii) An extension of the above results to solutions with matter contents would be rather straightforward. Recently, fluid generalizations[12] of NUT spaces have been manufactured, using the three-dimensional spin coefficient method, which brings out the nontrivial topological structure of such space-times. The expression of the metric in the case of a (timelike) stationary fluid flow has been explicitly given in Ref. 12. It follows that the topology of the $r = \text{const}$ hypersurfaces is S^3, the timelike Killing trajectories (fluid flow) being the S^1 (Hopf) fibers of S^3 over S^2. This metric reduces to the NUT metric in the matter-free limit. Since the bundle structure is not modified by the presence of matter, our results could be easily adapted in the case of (nonvacuum) solutions, which are (at least asymptotically) Lorentzian sections of complex solutions.

(iii) A reformulation of the above results using twistorial methods (Ref. 8) is required if a relation between our self- (anti-self-) dual gauge connections and self- (anti-self-) dual Yang–Mills fields [introduced as restrictions of curvature two-forms to β-planes or α-planes (in the sense of Penrose)] is to be exhibited within the context of curved twistor theory.[13]

ACKNOWLEDGMENTS

I would like to thank Professor Roger Penrose for suggestions, and P. Mazur and B. Schmidt for helpful discussions.

[1] J. A. Wheeler, in *Relativity, Groups and Topology*, edited by B. S. DeWitt and C. M. DeWitt (Gordon and Breach, New York, 1964).
[2] A. Magnon, "Mass, dual mass and gravitational entropy," submitted to J. Math. Phys.
[3] A. Magnon, "Global techniques, dual mass, and causality violation," J. Math. Phys. **27**, 1066 (1986).
[4] A. Magnon, "A Generalization of Penrose's helicity theorem for space-times with nonzero dual mass," J. Math. Phys. **27**, 1059 (1986).
[5] A. Ashtekar, "Asymptotic quantization of the gravitational field," Phys. Rev. Lett. **46**, 9 (1981).
[6] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, Proc. R. Soc. London Ser. A **362**, 425 (1978).
[7] R. Penrose, "Twistor theory, its aims and achievements," in *Quantum Gravity, an Oxford Symposium*, edited by C. J. Isham and R. Penrose (Oxford U. P., Oxford, 1975).
[8] R. Penrose and R. S. Ward, *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[9] C. N. Kozameh and E. T. Newman, Gen. Relativ. Gravit. **15**, 5 (1983).
[10] M. Ko, M. Ludvigsen, E. T. Newman, and K. P. Tod, *The Theory of $\mathscr{H}$-spaces* (North-Holland, Amsterdam, 1981).
[11] J. Plebański, J. Math. Phys. **16**, 2395 (1975).
[12] B. Luka'cs, E. T. Newman, G. Sparling, and J. Winicour, Gen. Relativ. Gravit. **15**, 5 (1983).
[13] R. Penrose, Gen. Relativ. Gravit. **7**, 1 (1976).

Total energy momentum in general relativity

Niall Ó Murchadha
Physics Department, University College, Cork, Ireland

(Received 31 January 1986; accepted for publication 2 April 1986)

The energy momentum of any asymptotically flat vacuum solution to the Einstein equations is a well-defined, conserved, Lorentz-covariant, timelike, future-pointing vector. The only requirement is that one be given asymptotically flat initial data that satisfy very weak continuity and falloff conditions; the three-metric must go flat faster than $r^{-1/2}$. A large class of such data exists, consistent with the constraints, and the constraints play a key role in guaranteeing that the energy momentum is well behaved.

I. INTRODUCTION

Many years ago people realized that asymptotically flat solutions to the Einstein equations possess conserved quantities. These conserved quantities were identified as the total energy momentum and the total angular momentum of the solution. Given a global singularity-free solution, and making plausible assumptions about the falloff of the metric, it was shown that the energy momentum was a conserved Lorentz four-vector.[1,2] Making similar assumptions (more or less that the metric falls off like $1/r$ at infinity), it was more recently demonstrated that this Lorentz vector was future pointing and timelike.[3–5]

In this paper we wish to prove that I get a well-defined, finite, conserved energy momentum, which forms a timelike, future-pointing Lorentz four-vector under much weaker assumptions about the asymptotic behavior of the metric. This improved result is an application of a recent proof that a very large class of asymptotically flat initial data possesses extended domains of development. These domains are large enough to permit coordinate transformations that become Lorentz transformations at infinity. Further, the space-time metric on this large domain inherits the falloff characteristics of the original data.

These conclusions ("the Boost Theorem")[6] mean that we need make no *a priori* assumptions about the space-time, either as regards its size, or the falloff properties of the metric. We can replace these with asymptotic conditions on the initial data. Essentially, I will show that if the energy momentum associated with a single initial slice is finite, then it will be globally well-defined.

Further, I will show that the standard asymptotic condition used ($1/r$ falloff in the metric) can be significantly relaxed. I will show that we only need that the metric approaches the flat metric faster than $r^{-1/2}$ to get a well-defined, conserved, finite energy momentum. This weaker asymptotic condition suffices because of the existence of constraints on the initial data. The conserved quantities are usually expressed as surface integrals at infinity (the ADM expressions[7]). If we only consider these surface integrals, then the $1/r$ falloff is the natural condition to guarantee finiteness. However, we can use the Gauss theorem to turn the surface integrals into volume integrals. The leading term in the volume integral expression vanishes, due to the constraints, and thus the volume integral expressions for the energy momentum converge using the weaker falloff condition.

This intimate connection between the conserved quantities and the constraints can be seen by considering the equivalent problem in electromagnetism. The total energy in general relativity is equivalent to, not one, but two independent quantities in electromagnetism, the total charge and the total energy, and it is illuminating to consider the asymptotic conditions necessary for the finiteness of each of these objects.

We know that the total charge can be expressed as a surface integral at infinity (the analog of the ADM integrals)

$$Q = \epsilon_0 \oint_\infty \mathbf{E} \cdot \mathbf{d}S, \tag{1.1}$$

where $\mathbf{E}$ is the electric field strength. This can be turned into a volume integral using the Gauss theorem

$$Q = \epsilon_0 \int_V \operatorname{div} \mathbf{E}\, d^3x. \tag{1.2}$$

On using the Maxwell initial value constraint

$$\epsilon_0 \operatorname{div} \mathbf{E} = \rho$$

this becomes

$$Q = \int_V \rho\, d^3x, \tag{1.3}$$

where ρ is the charge density.

If the initial value condition had been ignored the surface integral expression would have led one to the (incorrect) conclusion that for Q to be finite, $\mathbf{E}$ must fall off like $1/r^2$. The correct conclusion (on using the constraint) is that there is no direct restriction on $\mathbf{E}$, and that one only requires ρ to fall off faster than $1/r^3$. When this condition is substituted back into the constraint, it leads to the condition that the longitudinal part of $\mathbf{E}$ (the potential part) must fall off like $1/r^2$, and that the transverse part of $\mathbf{E}$ has no restrictions placed on it. This is only to be expected since electromagnetic waves are intrinsically uncharged.

When we look at the total energy of the electromagnetic field on the other hand we do get real restrictions on the asymptotic behavior of both the electric and magnetic fields. The total energy is

 0022-2488/86/082111-18$02.50

$$M=\frac{1}{2}\int_V\left(\epsilon_0E^2+\frac{B^2}{\mu_0}\right)d^3x. \tag{1.4}$$

Therefore, for M to be finite, **E** and **B** must fall off faster than $r^{-3/2}$. In terms of the four-potential A_μ this means $A_\mu\sim r^{-1/2}, A_{\mu,\nu}\sim r^{-3/2}$. This condition that the four-potential must fall off faster than $r^{-1/2}$ at infinity for finite total energy holds not only for Maxwell's equations but also for Yang–Mills theory.[8] The four-potential is the object that is directly analogous to the metric in general relativity and so it is very nice that the finite energy condition in general relativity is that the metric approach the Minkowski metric faster than $r^{-1/2}$.

The parallels between electromagnetism and general relativity run much deeper than this $r^{-1/2}$ coincidence, however. The mass in general relativity can be expressed as a surface integral [just like (1.1)]. This, when turned into a volume integral, can be reduced to a volume integral over the source using the initial value constraint [just like (1.3)]. Unfortunately the reduction in the case of gravity is not as clean as in the case of electromagnetism. The gravitational waves carry energy whereas the electromagnetic waves are not charged. This means that the gravitational analog of (1.3) is of the form

$$M\sim\int_V(``\rho+E^2+B^{2}")d^3x, \tag{1.5}$$

where "E^2+B^2" is shorthand for terms like $(g_{ij,k})^2$, which represent the wave energy density. This is why gravitational energy is a composite of the properties of electromagnetic charge and energy.

In field theories, conserved quantities are related to symmetries. In general relativity, the total energy momentum is generated not by any exact symmetry of the given solution, but rather by the time and space translational symmetry of the underlying Minkowski space. This is why only asymptotically flat solutions have a conserved energy momentum. If we have an asymptotically flat solution we do not have a unique underlying Minkowski space; rather we have a whole family of them. Thus to show that the energy momentum is well defined we have to consider not only transformations (boosts and rotations) of a given Minkowski space but also the effect of transforming from one Minkowski space to another. The class of flat spaces we have to consider is determined by the falloff characteristics of the given metric.

When we say that a metric g_{ij} goes flat faster than $r^{-\alpha}$ we mean that g_{ij} can be written in the form

$$g_{ij}=\delta_{ij}+h_{ij}, \tag{1.6}$$

and $r^\alpha h_{ij}\to 0$ at infinity, where r is a radial measure and it does not matter whether we measure it with respect to g_{ij} or δ_{ij}. Thus on a given manifold we have a curved metric g_{ij} and a flat metric δ_{ij}, which agree near infinity. Now if we perform a coordinate transformation f on the manifold that approaches the identity transformation at infinity we will have that $g'=f(g)$ is also asymptotically flat, i.e., g' can be written

$$g'_{ij}=\delta_{ij}+h'_{ij}, \tag{1.7}$$

with h'_{ij} vanishing near infinity. It is vital to notice that the δ_{ij} in (1.6) is a different flat metric from the δ_{ij} in (1.7) and that h'_{ij} is not the coordinate transformation of h_{ij} in (1.6). The two δ_{ij} are each diag(1,1,1) but in different coordinate systems.

However, we are given that the original h_{ij} falls off faster than $r^{-\alpha}$. Thus it seems perverse to allow coordinate transformations that do not preserve this property. In turn this means that we do not consider all coordinate transformations but only a restricted class, those which do not worsen the falloff characteristics of the metric. This also restricts the class of the flat spaces we have to consider.

Hence, the statement that the metric falls off faster than $r^{-1/2}$ is used twice. First, it is used to show that the energy momentum is finite and well behaved with respect to a given Minkowski background. Second, we have to show that the energy momentum is unchanged when we change the background. However, the set of allowed backgrounds is restricted also by the $r^{-1/2}$ falloff and this permits us to prove the necessary result. We will also show (by means of a counterexample) that if we relax either of these conditions, then the energy momentum goes crazy.

The theorems proved in this paper are all expressed in terms of weighted Sobolev spaces. We will attempt to give a precise definition for every quantity used in this paper but we will not aim at completeness in that we will only state those results that we need and that are available in the literature.

Thus, in Sec. II, I will give a very brief account of weighted Sobolev spaces and their useful properties. I will give a much more detailed discussion of what I would call the asymptotic structure group, the class of diffeomorphisms that preserve a given metric falloff. This group contains not only those coordinate transformations that reduce to the identity at infinity but also the rotations and translations of the Euclidean group for Riemannian manifolds (for pseudo-Riemannian manifolds we include the Poincaré group). We obviously will have to consider these rotations and boosts in any analysis of the energy momentum. Nothing surprising happens; the asymptotic symmetry group is the direct sum of the Euclidean group and of the transformations that reduce to the identity.

Section III consists just of a statement of the boost theorem.[6] Section IV is the central section of the paper. In it, by a careful reworking of the analysis in Weinberg,[2] we show that the energy momentum is well defined and finite under the weaker ($r^{-1/2}$) falloff conditions I use.

In Sec. V I show how this $r^{-1/2}$ falloff is compatible with the surface integral formulation of the energy momentum, which seems to demand an r^{-1} falloff. I strengthen the analogy with electromagnetism by showing that we can break the metric into two parts, a "wave" part, which falls off faster than $r^{-1/2}$, and a "Newtonian potential" part, which falls off like $1/r$. The "wave" part, just like the transverse part of the electric field, does not contribute to the surface integral. Only the "Newtonian potential" part contributes, and this must be finite. This breakup is most easily effected by using harmonic coordinates near infinity. To prove the existence of these harmonic coordinates we need to understand the properties of the Laplacian when operating in weighted Sobolev spaces. Thus in Sec. V we prove the desired results and use them to demonstrate the existence of the desired coordinate system.

Section VI is devoted to showing that the energy, in addition to being finite, is always positive. This, when combined with the boost theorem, shows that the energy momentum is both future pointing and timelike. It largely consists of a reworking of the standard positive energy proofs[3–5] to show that they work under the weaker asymptotic conditions used here.

Each of the three major sections IV, V, and VI are written so as to stand on their own, independent of one another. However, they each draw heavily on the two preliminary sections II and III.

II. WEIGHTED SOBOLEV SPACES

We will want to discuss two kinds of spaces in this article. The initial data for the gravitational field will be defined on the three-dimensional spacelike slices and therefore we will have to define functions and metrics on Riemannian spaces. In addition, the space-time itself will be a pseudo-Riemannian manifold, which will have to be handled differently.

A. Riemannian spaces

We define two classes of functions on an n-dimensional Riemannian manifold $\mathbb{R}^n$, specifying both differentiability (continuity) properties and falloff properties. We want, in particular, that each derivative (up to some order usually denoted by s) falls off faster (by one power of r) than the previous stage. In other words, we want the function to fall off like $r^{-\delta}$, the first derivatives to fall off like $r^{-(\delta+1)}$, the second derivative to fall off like $r^{-(\delta+2)}$ and so on.

To make this precise we want to be given a Euclidean (flat) metric e on $\mathbb{R}^n$ and we define the function σ on $\mathbb{R}^n$ to be

$$\sigma(x) = (1 + |x^2|)^{1/2},$$

where $x \in \mathbb{R}^n$, $|\cdot| \equiv \|\cdot\|_e$. Here σ looks just like the radial distance r at ∞ but it remains positive (in fact $\geqslant 1$) even at the origin. This is only to ensure that $\sigma^{-\alpha}$ is well behaved everywhere (whereas $r^{-\alpha}$ is not). The use of a Euclidean metric e permits us to consider non-Cartesian coordinates (spherical polars, or whatever), but little is lost if one restricts oneself to Cartesians. I will do so, and discuss at the end how to generalize.

The first set of functions we wish to consider are ordinary classical functions, which are differentiable s times and which fall off like $r^{-\delta}, r^{-(\delta+1)},\ldots,r^{-(\delta+s)}$, where s is a natural number and δ is a real number. We wish to consider metrics and other tensors so we will specify that the functions belong to some particular tensor class. The derivative operator will be the covariant derivative ∇ with respect to the given metric e.

Definition 2.1: $C^s_\delta(\mathbb{R}^n)$, $s \in \mathbb{N}$, $\delta \in \mathbb{R}$ is the Banach space of functions u in $\mathbb{R}^n$, with values in some finite dimensional vector space V of class C^s such that

$$\|u\|_{C^s_\delta(\mathbb{R}^n)} = \sup_{\mathbb{R}^n} \left\{ \sum_{|\alpha|<s} |\sigma^{\delta+|\alpha|}\nabla^\alpha u| \right\} < \infty .$$

In addition to these classical functions we also wish to consider distributions, functions that are not well defined at every point of space, but are well defined under integration. The norm we wish to use is a variant of the Sobolev norm, usually denoted by H_s, which involves squaring everything and integrating

$$\|u\|_{H_s} = \sum_{|\alpha|<s} \int_{\mathbb{R}^n} |\nabla^\alpha u|^2 \, d\mu(e),$$

where $d\mu(e)$ is the volume density related to e.

Such a norm, while it looks unnatural to the uneducated eye, is a much more physical object than the classical continuous functions. There is no way we can ever measure the value of the electric field at a point whereas $\int E^2 + B^2$ is a very important physical quantity. Therefore we wish to define the following.

Definition 2.2 (weighted Sobolov spaces): $H_{s,\delta}(\mathbb{R}^n)$, $s \in \mathbb{N}$, $\delta \in \mathbb{R}$, is the Hilbert space of functions u on $\mathbb{R}^n$ with values in V possessing weak derivatives up to order s, such that

$$\|u\|_{H_{s,\delta}(\mathbb{R}^n)} = \sum_{|\alpha|<s} \int_{\mathbb{R}^n} \sigma^{2(\delta+|\alpha|)} |\nabla^\alpha u|^2 \, d\mu(e) < \infty .$$

The Sobolev space is a Hilbert space because we can take the dot product of two functions

$$\|u \cdot v\| = \sum_{|\alpha|<s} \int_{\mathbb{R}^n} \sigma^{2(\delta+|\alpha|)} (\nabla u \cdot \nabla v) d\mu(e).$$

No such operation is naturally defined by the C^s_δ norm, which is why it is only a Banach space.

In our discussions, we will spend some time moving between the weighted Sobolev spaces and the classical weighted spaces. The relationship between the two sets is clearly complicated by the fact that $d\mu(e) \sim r^{+n-1} dr$ so that if $r^{\delta+n/2}u$ goes to zero at infinity, then $\int r^{2\delta}u^2$ is finite. This means that δ in one measure has to be replaced by $\delta + n/2$ in the other. This is not the major difficulty, however. This is due to the fact that the statement above is not reversible. The fact that $\int r^{2\delta}u^2$ is finite does not guarantee that $r^{\delta+n/2}u$ is even bounded.

A similar problem arises with differentiability. A strong differentiable function is also weakly differentiable, but a weakly differentiable function is not strongly differentiable. The Sobolev imbedding theorem[9] shows, however, that a function that is weakly differentiable to a high order is strongly differentiable to a lesser order. One loses $n/2$ degrees of differentiability. This means that in three-space if a function is weakly differentiable five times it is strongly differentiable three times.

In turns out that losing these $n/2$ degrees also resolves the falloff problem, and in three-space if a function satisfies a weak falloff condition to five orders, it satisfies a strong falloff condition to three orders. The following theorem has been proven.[10]

Theorem 2.1 (imbedding):

$$H_{s,\delta}(\mathbb{R}^n) \subset C^{s'}_{\delta'}(\mathbb{R}^n), \quad \text{if } s' < s - n/2, \quad \delta' < \delta + n/2,$$

$$C^{s'}_{\delta'}(\mathbb{R}^n) \subset H_{s,\delta}(\mathbb{R}^n), \quad \text{if } s' \geqslant s, \quad \delta' > \delta + n/2.$$

In the course of various calculations in this paper we will have to deal with complicated expressions like

$g^{am}g^{bn}g^{cp}g_{ab,c}\, g_{mn,p}$. Therefore we will have to multiply functions belonging to different weighted Sobolev spaces together. This is handled by the following theorem.

Theorem 2.2 (multiplication)[10]: Pointwise multiplication on $\mathbb{R}^n$ is a continuous, bilinear map

$$H_{s_1,\delta_1}(\mathbb{R}^n)\times H_{s_2,\delta_2}(\mathbb{R}^n)\to H_{s,\delta}(\mathbb{R}^n),$$

if

$$s_1,s_2\geqslant s,\quad s<s_1+s_2-n/2,\quad \delta<\delta_1+\delta_2+n/2.$$

The first two conditions are the standard ones from the multiplication theorem for regular Sobolev spaces.[11] The first one says that you cannot improve the differentiability of functions by multiplying them together. The second condition says that if you multiply functions of low differentiability together, you may worsen the differentiability of the combination. An extreme example: If u and v are square-integrable functions, the combination uv, while integrable, need not be square integrable.

The condition on δ is very easy to understand. All it says is that the falloff of the combination is the sum of the falloffs of the individual terms. This may be more obvious when the expression is written in the form

$$\delta+n/2<(\delta_1+n/2)+(\delta_2+n/2).$$

Corollary: $H_{s,\delta}(\mathbb{R}^n)$ is a Banach algebra if $s>n/2$, $\delta>-n/2$. This is just a fancy way of saying that if $s-n/2>0$, $\delta+n/2>0$, then $H_{s,\delta}\times H_{s,\delta}$ belongs to the same $H_{s,\delta}$. Here $\delta+n/2>0$ may be naively interpreted as functions that vanish at infinity. Thus, given two functions, each of which vanishes at infinity, the product also vanishes at infinity.

A key role in this paper will be taken by what we call the asymptotic symmetry group, the set of diffeomorphisms that preserve the asymptotic falloff of the metric. Part of the groundwork has been done already. We already have proved the following lemma.[6]

Lemma 2.1: If f is a diffeomorphism $\mathbb{R}^n\to\mathbb{R}^n$ and $f-\mathrm{id}\in H_{s+1,\delta-1}(\mathbb{R}^n)$, $s>n/2$, $\delta>-n/2$, then $f^{-1}-\mathrm{id}\in H_{s+1,\delta-1}(\mathbb{R}^n)$.

This lemma deals with diffeomorphisms that, near infinity, become the identity $x\to x$, plus a term that blows up slower than r. This means that the derivative of f is the identity matrix plus a term that belongs to $H_{s,\delta}$ ($s>n/2, \delta>-n/2$). The extra term (from the imbedding theorem) belongs to C^0_ϵ for some $\epsilon>0$, and so is pointwise well-defined and vanishes at infinity. The effect of these diffeomorphisms is reflected in the following theorem.

Theorem 2.3 (composition)[6]: If f is a diffeomorphism on $\mathbb{R}^n$ such that $f-\mathrm{id}\in H_{s+1,\delta-1}(\mathbb{R}^n)$, $s>n/2$, $\delta>-n/2$; u is a function on $\mathbb{R}^n$ belonging to $H_{s',\delta'}(\mathbb{R}^n)$ then composition $u\to u\cdot f$ is an isomorphism

$$H_{s',\delta'}(\mathbb{R}^n)\to H_{s',\delta'}(\mathbb{R}^n),\quad \text{for every } s'\leqslant s+1,\ \delta'\in\mathbb{R}.$$

This theorem says that these diffeomorphisms, which reduce to the identity at infinity, preserve asymptotic falloff ($\delta'\in\mathbb{R}$) and preserve differentiability as much as possible ($s'\leqslant s+1$). Lemma 2.1 and Theorem 2.3 now can be combined to prove[6] the following corollary.

Corollary: Let $D_{s+1,\delta-1}(\mathbb{R}^n)$ denote the set of diffeomorphisms f on $\mathbb{R}^n$ such that $f-\mathrm{id}\in H_{s+1,\delta-1}(\mathbb{R}^n)$ with $s>n/2$, $\delta>-n/2$. Then $D_{s+1,\delta-1}(\mathbb{R}^n)$ is a topological group with respect to composition.

Of course $D_{s+1,\delta-1}(\mathbb{R}^n)$ does not include all the diffeomorphisms that preserve falloff. We would expect that members of the Euclidean group (rigid rotations and translations) would have this property in addition to those diffeomorphisms that reduce to the identity at infinity. The new result in this section will be to show how the Euclidean group can be combined with $D_{s+1,\delta-1}$.

Lemma 2.2: If f is a member of the Euclidean group of $\mathbb{R}^n$ and $u\in H_{s,\delta}(\mathbb{R}^n)$ is a function on $\mathbb{R}^n$, then composition $u\to u\cdot f$ is an isomorphism $H_{s,\delta}(\mathbb{R}^n)\to H_{s,\delta}(\mathbb{R}^n)$ for every $s\in\mathbb{N}$, $\delta\in\mathbb{R}$.

Proof: Let us denote $y=f(x)$, $x\in\mathbb{R}^n$. Since f is a combination of a rotation and a translation, one can show there exist positive constants C_1,C_2 such that

$$C_1\sigma(x)\leqslant\sigma(y)\leqslant C_2\sigma(x),\quad \forall x\in\mathbb{R}^n.$$

We also have $d\mu(y)=d\mu(x)$. Hence there exist positive constants $\overline{C}_1,\overline{C}_2$ such that

$$\overline{C}_1\int\sigma(x)^{2\delta}|u\cdot f(x)|^2\,d\mu(x)$$

$$\leqslant\int\sigma(y)^{2\delta}|u(y)|^2\,d\mu(y)$$

$$\leqslant\overline{C}_2\int\sigma(x)^{2\delta}|u\cdot f(x)|^2\,d\mu(x).$$

This is sufficient to show $u\to u\cdot f$ is an isomorphism $H_{0,\delta}\to H_{0,\delta}$. In addition we have $|\nabla^\alpha u(y)|=|\nabla^\alpha u\cdot f(x)|$ for every $\alpha>0$. This is sufficient to prove the lemma.

Corollary: Let $E_{s+1,\delta-1}(\mathbb{R}^n)$ denote the direct product of $D_{s+1,\delta-1}(\mathbb{R}^n)$, $s>n/2$, $\delta>-n/2$ with the Euclidean group. Then $E_{s+1,\delta-1}(\mathbb{R}^n)$ is a group and $D_{s+1,\delta-1}(\mathbb{R}^n)$ is a normal subgroup. The factor group is the Euclidean group back again.

It is not enough for our purposes to just discuss the effects on scalar functions of the action of the diffeomorphisms. We will wish to deal with the situation where we have a Riemannian metric g on $\mathbb{R}^n$ (in addition to the Euclidean metric e). Here g and e will be asymptotically related by the fact that $g-e\in H_{s,\delta}(\mathbb{R}^n)$ for some $s>n/2$, $\delta>-n/2$. This is how we generically specify asymptotic flatness because the imbedding theorem guarantees that $g-e$ goes (classically) to zero at infinity. The rate of decay is determined by the particular value of δ specified.

We wish to investigate the behavior of g under various diffeomorphisms. If f is a diffeomorphism we will denote the image of g under f as $f'(g)$. We intend to prove that if $g-e\in H_{s,\delta}$ and $f\in E_{s+1,\delta-1}$ (with, of course, $s>n/2$, $\delta>-n/2$) then $f'(g)-e\in H_{s,\delta}$ (notice that the same s and δ are in H and E). Thus $E_{s+1,\delta-1}(\mathbb{R}^n)$ is the natural asymptotic symmetry group associated with g.

Before we discuss the effects of f on g, we wish to understand the effects of f on e. This is covered by the following lemma.

Lemma 2.3: If $f\in E_{s+1,\delta-1}(\mathbb{R}^n)$, then $f'(e)-e\in H_{s,\delta}(\mathbb{R}^n)$ for every $s>n/2$, $\delta>-n/2$.

Proof: If f is a member of the Euclidean group then $f'(e) = e$, which trivially proves the lemma. Hence we can restrict ourselves to the case where $f \in D_{s+1,\delta-1}(\mathbb{R}^n)$. Using the standard rule for coordinate transformations we have

$$f'(e) = e \cdot (Df)^2 \cdot f^{-1},$$

which can be rewritten

$$\begin{aligned} f'(e) &= e \cdot [I + D(f - \mathrm{id})]^2 \cdot f^{-1} \\ &= e \cdot [I + D(f - \mathrm{id})]^2 \\ &\quad + \mathrm{e} \cdot [I + D(f - \mathrm{id})]^2 \cdot (f^{-1} - \mathrm{id}), \end{aligned}$$

where I is the identity matrix and id is the identity map. This expression can be expanded out, and the multiplication theorem (together with Lemma 2.1) gives us the desired result.

All this lemma is saying is that if I have Cartesian coordinates and I make a coordinate transformation that reduces to the identity at infinity, the transformed coordinates are asymptotically Cartesian. The next piece that we need is an extension of the composition theorem (Theorem 2.3) to tensors. This is given by the following lemma.

Lemma 2.4: If $f \in E_{s+1,\delta-1}(\mathbb{R}^n)$, $s > n/2$, $\delta > -n/2$, and a tensor $t \in H_{s',\delta'}(\mathbb{R}^n)$, then $f'(t) \in H_{s',\delta'}(\mathbb{R}^n)$ if $s' \leqslant s$, $\delta' \in \mathbb{R}$.

Proof: If t is a pth-order tensor we have

$$f'(t) = t \cdot (Df)^p \cdot f^{-1}.$$

This, as in Lemma 2.3, can be expanded out, and use of the multiplication theorem and Lemma 2.1 gives the desired result.

All this lemma says is that any diffeomorphism that reduces to a Euclidean transformation at infinity will not change the falloff properties of a tensor ($\delta' \in \mathbb{R}^n$), nor will it change its differentiability so long as the diffeomorphism is as differentiable as the tensor $s' \leqslant s$. Note that the composition theorem had $s' \leqslant s + 1$; the difference is that the tensor transformation involves Df, whereas the scalar transformation involves only f.

Lemma 2.3 and Lemma 2.4 can now be combined to prove the following theorem.

Theorem 2.4 (asymptotic symmetry): If g is a two-tensor and if f belongs to $E_{s+1,\delta-1}(\mathbb{R}^n)$, $s > n/2$, $\delta > -n/2$, then

$$g - e \in H_{s',\delta'}(\mathbb{R}^n) \Rightarrow f'(g) - e \in H_{s',\delta'}(\mathbb{R}^n)$$
$$\text{for every } s' \leqslant s, \quad \delta' \leqslant \delta.$$

Proof: We can write

$$\begin{aligned} f'(g) - e &= f'(g) - f'(e) + f'(e) - e \\ &= f'(g - e) + [f'(e) - e]. \end{aligned}$$

We can immediately apply Lemma 2.4 to the first term and Lemma 2.3 to the last term to obtain the result.

The order in which this theorem should be read is that we are given g, which specifies s' and δ'. Then the largest set of diffeomorphisms that preserves $f'(g)$ in both falloff and differentiability is $E_{s'+1,\delta'-1}(\mathbb{R}^n)$. Of course we need $s' > n/2$, $\delta' > -n/2$. All this says is that g is naively asymptotically flat, i.e., there exists an ϵ such that $r^\epsilon(g - e) \to 0$ classically. The theorem does not deal with the falloff of the tensor $g - e$ (this is dealt with in Lemma 2.4). We have two different flat metrics $f'(e)$ and e on the same manifold and we can compare $f'(g)$ with either of them.

In the results we have proven here we have implicitly assumed that $e = \mathrm{diag}(1,1,...,1)$. This can be relaxed easily to the case where $e - \mathrm{diag}(1,1,...,1) \in H_{s'',\delta''}$ for some suitably large s'',δ''. Alternatively we could express it in terms of the mapping that maps e to $\mathrm{diag}(1,1,...,1)$ and have this mapping belonging to $D_{s''+1,\delta''-1}(\mathbb{R}^n)$ again for some suitably large s'',δ''.

A more difficult problem arises when we wish to deal with a non-Cartesian e (such as in spherical polar coordinates). In this case the results we have derived are clearly untrue. The easiest way to deal with this problem is to find the mapping h which brings one from the desired coordinates to Cartesian coordinates; identify the asymptotic symmetry group; and map back again. Thus, the diffeomorphism group we would deal with (and identify as the asymptotic symmetry group) is $h^{-1} \cdot E_{s+1,\delta-1} \cdot h$. All this does is reflect the fact that the Euclidean group in, say, polar coordinates looks very different from the same group in Cartesian coordinates.

B. Pseudo-Riemannian spaces

In general relativity we will be dealing with a four-dimensional manifold with a metric with signature $(+,+,+,-)$. We wish to extend the weighted Sobolev spaces to such manifolds and prove theorems analogous to those proven in Sec. II A. We begin by considering $\mathbb{R}^4$ with a Minkowski metric η. Associated with the Minkowski metric is a standard Euclidean metric $e = \eta + 2t \times t$, where t is a unit timelike Killing vector. In Cartesian coordinates, the Minkowski metric is $\mathrm{diag}(+1,+1,+1,-1)$, the timelike Killing vector is $(0,0,0,1)$, and therefore the Euclidean metric is $e = \mathrm{diag}(+1,+1,+1,+1)$. We need the Euclidean metric e to give us a positive norm on $\mathbb{R}^4$.

We define a function σ on $\mathbb{R}^4$ by

$$\sigma(x) = \{1 + |x|^2\}^{1/2}, \quad x \in \mathbb{R}^4, \quad |\cdot| = \|\cdot\|_e.$$

On any subset U of $\mathbb{R}^4$ we can define function spaces $C^s_\delta(U)$ and $H_{s,\delta}(U)$ analogous to Definitions 2.1 and 2.2 with (semi)norms

$$\|u\|_{C^s_\delta(U)} = \sup_U \left\{ \sum_{|\alpha| \leqslant s} \sigma^{|\alpha| + \delta} |\nabla^\alpha u| \right\},$$

$$\|u\|_{H_{s,\delta}(U)} = \sum_{|\alpha| \leqslant s} \int_U \sigma^{2(|\alpha| + \delta)} |\nabla^\alpha u|^2 \, d\mu(e).$$

They are seminorms rather than norms because they may not satisfy $\|A\| = 0 \Rightarrow A = 0$. We assume, as before, that u belongs to some given finite-dimensional vector space.

In this article we will be interested in special subsets of $\mathbb{R}^4$. The viewpoint we will adopt is that we consider an initial spacelike three-slice on which we give initial data for the gravitational field. We are interested in a subset of a Cauchy development of this data. We wish to consider both translating the original slice in time and tilting it (corresponding to a Lorentz boost). Thus we wish to consider subsets of $\mathbb{R}^4$ that are wedge shaped near infinity.

In standard coordinates let Ω_θ be the domain defined by

$$\Omega_\theta = \{x \in \mathbb{R}^4 \colon |x^0|/\sigma(x) < \theta\}, \quad 0 < \theta < 1/\sqrt{2}.$$

These domains are large enough to permit Lorentz boosts, and the closer θ is to $1/\sqrt{2}$, the closer the boost can be made to light speed. It has been shown[10] that the imbedding theorem, the multiplication theorem, Lemma 2.1, and the composition theorem are again correct in any such domain (with, of course, $\mathbb{R}^n$ replaced by Ω_θ).

This means that we have the group of diffeomorphisms $D_{s+1,\delta-1}(\Omega_\theta)$, which reduce to the identity at infinity, and preserve falloff. In Sec. II A we added the Euclidean group to $D_{s+1,\delta-1}$ to get the asymptotic symmetry group. The symmetry group of Minkowski space is the Poincaré group. Therefore in this section we wish to add the Poincaré group to $D_{s+1,\delta-1}(\Omega_\theta)$ and derive results analogous to Lemmas 2.2–2.4 and Theorem 2.4.

The first difficulty that arises is that if we have a set Ω_θ in some set of standard (Cartesian) coordinates, and Lorentz transform these coordinates, the original set Ω_θ is not an Ω_θ of the transformed coordinates. Rather it is contained in an Ω_{θ_1} and contains an Ω_{θ_2} $(\theta_1,\theta_2 < 1/\sqrt{2})$ of the transformed coordinates. In fact the imbedding, multiplication, and composition theorems are still true in the Lorentz-transformed Ω_θ.

Analogous to Lemma 2.2 we now wish to prove the following lemma.

Lemma 2.5: If f is a member of the Poincaré group and u is a scalar function on Ω_θ $(\theta \in (0,1/\sqrt{2}))$, then the composition $u \to u \cdot f$ is an isomorphism

$$H_{s,\delta}(\Omega_\theta) \to H_{s,\delta}(\Omega_\theta), \quad \text{for every } s \in \mathbb{N}, \quad \delta \in \mathbb{R}.$$

Proof: Let us denote $y = f(x)$. Since $x \in \Omega_\theta$, we can show that there exist positive constants C_1, C_2 such that $C_1\sigma(z) \leqslant \sigma(y) \leqslant C_2\sigma(x)$. In standard coordinates $d\mu(x) = d\mu(y)$. Hence there exist positive constants $\overline{C}_1, \overline{C}_2$ such that

$$\overline{C}_1 \int \sigma(x)^{2\delta} |u \cdot f(x)|^2 \, d\mu(x)$$

$$\leqslant \int \sigma(y)^{2\delta} |u(y)|^2 \, d\mu(y)$$

$$\leqslant \overline{C}_2 \int \sigma(x)^{2\sigma} |u \cdot f(x)|^2 \, d\mu(x).$$

This guarantees $H_{0,\delta} \to H_{0,\delta}$ is an isomorphism. In addition, for every α there exist positive constants $C_{1\alpha}, C_{2\alpha}$ such that

$$C_{1\alpha} |D^\alpha u(x)| \leqslant |D^\alpha u(y)| \leqslant C_{2\alpha} |D^\alpha u(x)|.$$

This is sufficient to prove the lemma.

Corollary: Let $D_{s+1,\delta-1}(\Omega_\theta)$ denote the set of diffeomorphisms on Ω_θ such that $f - \mathrm{id} \in H_{s+1,\delta-1}(\Omega_\theta)$, $s > 2$, $\delta > -2$. Let $P_{s+1,\delta-1}(\Omega_\theta)$ be the direct sum of $D_{s+1,\delta-1}$ and the Poincaré group. Here $P_{s+1,\delta-1}(\Omega_\theta)$ is a group and $D_{s+1,\delta-1}$ is a normal subgroup. In exact parallel with Lemmas 2.3 and 2.4 we can also prove the following lemma.

Lemma 2.6: If $f \in P_{s+1,\delta-1}(\Omega_\theta)$, $s > 2$, $\delta > -2$, then

$$f'(\eta) - \eta \in H_{s,\delta}(\Omega_\theta).$$

Lemma 2.7: If $f \in P_{s+1,\delta-1}(\Omega_\theta)$ and t is a tensor which belongs to $H_{s',\delta'}(\Omega_\theta)$, then $f'(t) \in H_{s',\delta'}(\Omega_\theta)$ for any $s' \leqslant s$, $\delta' \in \mathbb{R}$. These lemmas may now be combined to prove the following theorem.

Theorem 2.5 (asymptotic symmetry): If γ is a symmetric two-tensor and $f \in P_{s+1,\delta-1}(\Omega_\theta)$, $s > 2$, $\delta > -2$, then

$$\gamma - \eta \in H_{s',\delta'}(\Omega_\theta) \Rightarrow f'(\gamma) - \eta \in H_{s',\delta'}(\Omega_{\theta'}),$$

for every $s' \leqslant s$, $\delta' \leqslant \delta$ and some $\theta' \in (0,1/\sqrt{2})$.

Proof: The proof of this theorem is almost exactly the same as Theorem 2.4.

Again, the obvious way to understand this theorem is to be given a metric γ such that $\gamma - \eta \in H_{s',\delta'}(\Omega_\theta)$, $s' > 2$, $\delta' > -2$. This means that $\gamma - \eta$ goes to zero at infinity classically. The class of diffeomorphisms that preserve both the falloff and differentiability of γ include $P_{s+1,\delta-1}(\Omega_\theta)$ for any $s \geqslant s'$, $\delta \geqslant \delta'$. This means the Poincaré group, together with any diffeomorphisms that reduce to the identity at infinity, are continuous enough $(s \geqslant s')$ and fall off rapidly enough $(\delta \geqslant \delta')$. We need the falloff condition because we are not comparing $f'(\gamma)$ with $f'(\eta)$ (this is done in Lemma 2.7), we are comparing $f'(\gamma)$ with the old η, which is why we need Lemma 2.6.

We have two sets of weighted Sobolev spaces, three-dimensional Riemannian ones and four-dimensional pseudo-Riemannian ones. We wish to connect these together, because we wish to consider three-dimensional Riemannian hypersurfaces imbedded in the four-space.

We introduce a natural foliation of the sets Ω_θ induced by the function $\sigma(x)$, which we defined at the beginning of Sec. II B. We define the function

$$\tau(x) = x^0/\sigma(x), \quad x \in \Omega_\theta.$$

This function induces a foliation of Ω_θ:

$$\Omega_\theta = \bigcup_{\tau \in I_\theta} \Sigma_\tau, \quad I_\theta =]-\theta,\theta[.$$

Each Σ_τ, the set with constant τ, is a hypersurface imbedded in Ω_θ, and is a complete $\mathbb{R}^3$ with an induced Riemannian metric. We define the following restriction norm:

$$\|u\|_{H_{s,\delta}(\Sigma_\tau,\Omega_\theta)} = \left(\sum_{\alpha=0}^{s} \|D^\alpha u|_{\Sigma_\tau}\|_{H_{s-\alpha,\delta+\alpha}(\mathbb{R}^3)} \right)^{1/2}.$$

The following restriction lemma was proved (Ref. 12).

Lemma 2.8 (restriction)[12]*:* For each $\tau \in I_\theta$, the following inclusion holds and is continuous:

$$H_{s+1,\delta}(\Omega_\theta) \subset H_{s,\delta+1/2}(\Sigma_\tau,\Omega_\theta),$$

for every $s \in \mathbb{N}$, $\delta \in \mathbb{R}$.

The restriction norm allows us to take a function u on Ω_θ and look at it on Σ_τ. However, the restriction norm not only looks at the derivatives of u along Σ_τ but also its derivatives out of Σ_τ (at Σ_τ). Therefore it is not the same as $H_{s,\delta}(\mathbb{R}^3)$.

In $\mathbb{R}^4$, an $\mathbb{R}^3$ hypersurface has measure zero. Therefore a function could be square integrable on $\mathbb{R}^4$ and yet blow up on a set of nonzero measure on $\mathbb{R}^3$. Therefore an H_0 function on $\mathbb{R}^4$ may not be an H_0 function on $\mathbb{R}^3$. However, there is a Sobolev restriction theorem[13] that one only loses one degree of differentiability. Therefore a H_5 on $\mathbb{R}^4$ (or Ω_θ) induces a H_4 on a hypersurface. This is why $s+1 \to s$.

The $\delta \to \delta + \frac{1}{2}$ can be easily understood by considering ordinary functions with falloff. Say we have a function u on $\mathbf{R}^4$, which falls off faster than $r^{-\alpha}$ for some α. This means that $u \in H_{0,\delta}(\mathbf{R}^4)$ for every $\delta < \alpha - 2$. Now consider u_R the restriction of u to some $\mathbf{R}^3$ hypersurface of $\mathbf{R}^4$. Also u_R will fall off faster than $r^{-\alpha}$ so therefore $u_R \in H_{0,\delta'}(\mathbf{R}^3)$ for every $\delta' < \alpha - \frac{3}{2}$. The volume measure in one case goes like $r^3\,dr$ and in the second case like $r^2\,dr$.

III. THE BOOST PROBLEM

Initial data for the vacuum Einstein equations consist of giving a Riemannian metric g and a two-covariant symmetric tensor field k on $\mathbf{R}^3$. This $\mathbf{R}^3$ is to form the $x^0 = 0$ hypersurface of an $\mathbf{R}^4$ with a pseudo-Riemannian metric γ, which is a solution to the vacuum Einstein equations. The given g and k are to be, respectively, the first and second fundamental forms of $x^0 = 0$. These quantities cannot be given independently, but must satisfy the initial value constraints[7]

$$\nabla_g\{k - (\mathrm{tr}_g\, k)g\} = 0, \tag{3.1}$$

$$R(g) - |k|_g^2 + (\mathrm{tr}_g\, k)^2 = 0. \tag{3.2}$$

We have proven the following theorems.[6]

Theorem 3.1 (the boost theorem)[6]: Let g be a Riemannian metric and k a two-covariant symmetric tensor on a three-manifold Σ. If (g,k) satisfy the initial value constraints and

$$g - e \in H_{s,\delta+1/2}(\Sigma), \quad k \in H_{s-1,\delta+3/2}(\Sigma),$$
$$s \geqslant 4, \quad \delta > -2,$$

then there exists a $\theta > 0$ and a solution γ to the Einstein equations in $\Omega_\theta, \gamma - \eta \in H_{s,\delta}(\Omega_\theta)$ such that g and k are, respectively, the first and second fundamental form of Σ relative to γ, where γ is a regularly hyperbolic metric on Ω_θ.

Theorem 3.2 (completeness of spacelike infinity)[6]: Given an initial data system (Σ,g,k) satisfying the initial value constraints, $g - e \in H_{s,\delta+1/2}(\Sigma)$, $k \in H_{s-1,\delta+3/2}(\Sigma)$, $s \geqslant 4$, $\delta > -2$, and given $\theta < 1/\sqrt{2}$, there exists a finite R and a set Ω_θ^R such that $x \in \Omega_\theta^R \Leftrightarrow x \in \mathbf{R}^4$, $|x| > R$, $|x^0|/\sigma(x) < \theta$. There exists a regularly hyperbolic solution γ to the Einstein equations on Ω_θ^R $\gamma - \eta \in H_{s,\delta}(\Omega_\theta^R)$ such that g and k are, respectively, the first and second fundamental forms of Σ with respect to γ.

Remark: Theorem 3.1 shows that if the initial data is (very weakly) asymptotically flat, the Cauchy extension is large enough to permit a global Lorentz-boosted slice. Theorem 3.2 states that the Cauchy extension is large enough at spacelike infinity to permit slices that are boosted all the way up to (but not including) the light cone.

In proving Theorem 3.1 we showed that the value of θ we got was determined by the size of the initial data, in a suitable norm. This meant that we could only boost by a finite amount. In Theorem 3.2, we cut out a sphere of radius R out of the initial hypersurface, so as to reduce the measure of the initial data to any desired value. The Cauchy development of this data is large enough to contain a domain that is wedge shaped near infinity but is pinched off in the middle (Ω_θ^R).

From the imbedding theorem we get that initial data that satisfies the boost theorem also satisfies $g - e \in C^2_{2+\epsilon}(\mathbf{R}^3)$, $k \in C^1_{1+\epsilon}(\mathbf{R}^3)$ for some $\epsilon > 0$. Near infinity, the Brill–Deser expression[14]

$$\rho = (1/16\pi)\left[\tfrac{1}{4}(g_{ij,k})^2 + (k_{ij})^2\right]$$

accurately represents the energy density in the gravitational field and hence $\rho \sim r^{-(2+\epsilon)}$ near infinity. This means that the total energy of the solution will be infinite, although the solution is asymptotically flat.

When we seek to construct the Cauchy development of an initial data set we can integrate forward at least until we hit our first singularity. The boost theorem tells us that we can integrate forward much further near infinity than in the middle. It is interesting to see how $\rho \sim r^{-(2+\epsilon)}$ is compatible with this behavior. Singularities will form if a large amount of energy is forced into a small volume. Consider a small sphere of radius r_0 at a time t in the future from the original slice and at a radius r. The maximum energy that can collect in this sphere can be estimated by drawing the past light cone of the sphere, and seeing where it intersects the original slice.

This intersection consists of a shell of radius t, thickness r_0 at radius r. The volume of the shell is $4\pi t^2 r_0$ and the energy in it is $4\pi t^2 r_0 \rho(r)$. If we call the energy in the sphere of radius r_0, $E(r_0,r,t)$, we have

$$E(r_0,r,t) \leqslant 4\pi t^2 r_0 \rho(r),$$

at large r we have

$$\rho(r) \leqslant \rho_0 r^{-(2+\epsilon)}.$$

Hence

$$E(r_0,r,t)/r_0 \leqslant 4\pi\rho_0 (t/r)^2 r^{-\epsilon}.$$

Therefore, as we go out along a line of constant t/r (a boosted slice) we get that $E/r_0 \to 0$. But of course E/r_0 finite is the criterion for the formation of a black hole. Therefore, if the energy density in the original slice falls off faster than r^{-2}, we will not get black holes far out along any boosted slice.

IV. THE TOTAL ENERGY MOMENTUM

As is clear from the Introduction we will wish to consider space-times where $(\gamma - \eta)r^{1/2} \to 0$. In Theorems 3.1 and 3.2 (the boost theorems) we require $\delta > -2$. Theorem 2.1 (the imbedding theorem) gives us that if $\gamma - \eta \in H_{4,\delta}(\Omega_\theta)$ for any $\delta > -2$, then $(\gamma - \eta)r^\epsilon \to 0$ for an $\epsilon > 0$. Therefore the boost theorem allows us to consider solutions of the Einstein equations that fall off so slowly at infinity that they must be regarded as infinite energy, but asymptotically flat solutions.

To restrict attention to finite energy solutions we will only consider those solutions for which $\delta > -\frac{3}{2}$ (which we are free to do). This will give us the desired $r^{-1/2}$ falloff and we will assume this condition consistently for the rest of this paper. Hence, let us be given initial data (g,k) for the vacuum Einstein equations on some three-manifold Σ. We assume (g,k) satisfy the constraints. We further assume

$$g - \eta \in H_{s,\delta+1/2}(\Sigma), \quad k \in H_{s-1,\delta+3/2}(\Sigma),$$
$$s \geqslant 4, \quad \delta > -\tfrac{3}{2}.$$

We have shown[6] that a large set of such data exist.

These data will generate a solution γ of the vacuum Einstein equations on a manifold M that is large enough at infinity to permit a complete set of Lorentz boosts. The boost theorems further show that $\gamma - \eta \in H_{s,\delta}$. We will show that we can associate a well-defined, finite, conserved, future-pointing total energy momentum with γ.

The proof involves nothing more than a reworking of the standard approach adopted by Weinberg[2] (Chap. 7.6). He deals with a $1/r$ falloff so the only real care we need to take is to show that various integrals still converge with $r^{-1/2}$ falloff. We write $h = \gamma - \eta$ and identify the part of the Ricci tensor linear in h (raising and lowering with η)

$$R_{\mu\nu}^{(1)} = \tfrac{1}{2}(\partial_\mu \partial_\lambda h^\lambda_\nu - \partial_\lambda \partial_\nu h^\lambda_\mu - \partial_\mu \partial_\nu h^\lambda_\lambda + \eta^{\alpha\beta} \partial_\alpha \partial_\beta h_{\mu\nu}) \tag{4.1}$$

and rewrite the Einstein vacuum equation $G_{\mu\nu} = 0$ as

$$R_{\mu\nu}^{(1)} - \tfrac{1}{2}\eta_{\mu\nu} R_\lambda^{\lambda(1)} = 8\pi t_{\mu\nu}, \tag{4.2}$$

where $t_{\mu\nu}$ represents the terms in the Einstein equation of second or higher order in h. The second-order terms are of the form $h\partial\partial h$ and $(\partial h)^2$. Using the multiplication theorem it can be shown $h\partial\partial h$ and $(\partial h)^2$ belong to $H_{s',\delta'}(M)$ for every $s' \leqslant s-2$, $\delta' < 2\delta - 4$. In particular we can choose $s' = 2$ and $\delta' = 1$. We can prove a similar result for the higher-order terms in t. Thus we show $t \in H_{2,1}(M)$. The imbedding theorem then allows us to conclude that t is a C^0 (classical) function that falls off faster than σ^{-3}.

Remark: This key falloff result could be naively deduced from the fact that

$$hr^{1/2} \to 0, \quad \partial h r^{3/2} \to 0,$$

$$\partial\partial h r^{5/2} \to 0 \Rightarrow [h\partial\partial h + (\partial h)^2] r^3 \to 0.$$

$R_{\mu\nu}^{(1)}$ satisfies the linearized Bianchi identities and hence $t^{\mu\nu}$ is locally conserved:

$$t^{\mu\nu}{}_{,\mu} = 0. \tag{4.3}$$

We now consider some spacelike hypersurface V imbedded in M, and on it we define

$$\mathbf{P}_{(V)} = \int_V t^{\mu\nu} n_\nu \, d^3x, \tag{4.4}$$

where n is the unit timelike normal to V. Notice that there is no $\sqrt{g}$ in the definition of $\mathbf{P}$. We will call $\mathbf{P}_{(V)}$ the total energy momentum associated with V. We will naturally consider V as being one of the flat hypersurface of the given underlying Minkowski space, with metric η. We have that $t^{\mu\nu}$ falls off faster than r^{-3} as a classical function. Therefore the integral in (4.4) will converge and $\mathbf{P}_{(V)}$ is finite.

Remark: This result could also be obtained by starting with $t \in H_{2,1}(M)$ and applying first the restriction lemma and then the inclusion theorem. However, we would require stronger versions of these theorems because we would need to deal with nonintegral degrees of differentiability. In short, we would get

$$t \in H_{2,1}(M) \Rightarrow t \in H_{3/2,3/2}(V) \Rightarrow t \in C^0_{3+\epsilon}(V),$$

for some $\epsilon > 0$.

Having shown that $\mathbf{P}_{(V)}$ is finite, we now wish to prove that $\mathbf{P}_{(V)}$ is a constant irrespective of which flat hypersurface of the underlying Minkowski space is chosen. Let us pick two of them, V_0 and V_1 (V_1 may be boosted relative to V_0). These will enclose a part of Minkowski space and part of the four-manifold. Hence we have a set M^1 bounded by V_0, V_1 and a surface at spacelike infinity S_∞, which has topology $S_2 \times \mathbb{R}$.

Now we use the Gauss theorem on this M^1, regarded as part of Minkowski space

$$\begin{aligned} 0 &= \int_{M'} \partial_\mu t^{\mu\nu} \, d^4x \\ &= \int_{V_0} t^{\mu\nu} n_\mu \, d^3x + \int_{V_1} t^{\mu\nu} n_\mu \, d^3x + \int_{S_\infty} t^{\mu\nu} n_\mu \, d^3x. \end{aligned} \tag{4.5}$$

The normals in the surface integrals will all be pointing out of M'. Therefore, if the normal on V_1 is future pointing, the normal on V_0 will be past pointing. This means we can write (4.5) as

$$\mathbf{P}_{(V_0)} = \mathbf{P}_{(V_1)} + \int_{S_\infty} t^{\mu\nu} n_\mu \, d^3x. \tag{4.6}$$

We can do the integration over S_∞ in two stages, first the two-sphere integral and then the integral over the line. Since t falls off faster than r^{-3}, the two-sphere integral falls off faster than r^{-1}. The real-line interval is either finite (for time translations) or grows linearly with r (for boosts). In either case $\int_{S_\infty} t^{\mu\nu} n_\nu \, d^3x$ vanishes and

$$\mathbf{P}_{(V_0)} = \mathbf{P}_{(V_1)}. \tag{4.7}$$

Thus $\mathbf{P}$ is both finite and slice independent, with respect to a given Minkowski space background.

The final challenge is to show that the energy momentum is invariant with respect to changes of the Minkowski space background. If we make a coordinate transformation which simultaneously acts on γ and η, then the invariance of $\mathbf{P}_{(V)}$ is simply a consequence of the coordinate invariance of the integral. The difficulty arises in the kind of transformation which changes η but not γ, or vice versa.

These kind of transformations arise from the fact (discussed at length in Sec. II) that the underlying Minkowski space is only asymptotically fixed for an asymptotically flat space-time. The arbitrariness in the Minkowski space is represented by the asymptotic symmetry group $P_{s+1,\delta-1}(M)$ (defined in the corollary to Lemma 2.5). We must show that $\mathbf{P}$ is invariant under the action of $P_{s+1,\delta-1}(M)$.

We have a metric γ in some coordinate system such that $\gamma \to \mathrm{diag}(+1,+1,+1,-1)$ at infinity. We define $\eta = \mathrm{diag}(+1,+1,+1,-1)$ everywhere in the same coordinate system. Let f be a member of $P_{s+1,\delta-1}$ and consider $f'\gamma$, the transformation of γ under f. We have that $f'\gamma$ still goes to $\mathrm{diag}(+1,+1,+1,-1)$. We therefore define $\eta' = \mathrm{diag}(+1,+1,+1,-1)$ everywhere in the new coordinate system. Is $\mathbf{P}$ as measured on the flat slices of η' the same as $\mathbf{P}$ measured on the flat slices of η?

This is exactly equivalent to asking whether $\mathbf{P}$ as measured on one slice of M equals $\mathbf{P}$ measured on any other. The only restriction on the choice of slice is that they are related by an element of $P_{s+1,\delta-1}(M)$, $s \geqslant 4$, $\delta > -\tfrac{3}{2}$. If the two slices were such that they could be regarded as the $t=0$ and $t=1$ slice in some coordinates system then (4.7) would give an answer. However, if we had two slices that cross one another several times, we cannot find a coordinate system in which they are the flat slices of a Minkowski space. Further, even if we could find a Minkowski space, it would not be unique.

Therefore, we have to consider the invariance of **P** under several different kinds of transformation. First, given a slice in the manifold M, we can consider it as the $t=0$ slice of many Minkowski spaces. Thus we have to consider elements f of $P_{s+1,\delta-1}(M)$, which leave a given slice unchanged. These f's, at the slice in question, reduce to three-dimensional coordinate transformations on the slice. We also have to consider changes in the slice itself.

Weinberg[2] discusses these kinds of transformations and shows that the neatest way of dealing with them is not to consider **P** as defined by (4.4) but to substitute (4.2) into (4.4) to give

$$\mathbf{P}_{(\nu)} = \frac{1}{8\pi}\int_V \left[R^{(1)\mu\nu} - \frac{1}{2}\eta^{\mu\nu}R^{\lambda(1)}_{\lambda}\right]\eta_\nu\, d^3x. \quad (4.8)$$

He shows that this is a total divergence, which allows us to turn it into a surface integral and obtain the standard ADM energy-momentum formulas[7]

$$P^0 = \frac{1}{16\pi}\oint_\infty (g_{ij,j} - g_{jj,i})dS_i, \quad (4.9)$$

$$P^i = \frac{1}{8\pi}\oint_\infty [k^{ij} - k^m_m g^{ij}]\, dS_j. \quad (4.10)$$

Let us assume $f\in D_{s+1,\delta-1}(M)$, $s\geqslant 4$, $\delta > -\frac{3}{2}$. We know

$$\begin{aligned}\mathscr{L}_f \gamma &= \nabla_\alpha f_\beta + \nabla_\beta f_\alpha \\ &= f_{\alpha,\beta} + f_{\beta,\gamma} - 2f_\gamma \Gamma^\gamma_{\alpha\beta},\end{aligned} \quad (4.11)$$

where $\mathscr{L}$ is the Lie derivative. Let us first deal with the $f\Gamma$ term. We can write it as

$$f\Gamma = (f - \mathrm{id})\Gamma + \mathrm{id}\,\Gamma.$$

On using the multiplication theorem we can show $f\Gamma\in H_{3,\delta'}$ for some $\delta' > -1$. The imbedding theorem then gives $f\Gamma \in C^1_\alpha$ for some $\alpha > 1$. This implies that $f\Gamma$ falls off faster than r^{-1}, and $(f\Gamma)_{,\alpha}$ falls off faster than $1/r^2$. This means that when we wish to calculate the changes in the surface integrals (4.9) and (4.10) due to f, we can ignore the changes due to $f\Gamma$ and need only worry about the changes caused by

$$\delta\gamma_{\mu\nu} = f_{\mu,\nu} + f_{\nu,\mu}.$$

Weinberg[2] does the calculation in detail for such shifts and shows that the extra terms in (4.9) and (4.10) are a total curl and thus the surface integral must vanish, independent of any falloff argument.

This can be easily seen in the case of (4.9) if we consider a three-dimensional coordinate transformation

$$\begin{aligned}\delta g_{ij} &= f_{i,j} + f_{j,i}, \\ \delta P^0 &= \oint_\infty (f_{i,jj} + f_{j,ij} - 2f_{j,ji})dS_i \\ &= \oint_\infty (f_{i,j} - f_{j,i})_{,j}\, dS_i = 0.\end{aligned} \quad (4.12)$$

Therefore, the total energy momentum is invariant under any infinitesimal coordinate transformation that belongs to $D_{s+1,\delta-1}(M)$. However, any finite transformation that is connected to the identity can be regarded as an integral of infinitesimals. Thus the energy momentum is invariant under transformations that belong to $D_{s+1,\delta-1}(M)$.

We have effectively dealt with the Lorentz covariance of the energy momentum already. We have shown that the energy momentum on a given boosted slice of a given Minkowski space is constant. Now make a Lorentz transformation of everything, so that boosted slice becomes the new $t=0$ slice. Then the energy momentum, under this Lorentz transformation, will transform like a Lorentz four-vector. Thus **P** is well behaved under the action of $P_{s+1,\delta-1}(M)$.

V. ADM MASS

In Sec. IV we have shown that the energy momentum is finite and well defined in any space-time that is generated by initial data (g,k) belonging to

$$g - e\in H_{s,\delta}(\Sigma), \quad k\in H_{s-1,\delta+1}(\Sigma), \quad s\geqslant 4, \quad \delta > -1.$$

The argument in Sec. IV was a four-dimensional argument, and only at the very end [in using (4.9) and (4.10)] did any element of a 3 + 1 analysis creep in. It is interesting and enlightening to do the finiteness of energy proof entirely as an initial data problem. This is the motivation behind this section.

On a three-dimensional manifold the imbedding theorem gives

$$g - e\in H_{s,\delta}(\Sigma) \Rightarrow g - e\in C^2_{1/2+\epsilon},$$

for some $\epsilon > 0$, which implies that $g - e$ falls off faster than $r^{-1/2}$, its derivative faster than $r^{-3/2}$, and k falls off faster than $r^{-3/2}$. Traditionally, the ADM energy (4.9) has been identified with the mass of the Schwarzschild solution, which appears as a M/r term in the metric. Further, it appears that the surface integral expressions (4.9) and (4.10) would blow up if either $g_{ij,k}$ or k_{ij} fell off slower than r^{-2}.

We will show that this r^{-2} falloff, which is obtained by directly inspecting the surface integrals, is misleading. Just as in electromagnetism, if we turn the surface integrals into volume integrals, and use the field equations (in this case the constraints) we will show that we can get by with weaker falloff conditions. Thus, we are not looking for pairs (g,k) that give finite integrals (4.9) and (4.10). Rather we are looking for pairs (g,k) that solve the constraints and give us finite (4.9) and (4.10).

In fact, we can strengthen the analogy to electromagnetism by showing that we can identify a unique M/r term in the metric ($M = P^0 = \text{const}$) even when the metric only falls off faster than $r^{-1/2}$. The M/r term (which is the analog of the Newtonian potential) is the only term that contributes to the surface integral (4.9), the rest of the metric, which may fall off slower than $1/r$, appears in the surface integral as a total curl, and so integrates to zero. One way of doing this is by introducing three-harmonic coordinates.

To show the existence of such a coordinate sytem we need to prove a number of new results. In this section we will restrict our attention to Riemannian metrics on a three-manifold (although the theorems can be easily generalized).

Lemma 5.1:

$$g-e\in H_{s,\delta}(\Sigma)\Rightarrow g^{ab}-e\in H_{s,\delta}(\Sigma),$$
$$\text{for any } s\geqslant 2,\ \delta>-\tfrac{3}{2}.$$

Proof: From the corollary to the multiplication theorem (Theorem 2.2) we have

$$g-e\in H_{s,\delta}(\Sigma)\Rightarrow \det g-1\in H_{s,\delta}(\Sigma),\quad \text{if } s\geqslant 2,\ \delta>-\tfrac{3}{2}. \tag{5.1}$$

Since we assume g Riemannian we have $\det g>0$. From the imbedding theorem we have $\det g-1\in C^0_\beta$ for some $\beta>0$. Hence there exists constants C_1,C_2 such that

$$0<C_1\leqslant \det g\leqslant C_2. \tag{5.2}$$

Now $(\det g)g^{ab}=$ cofactor of g_{ab}. Hence from the corollary to the multiplication theorem

$$g-e\in H_{s,\delta}(\Sigma)\Rightarrow(\det g)g^{ab}-e\in H_{s,\delta}(\Sigma),$$
$$\text{if } s\geqslant 2,\quad \delta>-\tfrac{3}{2}. \tag{5.3}$$

Together (5.1)–(5.3) prove the lemma. The following theorems have been proven about the Laplacian.

Theorem 5.1 (McOwen)[15] (the flat space Laplacian): The map

$$\Delta_e:\ H_{s,\delta}(\mathbb{R}^3)\rightarrow H_{s-2,\delta+2}(\mathbb{R}^3),\quad s\geqslant 2,$$

(a) is an isomorphism if $-\frac{3}{2}<\delta<-\frac{1}{2}$; (b) is an injection with closed range given by

$$R_1=\left\{f\in H_{s-2,\delta+2}(\mathbb{R}^3),\quad \int f\,d\mu(e)=0\right\}\quad \text{if } -\tfrac{1}{2}<\delta<\tfrac{1}{2};$$

and (c) is a surjection with kernel equal the constant functions if

$$-\tfrac{5}{2}<\delta<-\tfrac{3}{2}.$$

Theorem 5.2a (the curved space Laplacian)[6,10]: Let g be a Riemannian metric on $\mathbb{R}^3$ such that $g-e\in H_{s',\delta'}(\mathbb{R}^3)$, $s'\geqslant 2$, $\delta'>-\frac{3}{2}$. Then Δ_g (acting on scalar functions) is an isomorphism $H_{s,\delta}\rightarrow H_{s-2,\delta+2}$ for each $2\leqslant s\leqslant s'+1$, $-\frac{3}{2}<\delta<-\frac{1}{2}$.

Theorem 5.2a is clearly the direct analog of Theorem 5.1(a). They both say that if the source falls off between $1/r^2$ and $1/r^3$, the potential falls off at infinity but slower than $1/r$ (and vice versa). The extra bits in 5.2a are there to deal with the $\Gamma\phi_{,a}$ term in $\Delta_g\,\phi$. In this article we will prove the theorems equivalent to 5.1(b) and 5.1(c). Theorem 5.1(b) says that if the potential falls off faster than $1/r$, then the source must fall off faster than $1/r^3$ and its integral must vanish. Theorem 5.1(c) deals with slowly falling-off potentials and sources. It says that if the potential blows up slower than r, the source falls off faster than $1/r$. However, we can always add a constant to the potential.

Theorem 5.2b: Let g ge a Riemannian metric on $\mathbb{R}^3$ such that $g-e\in H_{s',\delta'}(\mathbb{R}^3)$, $s'\geqslant 3$, $\delta'>-\frac{3}{2}$. Then Δ_g: $H_{s,\delta}(\mathbb{R}^3)\rightarrow H_{s-2,\delta+2}(\mathbb{R}^3)$ is an injection with closed range given by

$$R_1(g)=\left\{f\in H_{s-2,\delta+2},\ \int f\,d\mu(g)=0\right\},$$
$$\text{if}\quad 2\leqslant s\leqslant s'+1,\quad -\tfrac{1}{2}<\delta<\tfrac{1}{2}.$$

Proof: Let us be given $f\in R_1(g)$. Theorem 5.2a tells us we can solve $\Delta_g\,\phi=f$ with $\phi\in H_{s,\bar\delta}$ for every $\bar\delta<-\frac{1}{2}$. Further $\oint_\infty \nabla_g\,\phi\cdot dS=0$ since $f\in R_1(g)$. From the imbedding theorem we have that $g-e\in C^0_\beta$ for some $\beta>0$. Hence

$$\oint_\infty \sqrt{g}\phi_{,a}\,dS^a=\oint_\infty \phi_{,a}\,dS^a=0. \tag{5.4}$$

Now

$$\Delta_g\,\phi=\Delta_e\phi+(g^{ab}-e)\partial_a\partial_b\,\phi-g^{ab}\Gamma^c_{ab}\,\phi_{,c}, \tag{5.5}$$

$$\Rightarrow\Delta_e\,\phi=f-(g^{ab}-e)\partial_a\partial_b\,\phi+g^{ab}\Gamma^c_{ab}\,\phi_{,c}. \tag{5.6}$$

The multiplication theorem shows that the right-hand side of (5.6) belongs to $H_{s-2,\bar{\bar\delta}+2}$ for some $\bar{\bar\delta}>-\frac{1}{2}$. Further, the right-hand side of (5.6) belongs to $R_1(e)$ because of (5.4). Therefore Theorem 5.1(b) gives us that $\phi\in H_{s,\bar{\bar\delta}}$ for some $\bar{\bar\delta}>-\frac{1}{2}$. Thus we have improved the falloff on ϕ from $<-\frac{1}{2}$ to $>-\frac{1}{2}$. Now we can iterate on Eq. (5.6), by substituting the faster falloff ϕ into the right-hand side. Since $g-e\in H_{s',\delta'}$, $\delta'>-\frac{3}{2}$, there exists an $\epsilon>0$ so that $g-e\in H_{s',-3/2+2\epsilon}$. When we substitute this, together with the fact that $\phi\in H_{s,\bar{\bar\delta}}$ into the right-hand side of (5.6), we can prove it belongs to $H_{s-2,\gamma+2}$ where $\gamma=\min(\delta,\bar{\bar\delta}+\epsilon)$. So long as $\bar{\bar\delta}<\delta$ we gain an extra ϵ every time round, to finally conclude $\phi\in H_{s,\delta}(\mathbb{R}^3)$.

Theorem 5.2c: Let g be a Riemannian metric on $\mathbb{R}^3$ with $g-e\in H_{s',\delta'}(\mathbb{R}^3)$, $s'\geqslant 3$, $\delta'>-\frac{3}{2}$. Then Δ_g: $H_{s,\delta}\rightarrow H_{s-2,\delta+2}(\mathbb{R}^3)$ is a surjection with kernel equal the constant functions if $2\leqslant s\leqslant s'+1$, $-\frac{5}{2}<\delta<-\frac{3}{2}$.

Proof: Let us be given an $f\in H_{s-2,\delta+2}$ with $2\leqslant s\leqslant s'+1$, $-\frac{5}{2}<\delta<-\frac{3}{2}$. Theorem 5.1(c) tells us that there exists a $\phi_0\in H_{s,\delta}$, which solves

$$\Delta_e\,\phi_0=f. \tag{5.7}$$

Now consider

$$\Delta_e\,\phi_1=-(g-e)\partial\,\partial\phi_0+g\Gamma\phi_0. \tag{5.8}$$

We know there exists an $\epsilon>0$ such that $g-e\in H_{s',-3/2+2\epsilon}(\mathbb{R}^3)$. The multiplication theorem now can be used to show that the right-hand side of (5.8) belongs to $H_{s-2,\delta+2+\epsilon}$. Theorem 5.1(c) tells us now that we can solve (5.8) for ϕ_1 belonging to $H_{s,\delta+\epsilon}$. We now iterate solving

$$\Delta_e\,\phi_2=-(g-e)\partial\partial\phi_1+g\Gamma\partial\phi_1, \tag{5.9}$$

for $\phi_2\in H_{s,\delta+2\epsilon}$. We keep iterating until we get $\phi_n\in H_{s,\delta+n\epsilon}$ with $\delta+n\epsilon>-\frac{3}{2}$. We now solve

$$\Delta_g\,\phi_{n+1}=-(g-e)\partial\partial\phi_n+g\Gamma\partial\phi_n. \tag{5.10}$$

The right-hand side of (5.10) now belongs to $H_{s-2,1/2+\epsilon}$. Theorem 5.2a now guarantees that we can solve (5.10) for $\phi_{n+1}\in H_{s,-3/2+\epsilon}$. Summing (5.7)–(5.10) now gives us

$$\Delta_g(\phi_0+\phi_1+\cdots+\phi_{n+1})=f, \tag{5.11}$$

with, of course, $\phi_0+\phi_1+\cdots+\phi_{n+1}\in H_{s,\delta}$. Hence Δ_g is a surjection. It is clear from the construction [and Theorem 5.1(c)] that the kernel is the constant functions.

To return to the problem at hand, we have that the initial data (g,k) are not independent, they must satisfy the constraints. The constraints are elliptic equations, the exact analog of the Poisson equation in electromagnetism. The surface integral expressions (4.9) and (4.10) are locked into these elliptic equations. They detect only the energy and momentum potentials (the dependent parts of the elliptic equations) and ignore the independent parts of the initial data.

Expression (4.9) expresses P^0 just as a function of the initial data

$$P^0=\frac{1}{16\pi}\oint_\infty \partial^{ij}\partial^{kl}(g_{ik,j}-g_{ij,k})dS_l. \qquad (5.12)$$

We have $g_{ab}-e$, $\det g-1$, and $g^{ab}-e\in C^2_{1/2+\epsilon}$ for some $\epsilon>0$. Therefore in (5.12) we can replace $\partial^{ij}\partial^{kl}$ with $\sqrt{g}g^{ij}g^{kl}$, without changing the value of the integral, to give

$$P^0=\frac{1}{16\pi}\oint_\infty \sqrt{g}g^{ij}g^{kl}(g_{ik,j}-g_{ij,k})dS_l. \qquad (5.13)$$

When this is turned into a volume integral, using the Gauss theorem, we get a term

$$g^{ij}g^{kl}(g_{ik,jl}-g_{ij,kl}),$$

plus other, first metric derivative terms. This is the second derivative term in the scalar curvature ${}^{(3)}R$. Thus the volume integral of (5.13) can be written as

$$P^0=\frac{1}{16\pi}\int_\Sigma \sqrt{g}\,\{{}^{(3)}R+\tfrac{1}{4}g^{mn}g^{ab}g^{cd}(2g_{mn,d}\,g_{ac,b}-2g_{ma,d}\,g_{nc,b}+g_{dm,b}\,g_{nc,a}-g_{cd,a}\,g_{mn,b})\}d^3x. \qquad (5.14)$$

On using the initial value constraint

$${}^{(3)}R=k\cdot k-(\operatorname{tr}k)^2,$$

this can be written as

$$P^0=\frac{1}{16\pi}\int_\Sigma \sqrt{g}\,\{k\cdot k-(\operatorname{tr}k)^2+\tfrac{1}{4}g^{mn}g^{ab}g^{cd}(2g_{mn,d}\,g_{ac,b}-2g_{ma,d}\,g_{nc,b}+g_{dm,b}\,g_{nc,a}-g_{cd,a}\,g_{mn,b})\}d^3x. \qquad (5.15)$$

The multiplication theorem immediately guarantees that this is finite. An exactly equivalent calculation shows that the momentum constraint $\nabla_a(k^{ab}-g^{ab}k^c_c)=0$ guarantees the finiteness of P^i as defined by (4.10). Thus, the constraints guarantee that the energy momentum, defined on the initial slice, is finite.

To show that the energy momentum is well defined we would have to calculate the change of P^0 and P^i due to changing the hypersurface. This means using the Einstein dynamical equations with essentially arbitrary lapse N and shift N_2 functions to calculate the rate of change of g and k in propagating off the hypersurface. The only real restrictions we would place on N and N_2 is that they asymptotically blow up slower than $r^{1/2}$. This is equivalent to demanding that the coordinate transformation belongs to $D_{s+1,\delta-1}(\mathbb{R}^4)$. It follows from this [see the part of Sec. IV after (4.9)] that the energy momentum is conserved.

To extend to the Poincaré group we have to consider

$$N\sim\alpha x+\beta y+\gamma z+O(r^{1/2-\epsilon}),$$

where α,β,γ are constants (a general boost), and

$$N_a\sim A(0,-z,+y)+B(z,0,-x)+C(-y,+x,0)+O(r^{1/2-\epsilon}),$$

with A, B, and C constant (a general rotation). The energy momentum P^μ transforms in exactly the correct way.

The constraint equations not only force the energy momentum to be finite, but also allow us to identify part of the metric to be a potential term (analogous to the Newtonian potential). The independent, freely specifiable parts of the metric[16] can fall off like $r^{-1/2-\epsilon}$ but the dependent part (the potential) must fall off like $1/r$. Not only does it fall off like $1/r$, it must be of the form P^0/r, where P^0 is a constant, the total energy.

This breakup of the metric into independent and dependent parts is related to the choice of coordinates. One natural choice of coordinates is a three-harmonic coordinate system. The three-harmonic coordinate condition is that

$$\Delta_g x^a=0,$$

for each coordinate x^a. We wish to prove that on all asymptotically flat three-manifolds we can choose (asymptotic) three-harmonic coordinates.

Let us be given nonharmonic coordinates X' and we seek a coordinate transformation $X'\to X=\xi(X')$ such that X is harmonic. This is equivalent to

$$\Delta_g X=0,$$

which implies

$$\Delta_{g'}X=0=\Delta_{g'}\xi(X'). \qquad (5.16)$$

Let us set $\xi^a=X^{a'}+f^a$. Equation (5.16) then becomes

$$\Delta_{g'}f^a=\Gamma^a. \qquad (5.17)$$

If $g'-e\in H_{s,\delta}(\Sigma)$, $s\geqslant-4$, $\delta>-\frac{3}{2}$ then $\Gamma\in H_{s-1,\delta+1}$. Theorem 5.2c tells us that there exists an $f^a\in H_{s+1,\delta-1}(\Sigma)$, which solves (5.17) and is unique up to constants.

The imbedding theorem tells us that $Df\in C^0_\beta$ $(\beta>0)$. We can find a C^∞ non-negative function $\theta_R\leqslant1$ with the following properties on Σ:

$$\theta_R(x)=0,\quad \forall|x|<R,$$

$$\theta_R(x)=1,\quad \forall|x|>2R,$$

and in the region between R and $2R$, $|D\theta_R|<2/R$. Hence, there exists an R_0 such that $|\det D(\theta_{R_0}f)|<1$. We cannot generate a global coordinate transformation with f because Df may be too large somewhere and the Jacobian may go negative. However, since Df falls off we can multiply f by a smooth function θ_{R_0}, which eliminates f in the center and leaves f unchanged in the exterior region. This guarantees that the transformation

$$\bar{\xi}: x'\to x=x'+\theta_{R_0}f$$

is a diffeomorphism. In fact, $\bar{\xi}\in D_{s+1,\delta-1}(\Sigma)$ so the energy momentum is unchanged. Further, outside of $2R_0$, the coordinates X are three-harmonic. Hence we have proved the following theorem.

Theorem 5.3 (three-harmonic coordinates): Given a three-manifold Σ with Riemannian metric g such that $g-e\in H_{s,\delta}(\Sigma)$, $s\geqslant4$, $\delta>-\frac{3}{2}$, there exists a diffeomorphism $\xi\in D_{s+1,\delta-1}$ and a constant C such that $\xi'g$ is three-harmonic outside a sphere of radius C.

The three-harmonic coordinate condition $\Delta_g x=0$ reduces to

$$\Gamma^a=g^{bc}\Gamma^a_{bc}=g^{bc}g^{ad}(g_{bd,c}-\tfrac{1}{2}g_{bc,d})=0.$$

We can now use this condition to simplify the ADM energy expression (5.13). It now can be written

$$P^0 = -\frac{1}{32\pi}\oint_\infty g^{1/2}g^{ab}g^{cd}g_{ab,c}\,dS_d$$
$$= -\frac{1}{16\pi}\oint_\infty (g^{1/2})_{,c}\,dS^c. \tag{5.18}$$

Hence, we expect $g^{1/2}$ to be the mass potential and to go like $g^{1/2}\sim 1 + 4P_0/r$, even when $g_{ab}\sim\delta_{ab} + O(r^{-1/2-\epsilon})$.

In harmonic coordinates, we can show that $R_{ab} = -\frac{1}{2}g^{mn}g_{ab,mn} + P_{ab}$, where P_{ab} is quadratic in $g_{ab,c}$. Hence

$$R = g^{ab}R_{ab} = -\tfrac{1}{2}g^{ab}g^{mn}g_{ab,mn} + g^{ab}P_{ab}. \tag{5.19}$$

We also have

$$g^{mn}\,\partial_m\,\partial_n\,g^{1/2} = \tfrac{1}{2}g^{1/2}\{g^{ab}g^{mn}g_{ab,mn} + Q\}. \tag{5.20}$$

Combining (5.19) and (5.20), we have

$$g^{1/2}\,{}^{(3)}R = -g^{mn}\,\partial_m\,\partial_n\,g^{1/2} + S,$$

where S is quadratic in $g_{ab,c}$. Finally, the initial value constraint gives

$$-g^{mn}\,\partial_m\,\partial_n\,g^{1/2} = g^{1/2}[k\cdot k - (\operatorname{tr} k)^2] - S$$

or (5.21)

$$-\partial_m(g^{mn}\,\partial_n\,g^{1/2}) = g^{1/2}[k\cdot k - (\operatorname{tr} k)^2] - S',$$

where S' is quadratic in $g_{ab,c}$. Equations (5.18) and (5.15) are compatible because the right-hand side of (5.21) is identical to the expression (5.15). The multiplication theorem and Lemma 5.1 show that it belongs to $H_{3,3/2+\epsilon}(\Sigma)$ for some $\epsilon > 0$.

Naively, Eq. (5.21) looks like $\Delta^2\phi = \rho$, where $\phi = g^{1/2}$, and $\rho\sim r^{-3-\epsilon}$. We are seeking a solution with $\phi\to 1$ at ∞. We expect a solution $\phi = 1 - \alpha/4\pi r + O(r^{-1-\epsilon})$, where $\alpha = \int\rho$. The theorems we have in terms of weighted Sobolev spaces, which deal with fast falloff sources [Theorems 5.1(b) and 5.2b], deal with the situation where we eliminate the α/r term by demanding that $\int\rho\,d^3x = 0$. The trick is to take out the α/r term first by hand and then apply the theorems to the residue.

To that end, it is easy to show

$$\partial_m(g^{mn}\,\partial_n\sigma^{-1})\in H_{3,2}(\Sigma).$$

Now consider

$$\partial_m\left[g^{mn}\,\partial_n\left\{g^{1/2}\left(1 - \frac{4P^0}{\sigma}\right) - 1\right\}\right]$$
$$= -g^{1/2}\,\partial_m\left[g^{mn}\,\partial_n\left(\frac{4P^0}{\sigma}\right)\right]$$
$$- 2g^{mn}\,\partial_n\,g^{1/2}\,\partial_m\left(\frac{4P^0}{\sigma}\right)$$
$$+ \left(1 - \frac{4P^0}{\sigma}\right)\partial_m\,[g^{mn}\,\partial_n\,g^{1/2}]. \tag{5.22}$$

From (5.21) the third term on the right-hand side of (5.22) belongs to $H_{3,3/2+\epsilon}$. The other two terms belong to $H_{3,2}$. Thus the sum belongs to $H_{3,3/2+\epsilon}(\Sigma)$. Now consider

$$\oint_\infty g^{mn}\,\partial_n\left\{g^{1/2}\left(1 - \frac{4P^0}{\sigma}\right) - 1\right\}dS_m.$$

This is zero because of (5.18). Hence the right-hand side of (5.22) not only belongs to $H_{3,3/2+\epsilon}$ but also to $R_1(g)$. Therefore, we can apply Theorem 5.2b to have that

$$g^{1/2}(1 - 4P^0/\sigma) - 1\in H_{5,-1/2+\epsilon}(\Sigma)$$

or (5.23)

$$g^{1/2} = 1 + 4P^0/\sigma + h,$$

where $h\in H_{5,-1/2+\epsilon}$. Thus h is a classical function that falls off faster than $1/r$. Notice that not only does the constraint (in three-harmonic coordinates) force $g^{1/2}$ to fall off more rapidly than g_{ab}; it also forces it to be smoother.

This result, that in three-harmonic coordinates the determinant of g acts as the gravitational potential and falls off like P^0/r, is not unique to this coordinate choice. For example, another coordinate choice is the so-called TT coordinates, where the metric is written in the form

$$g_{ab} = \delta_{ab} + \theta\delta_{ab} + h^{\mathrm{TT}}_{ab},$$

where h^{TT}_{ab} is both divergence-free and trace-free. Everything we have proven for three-harmonic coordinates can also be proven for TT coordinates; they exist (asymptotically); the energy expression (5.12) reduces to

$$P^0 = -\frac{1}{8\pi}\oint_\infty \nabla\theta\cdot d\mathbf{S}, \tag{5.24}$$

and the constraint reduces to

$$\nabla^2\theta = \rho,\quad \rho\in H_{3,3/2+\epsilon},$$

which implies $\theta = 2P^0/r + O(r^{-1-\epsilon})$. Therefore, in the TT coordinates, θ, which can be regarded as either the trace of $g_{ab} - \delta_{ab}$ or as a conformal factor, acts as the Newtonian potential.

VI. POSITIVITY OF MASS

One of the long-outstanding problems related to the ADM mass was whether it was positive. Recently, two proofs of the positivity of mass have been obtained, one by Witten,[5] one by Schoen and Yau.[3,4] Both the Witten proof and the Schoen and Yau proof explicitly assume that the initial data is such that the metric falls off like $1/r$, and the extrinsic curvature falls off like $1/r^2$. In this section we wish to show that the ADM mass associated with the more slowly falling-off initial data we have been dealing with in this paper,

$$g - e\in H_{s,\delta}(\Sigma),\quad k\in H_{s-1,\delta+1}(\Sigma),\quad s\geqslant 4,\quad \delta > -1, \tag{6.1}$$

is positive definite.

This result, when combined with the boost theorem, says that the ADM energy is finite and positive in every Lorentz frame. This is sufficient to show that the total energy momentum is timelike and future pointing.

The way we approach the problem is by redoing the Schoen and Yau proof, but being careful to assume only that the initial data satisfy (6.1) rather than the more restrictive Schoen and Yau assumptions. The Schoen and Yau proof breaks up naturally into a set of independent stages.

(A) If we are given nonmaximal initial data, i.e., data with $g^{ab}k_{ab}\neq 0$, we cannot assume that the three-scalar curvature ${}^{(3)}R$ is positive because the constraint (3.2) only gives ${}^{(3)}R = k\cdot k - (\operatorname{tr} k)^2$. The first stage is to eliminate the tr k

term to find a three-metric with positive scalar curvature, with the same ADM energy.

(B) Given a metric with positive scalar curvature (which we are guaranteed if tr $k=0$), then we can show the existence of a conformally related three-manifold with zero scalar curvature and less energy. Of course, any three-manifold with zero scalar curvature is a solution to the constraints with $k_{ab}=0$. Physically, what we are saying is that if we eliminate the extrinsic curvature we lower the energy.

(C) Now one eliminates the independent parts of the metric outside a region of compact support and further lowers the energy. We are then left with a three-manifold with zero scalar curvature, which is conformally flat outside a region of compact support, with less energy. We cannot have it flat outside the region of compact support because we cannot eliminate the Newtonian potential. However, the new metric is Schwarzschildean at infinity.

(D) Schoen and Yau finally show that all solutions that are Schwarzschildean outside a region of compact support have positive mass.

The detailed falloff conditions play relatively little role in part (A) and the differences have been eliminated by part (D), so in this section we will only work through parts (B) and (C).

Part (A): Let us be given an initial data set (g'',k'') satisfying both the constraints (3.1) and (3.2) and (6.1). From the constraints (3.2), the multiplication theorem (Theorem 2.2) shows us that $R(g'')\in H_{s',\delta'}(\Sigma)$, $s'\geqslant 3$, $\delta'>\frac{3}{2}$. [The imbedding theorem gives us k'' falling off faster than $r^{-3/2}$, and since $R\sim k^2$, we have that R falls off faster than r^{-3}, whereas, reading directly from (6.1) we would only get R falling off faster than $r^{-5/2}$).] Schoen and Yau[4] show how to construct a metric g', $g'-e\in H_{s,\delta}(\Sigma)$, $s\geqslant 4$, $\delta>-1$, such that $R(g')\geqslant 0$, that $R(g')\in H_{s',\delta'}(\Sigma)$, $s'\geqslant 3$, $\delta'>\frac{3}{2}$, and that the ADM energy associated with g' equals the ADM energy of g''.

Part (B): Let us be given a metric g' such that $g'-e\in H_{s,\delta}(\Sigma)$, $s\geqslant 4$, $\delta>-1$, such that $R(g')\geqslant 0$ and $R(g')\in H_{s',\delta'}$, $s'\geqslant 3$, $\delta'>\frac{3}{2}$. This can either come from part (A), or directly from an initial data set which satisfies $|k'|^2_{g'}\geqslant(\mathrm{tr}_{g'}\,k')^2$.

We have already proven one relevant result for metrics satisfying the above conditions (Lemma 3.2 of Ref. 6).

Lemma 6.1: Let g' be a Riemannian metric on $\mathbb{R}^3$ such that $g'-e\in H_{s,\delta}$, $s\geqslant 4$, $\delta>-\frac{3}{2}$, and $R(g')\geqslant 0$. Then there exists a unique Riemannian metric g, conformally equivalent to g', such that $g-e\in H_{s,\delta}$, $R(g)=0$.

If we call $P^0(g)$ the value of the ADM energy integral (5.13) associated with a metric g, we can prove the following lemma.

Lemma 6.2: Let g' be a Riemannian metric on $\mathbb{R}^3$ such that $g'-e\in H_{s,\delta}$, $s\geqslant 4$, $\delta>-1$, with $R(g')\geqslant 0$, $R(g')\in H_{s',\delta'}$, $s'\geqslant 3$, $\delta'>\frac{3}{2}$. Let g be the unique Riemannian metric conformally equivalent to g' with $R(g)=0$. Then

$$P^0(g)\leqslant P^0(g')<\infty.$$

Proof: $P^0(g')$ is finite, this follows from (5.14). Lemma 6.1 guarantees the existence of a positive function ϕ, $\phi-1\in H_{s,\delta}$, where ϕ is the unique solution to

$$8\Delta_{g'}\,\phi-R(g')\phi=0,\quad \phi\to 1\quad\text{at }\infty. \tag{6.2}$$

Then $g=\phi^4 g'$ satisfies $R(g)=0$, $g-e\in H_{s,\delta}$. On substituting this expression into (5.13) we can easily show

$$P^0(g')-P^0(g)=\frac{1}{16\pi}\oint_\infty 8\nabla\phi\cdot d\mathbf{S}.$$

Turning the surface integral into a volume integral gives

$$P^0(g')-P^0(g)=\frac{1}{16\pi}\int 8\Delta_{g'}\,\phi\,d\mu(g')$$
$$=\frac{1}{16\pi}\int R(g')\phi\,d\mu(g'). \tag{6.3}$$

The integrand in (6.3) is positive, and the integral is finite, thus proving the lemma. This result was originally derived by Ó Murchadha and York.[17] The major difference is that we deal with slow falloff data. Hence, we have shown that eliminating the extrinsic curvature must reduce the energy.

Part (C): Let us be given a Riemannian metric g, $g-e\in H_{s,\delta}$, $s\geqslant 4$, $\delta>-1$, with $R(g)\equiv 0$. This trivially satisfies the constraints (with $k=0$) and is called a moment of time symmetry solution. Now we will follow Schoen and Yau[3] and show that there exists an asymptotically Schwarzschildean metric $\bar g$ with $R(\bar g)=0$ and $P^0(\bar g)\approx P^0(g)$. More precisely, given any $\epsilon>0$, there exists a Riemannian metric $\bar g$ [with $R(\bar g)=0$] and a constant α with

$$\bar g-(1+\alpha/2\sigma)^4\delta_{ij}\in H_{s,\delta'},\quad \delta'>-\tfrac{1}{2}, \tag{6.4}$$

and

$$|P^0(g)-\alpha|<\epsilon. \tag{6.5}$$

The imbedding theorem says that $\bar g-(1+\alpha/2\sigma)^4\delta$ falls off faster than $1/r$. Further, only the $1/r$ part of $\bar g$ will contribute to the ADM energy integral (5.13) and it is easy to show $P^0(g)=\alpha$. In fact, not only does $\bar g-(1+\alpha/2\sigma)^4\delta$ fall off faster than r^{-1}, it actually falls off like r^{-2}.

The technique involves taking the given metric g and multiplying it by a suitable cutoff function so that it becomes flat outside some large radius R. Now the resultant metric is conformally transformed into one with vanishing scalar curvature. Since we have only eliminated the gravitational waves outside R, the total energy of the new solution will not differ much from the total energy of the original solution.

Let us choose a family of smooth cutoff functions $\zeta_\theta(\sigma)$ with $\zeta_\theta\in C^\infty$ such that

$$\zeta_\theta(\sigma)=1,\quad \sigma<\theta,$$
$$\zeta_\theta(\sigma)=0,\quad \sigma>2\theta,$$
$$0\leqslant\zeta_\theta\leqslant 1,$$
$$|D\zeta_\theta|\leqslant C(1+\sigma)^{-1},$$
$$|DD\zeta_\theta|\leqslant C(1+\sigma)^{-2},$$

for some constant C. We now define a modified metric

$$g^\theta=\left(1+\frac{P^0(g)}{2r}\right)^4 e+\zeta_\theta\left\{g-\left(1+\frac{P^0(g)}{2r}\right)^4 e\right\}. \tag{6.6}$$

When θ is large enough g^θ will be Riemannian. It is clear that

$$g^\theta=g,\quad r<\theta,$$
$$g^\theta=(1+P^0(g)/2r)^4 e,\quad r>2\theta,$$

$$R(g^\theta)=0,\quad r<\theta \quad \text{(from definition of } g\text{)},$$
$$R(g^\theta)=0,\quad r>2\theta \quad \text{(from Schwarzschildean form of } g\text{)}, \tag{6.7a}$$
and the key result
$$|R(g^\theta)|\leqslant C_1|1+r|^{-(5/2+\epsilon_1)},\quad \theta\leqslant r\leqslant 2\theta,\quad \text{for some } \epsilon_1>0. \tag{6.7b}$$
To see this we write (6.6) as
$$g^\theta=\zeta_\theta(g-e)+O(1/r).$$
Then
$$g^\theta_{,ab}=\zeta_{\theta,ab}(g-e)+\zeta_{\theta,a}(g-e)_{,b}+\zeta_{\theta,b}(g-e)_{,a}+\zeta_\theta(g-e)_{,ab}+O(1/r^3),$$
a power-counting argument shows that each of these terms falls off faster than $r^{-5/2}$. In particular we now can show
$$\left[\int_{\mathbf{R}^3}|R(g^\theta)|^{3/2}\,d\mu(g^\theta)\right]^{2/3}\leqslant C_2\theta^{-1/2} \tag{6.8}$$
$$\left[\int_{\mathbf{R}^2}|R(g^\theta)|^{6/5}\,d\mu(g^\theta)\right]^{5/6}\leqslant C_3\theta^{-\epsilon},\quad \text{for some } \epsilon>0, \tag{6.9}$$
where C_2 and C_3 are constants independent of θ.

It is not clear from (6.7b) that $\int|R(g^\theta)|d\mu(g^\theta)$ remains finite as $\theta\to\infty$. It seems to blow up like $\theta^{1/2}$. Nevertheless we can still prove that
$$\int_{\mathbf{R}^3}R(g^\theta)d\mu(g^\theta)\to 0\quad \text{as}\quad \theta\to\infty. \tag{6.10}$$
The way to prove this is to realize that the ADM mass associated with g^θ is $P^0(g)$, the ADM mass associated with g. Consider the ADM surface integral in the form (5.13), and evaluate it on the surface $r=2\theta$, it is easy to show
$$\frac{1}{16\pi}\oint_{r=2\theta}\sqrt{g^\theta}g^\theta g^{\theta^{ij}}g^{\theta^{kl}}(g^\theta_{ik,j}-g^\theta_{ij,k})dS_l=\left(1+\frac{P^0(g)}{4\theta}\right)P^0(g).$$
Turn this surface integral into a volume integral just like (5.14) to give
$$\left(1+\frac{P^0(g)}{4\theta}\right)P^0(g)=\frac{1}{16\pi}\int_{r<2\theta}[R(g^\theta)+\tfrac14 g^{\theta^{mn}}\cdots(2g^\theta_{mn,d}\,g^\theta_{ac,b}+\cdots)]d\mu(g^\theta). \tag{6.11}$$
We also have the expression for $P^0(g)$ in terms of g:
$$P^0(g)=\frac{1}{16\pi}\int_{\mathbf{R}^3}[R(g)+\tfrac14 g^{mn}\cdots(2g_{mn,d}\,g_{ac,b}+\cdots)]d\mu(g)=\frac{1}{16\pi}\int_{\mathbf{R}^3}\left[\frac14 g^{mn}\cdots(2g_{mn,d}\,g_{ac,b}+\cdots)\right]d\mu(g). \tag{6.12}$$
Subtract (6.12) from (6.11), remembering that $g=g^\theta$ inside $r=\theta$. We then get
$$[P^0(g)]^2/4\theta=\frac{1}{16\pi}\int R(g^\theta)d\mu(g^\theta)+\frac{1}{64\pi}\int_{r=\theta}^{r=2\theta}g^{\theta^{mn}}\cdots(2g^\theta_{mn,d}\,g^\theta_{ac,b}+\cdots)d\mu(g^\theta)-\frac{1}{64\pi}\int_{r=\theta}^{\infty}g^{mn}\cdots(2g_{mn,d}\,g_{ac,b}+\cdots)d\mu(g). \tag{6.13}$$
It is easy to see that the last two expressions in (6.13) fall off like $\theta^{-\epsilon}$ for some $\epsilon>0$ as $\theta\to\infty$. The left-hand side falls off like θ^{-1} and hence the integral of $R(g^\theta)$ must also fall off as $\theta^{-\epsilon}$.

We have written down a metric g^θ, which is asymptotically Schwarzschildean. However, $R(g^\theta)$, although small, is not identically zero. We now conformally transform g^θ into an asymptotically flat metric $\bar g^\theta$ that satisfies $R(\bar g^\theta)=0$. This means that we seek a conformal factor χ, which satisfies an equation similar to (6.2):
$$8\Delta_{g^\theta}\chi-R(g^\theta)\chi=0,\quad \chi\to 1\quad \text{at}\quad \infty. \tag{6.14}$$
Hence $\bar g^\theta=\chi^4 g$ satisfies $R(\bar g^\theta)\equiv 0$.

Since $R(g^\theta)$ has compact support χ asymptotically will look like $\chi=1+\beta/2r$, where β is some constant. Hence $\bar g^\theta$ will be asymptotically Schwarzschildean. In Lemma (6.1), where we proved the existence of suitable solutions to (6.2), we needed that $R(g')\geqslant 0$. However, we have no guarantee that $R(g^\theta)\geqslant 0$. However, we do have that $R(g^\theta)$ is small, and that is sufficient to prove that a suitable solution exists. Further, since $R(g^\theta)$ is small, the deviation of χ from unity will be small, and the value of β will be small. We have
$$\bar g^\theta=\chi^4 g^\theta=(1+[P^0(g)+\beta]/2r)^4 e\quad \text{at}\quad \infty.$$
Hence
$$P^0(\bar g^\theta)=P^0(g)+\beta.$$
The trick then is to choose θ large enough so that β is less than the ϵ mentioned in expression (6.5). We use the smallness of $R(g^\theta)$ twice, and, happily, (6.8)–(6.10) are the norms in which we want $R(g^\theta)$ to be small. Not only do we want a solution to (6.14), we want a positive solution. Otherwise $\bar g^\theta=\chi^4 g^\theta$ would have a singularity.

To prove the existence of a suitable solution when θ is large enough we need to prove a few theorems.

Lemma 6.3[18]: Let g be a Riemannian metric, $g-e\in H_{s',\delta'}$, $s'\geqslant 4$, $\delta'>-\frac32$, and f some function, $f\in H_{s'-2,\delta'+2}$. The operator Δ_g-f is injective from $H_{s,\delta}\to H_{s-2,\delta+2}$, $s\geqslant 3$, $\delta>-1$, if
$$\left[\int_{\mathbf{R}^3}|f|^{3/2}\,d\mu\right]^{2/3}<\epsilon_0,$$
for some $\epsilon_0>0$.

Proof: Let us assume the contrary, i.e., there exists a function ϕ that solves
$$\Delta_g\phi-f\phi=0,\quad \phi\in H_{s,\delta}.$$
Multiply across by ϕ to give
$$\phi\Delta_g\phi-f\phi^2=0,$$
that is,
$$\operatorname{div}(\phi\nabla\phi)-(\nabla\phi)^2-f\phi^2=0. \tag{6.15}$$

Take the volume integral of (6.15). The term

$$\int_{\mathbf{R}^3} \operatorname{div}(\phi\nabla\phi)d\mu(g) = \oint_\infty \phi\nabla\phi\cdot d\mathbf{S} = 0,$$

since $\delta > -1$. Hence

$$\int (\nabla\phi)^2\,d\mu(g) + \int f\phi^2\,d\mu(g) = 0.$$

Now

$$\left|\int f\phi^2\,d\mu(g)\right| \leqslant \left[\int |f|^{3/2}\,d\mu(g)\right]^{2/3}\left[\int \phi^6\,d\mu(g)\right]^{1/3}. \tag{6.16}$$

Further we have the Sobolev inequality

$$\left[\int \phi^6\,d\mu(g)\right]^{1/3} \leqslant C_0\int (\nabla\phi)^2\,d\mu(g). \tag{6.17}$$

Therefore we can rewrite (6.16) as

$$\begin{aligned}\int (\nabla\phi)^2\,d\mu(g) &= \left|\int f\phi^2\,d\mu(g)\right|\\ &\leqslant C_0\left[\int |f^{3/2}|d\mu(g)\right]^{2/3}\int(\nabla\phi)^2\,d\mu(g).\end{aligned} \tag{6.18}$$

Therefore if

$$\left[\int |f^{3/2}|d\mu(g)\right]^{2/3} < \frac{1}{C_0},$$

we cannot satisfy (6.18) and hence the lemma must be true.

Aside: This is the old result that with the Schrödinger equation if we have a shallow well we cannot have a bound state. If we regard f as a potential, then the $\frac{3}{2}$ norm is the correct one. This result can be dressed up by splitting f into positive and negative parts f_+ and f_-. We really only need f_- small, i.e.,

$$\left[\int |f_-|^{3/2}\,d\mu(g)\right]^{2/3} < \frac{1}{C_0}$$

to prove the lemma. Again, only the negative part of the potential matters for a bound state.

Now, we have a uniqueness result for an elliptic equation. The obvious thing to do is to use the Fredholm alternative to give us existence. Rather, since we are dealing with noncompact manifolds we can use[10] the following theorem.

Theorem 6.1: Let L be a linear elliptic differential system such that L belongs to a continuous family L_t of such systems $t\in[0,1]$, $L_1 = L$. If each of the L_t is injective and if L_0 is an isomorphism $H_{s,\delta}\to H_{s-2,\delta+2}$ with $-\frac{3}{2}<\delta<-\frac{1}{2}$ then L is also an isomorphism $H_{s,\delta}\to H_{s-2,\delta+2}$. Lemma (6.3) and Theorem (6.1) can be used in combination to show that a unique solution to (6.14) exists. First of all we choose θ large enough so that

$$\left[\int_{\mathbf{R}^3} |R(g^\theta)|^{3/2}\,d\mu(g)\right]^{2/3} < \frac{8}{C_0}$$

[(6.8) shows that this can be done]. Now rewrite (6.14) in the form

$$8\Delta_{g^\theta}(\chi-1) - R(g^\theta)(\chi-1) = R(g^\theta). \tag{6.19}$$

The right-hand side of (6.19) has compact support and hence belongs to $H_{s-2,3/2}$. The operator $\Delta_{g^\theta} - R(g^\theta)$ from Lemma 6.3 is injective. Hence (from Theorem 6.1) a solution $\chi - 1$ exists and belongs to $H_{s,\delta}$ for every $\delta < -\frac{1}{2}$.

We still have to show that the solution χ to (6.14) remains positive. This is shown by the following lemma.

Lemma 6.4: Let χ be the solution to

$$8\Delta_{g^\theta}\chi - R(g^\theta)\chi = 0,\quad \chi - 1\in H_{s,\delta} \tag{6.14}$$

and if

$$\left[\int_{\mathbf{R}^3} |R(g^\theta)|^{3/2}\,d\mu(g^\theta)\right]^{2/3} < \frac{8}{C_0},$$

then $\chi > 0$.

Proof: Again, a proof by contradiction. Say there exists a point in $\mathbf{R}^3$ at which $\chi<0$. Hence there exists a subset S of $\mathbf{R}^3$ in which $\chi\leqslant 0$. This set S must be bounded (because $\chi\to 1$ at ∞), and on the boundary of S, ∂S, $\chi = 0$. Now, repeat the calculation of Lemma 6.3, but now only over the set S:

$$\begin{aligned}&8\Delta_{g^\theta}\chi - R(g^\theta)\chi = 0\\ &\quad\Rightarrow 8\chi\Delta_{g^\theta}\chi - R(g^\theta)\chi^2 = 0\\ &\quad\Rightarrow 8\operatorname{div}(\chi\nabla\chi) - 8(\nabla\chi)^2 - R(g^\theta)\chi^2 = 0.\end{aligned}$$

Integrate over the set S. Now

$$8\int_S \operatorname{div}(\chi\nabla\chi)d\mu(g^\theta) = 8\oint_{\partial S}\chi\nabla\chi\cdot d\mathbf{S} = 0,$$

since $\chi = 0$ on ∂S. Hence we must have

$$8\int_S(\nabla\chi)^2\,d\mu(g^\theta) + \int_S R\chi^2\,d\mu(g^\theta) = 0.$$

Analogous to Eq. (6.18) we have

$$\begin{aligned}8\int_S(\nabla\chi)^2\,d\mu(g^\theta) &\leqslant C_0\left[\int |R(g^\theta)|^{3/2}\,d\mu(g^\theta)\right]^{2/3}\\ &\quad\times\int_S(\nabla\chi)^2\,d\mu(g^\theta).\end{aligned}$$

This cannot be true under the assumptions of the lemma, hence S cannot exist, and so χ can never be negative.

The only other case we have to consider is that $\chi = 0$ at an isolated point. In this case χ, $\nabla\chi$, and $\Delta\chi$ are all zero at that point. This guarantees $\chi\equiv 0$, which does not satisfy the boundary condition $\chi\to 1$ at ∞. Hence $\chi > 0$ everywhere.

We have shown that a solution χ to (6.14) exists and is everywhere positive. We hve also shown $\chi - 1\in H_{s,\delta'}$ for every $\delta' < -\frac{1}{2}$. This (from the imbedding theorem) shows that $r^{1-\epsilon}(\chi-1)\to 0$ at infinity for every $\epsilon > 0$. We can do better than that, however, since the right-hand side of (6.19) has compact support. We can show that $\chi - 1$ falls off at infinity like $\beta/2r$, where β is a constant. This is nothing more than the calculation used to demonstrate Eq. (5.23) in Sec. V. Hence we can show

$$\chi - 1 + \beta/2r\in H_{s,-1/2+\epsilon}(\mathbf{R}^3),\quad \text{for some } \epsilon > 0. \tag{6.20}$$

(From the imbedding theorem we have that $\chi - 1 + \beta/2r$ falls off faster than $1/r$.) We also have that

$$\beta = -\frac{1}{16\pi}\int_{\mathbf{R}^3} R(g^\theta)\chi\, d\mu(g^\theta). \tag{6.21}$$

Therefore, we finally get that the metric $\bar{g}^\theta = \chi^4 g^\theta$ is asymptotically Schwarschildean and satisfies $R(\bar{g}^\theta) \equiv 0$.

The next problem is to show that the difference between the ADM mass associated with the original metric g and the ADM mass associated with $\bar{g}^\theta$ is small. Analogous to Eq. (6.3), it is easy to show

$$P^0(g) - P^0(\bar{g}^\theta) = \beta. \tag{6.22}$$

At a naive level it seems obvious that as θ gets large $R(g^\theta)$ gets small and $\chi \approx 1$. When this is pushed into (6.21) we expect that β would get small. To turn this into a precise argument is a nontrivial operation and we will follow Schoen and Yau,[3] who have worked it out in detail.

We have to show that

$$\beta = -\frac{1}{16\pi}\int_{\mathbf{R}^3} R(g^\theta)\chi\, d\mu(g^\theta) \tag{6.21}$$

gets small as θ gets large. We already know that $\int R(g^\theta) d\mu(g^\theta)$ gets small [(6.10)] and so it is sufficient to show that

$$\alpha = \int_{\mathbf{R}^3} R(g^\theta)(\chi - 1) d\mu(g^\theta)$$

gets small. We have that $v = \chi - 1$ satisfies (6.19),

$$8\Delta_{g^\theta} v - R(g^\theta)v = R(g^\theta). \tag{6.19}$$

Multiplying across by v and integrating gives

$$8\oint_\infty v\nabla v \cdot d\mathbf{S} - 8\int (\nabla v)^2\, d\mu(g^\theta) - \int R(g^\theta)v^2\, d\mu(g^\theta) = \int R(g^\theta)v d\mu(g^\theta).$$

The surface integral is zero, since v falls off like $1/r$ and so we get

$$8\int (\nabla v)^2\, d\mu(g^\theta) - \int R(g^\theta)v^2\, d\mu(g^\theta) - \int R(g^\theta)v\, d\mu(g^\theta). \tag{6.23}$$

Hence

$$8\int (\nabla v)^2\, d\mu(g^\theta) \leqslant \left[\int |R(g^\theta)|^{3/2}\, d\mu\right]^{2/3}\left[\int v^6\, d\mu(g^\theta)\right]^{1/3} + \left[\int |R(g^\theta)|^{6/5}\, d\mu\right]^{5/6}\left[\int v^6\, d\mu(g^\theta)\right]^{1/6}.$$

We now use the Sobolev inequality (6.17)

$$\left[\int v^6\, d\mu(g^\theta)\right]^{1/3} \leqslant C_0\int (\nabla v)^2\, d\mu(g^\theta) \tag{6.17}$$

to give

$$8\int (\nabla v)^2\, d\mu(g^\theta) \leqslant C_0\left[\int |R(g^\theta)|^{3/2}\, d\mu(g^\theta)\right]^{2/3}\int (\nabla v)^2\, d\mu(g^\theta) + \sqrt{C_0}\left[\int |R(g^\theta)|^{6/5}\, d\mu(g^\theta)\right]^{5/6} \times\left[\int (\nabla v)^2\, d\mu(g^\theta)\right]^{1/2}. \tag{6.24}$$

If we choose θ large enough we can have [see (6.8)]

$$\left[\int |R(g^\theta)|^{3/2}\, d\mu(g^\theta)\right]^{2/3} < \frac{8}{C_0}.$$

In that case we can rearrange (6.24) to have

$$\int (\nabla v)^2\, d\mu(g^\theta) \leqslant \frac{C_0[\int |R(g^\theta)|^{6/5}\, d\mu(g^\theta)]^{5/3}}{\{8 - C_0[\int |R(g^\theta|^{3/2}\, d\mu(g^\theta)]^{2/3}\}^2}.$$

From (6.8) and (6.9) it is clear that

$$\int (\nabla v)^2\, d\mu(g^\theta) \to 0 \quad \text{as} \quad \theta \to \infty. \tag{6.25}$$

Let us return to (6.23). It also can be written

$$\alpha = \int R(g^\theta)v\, d\mu(g^\theta) = -8\int (\nabla v)^2 d\mu(g^\theta) - \int R(g^\theta)v^2\, d\mu(g^\theta)$$

Hence

$$\alpha \leqslant -8\int (\nabla v)^2\, d\mu(g^\theta) + \left[\int |R(g^\theta)|^{3/2}\, d\mu(g^\theta)\right]^{2/3}\left[\int v^6\, d\mu(g^\theta)\right]^{1/3}$$
$$\leqslant -8\int (\nabla v)^2\, d\mu(g^\theta) + C_0\left[\int |R(g^\theta)^{3/2}\, d\mu(g^\theta)\right]^{2/3}\left[\int (\nabla v)^2\, d\mu(g^\theta)\right].$$

Therefore as $\theta \to \infty$, $\alpha \to 0$, hence $\beta \to 0$.

We have finally shown what we set out to do in part (c) of the Schoen and Yau program. This means that given any metric g, $g - e \in H_{s,\delta}(\mathbb{R}^3)$, $s \geqslant 4$, $\delta > -1$, with $R(g) = 0$ and any $\epsilon > 0$, there exists a metric g', which is Schwarzschildean at infinity so that $|P^0(g) - P^0(g')| < \epsilon$ and $R(g') = 0$.

The last stage (D) of the Schoen and Yau program consists of showing that all metrics g' that satisfy $R(g') = 0$ and are Schwarzschildean outside a region of compact support have positive mass. We will not discuss this result here. All we wish to do is point out that this result, when combined with part (C), shows that all metrics g, $g - e \in H_{s,\delta}$, $s \geqslant 4$, $\delta > -1$, with $R(g) = 0$ have positive mass. In turn part (B) now shows that all metrics g, $g - e \in H_{s,\delta}$, $s \geqslant 4$, $\delta > -1$, with $R(g) \geqslant 0$, $R(g) \in H_{s',\delta'}$, $s' \geqslant 3$, $\delta' > \frac{3}{2}$ have positive mass. Finally, using part (A), this suffices to show that all solutions to the constraints satisfying (6.1) have positive mass.

The boost theorem shows that initial data of the form (6.1) are preserved under Lorentz transformations. This means that the ADM energy is positive in every Lorentz frame. Hence the ADM energy momentum four-vector is timelike and future pointing.

VII. CONCLUSIONS

In this article we have shown that if we have initial data to the vacuum Einstein equations satisfying

$$g-e \in H_{s,\delta}(\mathbb{R}^3), \quad k \in H_{s-1,\delta+1}(\mathbb{R}^3), \quad s \geqslant 4, \quad \delta > -1, \tag{7.1}$$

then the space-time generated by this data has a well-defined, finite, conserved, Lorentz covariant, future-pointing, timelike energy momentum vector.

Can we relax conditions (7.1)? The falloff condition $\delta > -1$ in (7.1) cannot be significantly weakened. One demonstration of this is to consider the following coordinate transformation of flat space-time:

$$t' = t + \sqrt{8\alpha r}, \quad r' = r \quad (\alpha \quad \text{const}). \tag{7.2}$$

This gives $g'_{ij} = \delta_{ij} - 2\alpha x_i x_j / r^3$, where g'_{ij} is the induced three-metric of the $t' = 0$ slice. When this metric is substituted into the ADM expression (4.9) and (5.12), we immediately get $P^0 = -\alpha$. Of course, we know that the flat space energy must be zero. This apparent contradiction is resolved when we realize that in the coordinate system (7.2), k_{ij} falls off like $r^{-3/2}$ and hence belongs to every $H_{s,\delta}$ with $\delta < -1$, but does not belong to any $H_{s,\delta}$ with $\delta \geqslant -1$, and so does not satisfy (7.1). In this particular example the ADM momentum is actually zero due to the spherical symmetry of the transformation (7.2). This example shows clearly that the ADM energy momentum being finite is not equivalent to it being well defined.

An alternative demonstration of this same disease can be obtained by considering the Schwarzschild solution. The lapse function joining the regular $t = 0$ slice of the Schwarzschild solution to a constant time slice in Lemaitre coordinates is of the form $N \sim -\sqrt{8Mr}$. Now the constant time slices of the Lemaitre coordinates are flat so obviously $P^0 = 0$ and k falls off like $r^{-3/2}$. Again the momentum is zero from the spherical symmetry.

Thus, the only case we need to worry about is $\delta = -1$. This condition is sufficient to give a finite P^0 and P^i [from Eq. (5.15)]. However, we use the $\delta > -1$ condition in two places in this article. First, we use the fact that $t_{\mu 0}$ [as defined by Eq. (4.2)] falls off faster than $1/r^3$ to show that the energy momentum is conserved [Eq. (4.7)]. Second, we use $\delta > -1$ in the positive energy proofs, particularly in deriving the estimates (6.9) and (6.10), which are very important in the proof.

There exists, of course, an alternative method of proving the positivity of mass, the Witten proof.[5] This method also can be used to relax the asymptotic conditions and show that the energy momentum is well behaved under weak asymptotic conditions. The Witten proof is totally different from the Schoen and Yau proof and is, in many ways, much easier. An analysis of the Witten proof, with emphasis on weakening the asymptotic conditions, already has been carried out by Reula.[19]

The Reula analysis cannot be translated directly into the language of weighted Sobolev spaces because he assumes that the metric and extrinsic curvature are C^∞. With this condition all he needs is that $(g_{ab,c})^2$ and $(k_{ab})^2$ be both integrable. Thus, this result is simultaneously stronger and weaker than the result we obtain here. It would be desirable to redo the Reula analysis in the language of weighted Sobolev spaces because it seems to indicate very strongly that with $\delta = -1$ we can show $P^0 > 0$ and $(P^0)^2 > (P^i)^2$ on a single slice. This still leaves open the question of the energy momentum being conserved. Until this can be resolved, it may be better to stick with $\delta > -1$.

It is important to realize that in this paper we are only setting conditions on the independent gravitational degrees of freedom. The dependent parts of the gravitational field are determined by the constraints, which are elliptic equations.[16] This is why the natural asymptotic conditions are in terms of "falling off faster than" rather than "falling off like." It is the gravitational potentials[20] that emerge when we solve the constraints that have the specified falloff.

The relationship between the $r^{-1/2}$ falloff and finite mass is not new. It is implicit in the key perturbation calculation of Brill and Deser.[14] Sommers[21] discusses a source that has been radiating for an infinite time to the past. He assumes that the power radiated falls off like $(-t)^n$. If $n < -1$, the total power radiated will be finite. This solution will have the curvature falling off at spacelike infinity like $r^{n/2-2}$, i.e., for finite total power faster than $r^{-5/2}$. Hence the metric itself need only fall off faster than $r^{-1/2}$.

Schutz and Sorkin[22] have explicitly demonstrated that the mass is finite if the metric falls off faster than $r^{-1/2}$. In their calculation the mass is defined via a variational principle which replaces the initial value constraints. This approach could have been used in Sec. IV in place of the more direct method we did adopt.

In retrospect, it is amazing how accurately the perturbation analysis of Brill and Deser[14] represented the true state of affairs. They obtained

$$\delta^2 P^0 = \frac{1}{16\pi} \int \left[\frac{1}{4} (\delta g^{\mathrm{TT}}_{ij,k})^2 + (\delta \pi^{ij}_{\mathrm{TT}})^2 \right] d^3x.$$

Therefore, the energy in their analysis would be finite and positive if the TT parts of g_{ij} and k^{ij} are square integrable. Further, the perturbation analysis of Deser *et al.*[23] shows

$$\delta^2 P^i = -\frac{1}{16\pi} \int (\delta g^{\mathrm{TT}}_{jk,i} \delta \pi^{jk}_{\mathrm{TT}}) d^3x.$$

This immediately shows $(\delta^2 P^0)^2 \geqslant (\delta^2 P^i)^2$ and so we have the positivity and the future-pointing nature of the energy momentum. These expressions are correct for nonmaximal as well as maximal data, see Ó Murchadha.[24]

Finally, I would like to point out that in most of the proofs I have made no particular attempt to state minimal conditions on the degree of differentiability needed, especially in the existence of harmonic coordinates and the positivity of energy proofs. In both these cases I have been guided by the $s \geqslant 4$ condition in the boost theorem. Undoubtedly sharper results could be obtained if we were interested in these proofs for their own sakes.

ACKNOWLEDGMENT

I wish to acknowledge with gratitude the major assistance I got from Demetrios Christodoulou. Without his help the paper would never have reached its present state. Of course, I am solely responsible for any errors therein.

[1]C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[2]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[3]R. Schoen and S.-T. Yau, Commun. Math. Phys. **65**, 45 (1979).
[4]R. Schoen and S.-T. Yau, Commun. Math. Phys. **79**, 231 (1981).
[5]E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[6]D. Christodoulou and N. Ó Murchadha, Commun. Math. Phys. **80**, 271 (1981).
[7]R. Annowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[8]S. Deser, Phys. Lett. B **64**, 463 (1976).
[9]R. Adams, *Sobolev Spaces* (Academic, New York, 1975), p. 97.
[10]Y. Choquet-Bruhat and D. Christodoulou, Acta Math. **146**, 129 (1981).
[11]Reference 9, p. 115.
[12]D. Christodoulou, J. Math. Pur. Appl. **60**, 99 (1981).
[13]Reference 9, p. 97.
[14]D. Brill and S. Deser, Ann. Phys. (NY) **50**, 548 (1968).
[15]R. McOwen, Commun. Pure Appl. Math. **32**, 783 (1979).
[16]N. Ó Murchadha and J. W. York, Phys. Rev. D **10**, 428 (1974).
[17]N. Ó Murchadha and J. W. York, Phys. Rev. D **10**, 2345 (1974).
[18]M. Cantor and D. Brill, Compositio. Math. **43**, 317 (1981).
[19]O. Reula, J. Math. Phys. **23**, 810 (1982).
[20]N. Ó Murchadha and J. W. York, Gen. Relativ. Gravit. **7**, 257 (1976).
[21]P. Sommers, J. Math. Phys. **19**, 549 (1978).
[22]B. F. Schutz and R. Sorkin, Ann. Phys. (NY) **107**, 1 (1977).
[23]S. Deser, J. Trubatch, and S. Trubatch, Nuovo Cimento **39**, 1159 (1965).
[24]N. Ó Murchadha, Nuovo Cimento B **34**, 141 (1976).

The initial value problem for colliding gravitational and hydrodynamic waves

Basilis C. Xanthopoulos
Department of Physics, University of Crete and Research Center of Crete, Iraklion, Greece

(Received 17 January 1986; accepted for publication 26 March 1986)

The collision of two plane gravitational and hydrodynamic waves with parallel polarizations is studied. In the interaction region, to the future of the collision, the space-time admits two hypersurface orthogonal Killing fields, and the problem reduces to solving two (decoupled!) linear partial differential equations. The characteristic initial value problem for these equations is explicitly solved by means of Riemann's method. In the appendices the relevant field equations are written in two different coordinate systems that have been proved useful in the studies of colliding waves, relationships among the solutions of the gravitational and the hydrodynamic equations are obtained, separable solutions involving Bessel functions are constructed, and integral identities among the Legendre functions are obtained.

I. INTRODUCTION

Stationary axisymmetric space-times, as well as space-times with two spacelike commuting Killing fields, have been considered extensively in general relativity: The former describe the space-time of uniformly rotating stars and the latter—except from cylindrically symmetrical situations that will not be considered in the present paper—describe the interaction region of two colliding plane gravitational waves, i.e., the part of space-time to the future of the collision. The common geometrical characteristic of the two cases is the existence of two commuting Killing vector fields, so one expects to encounter the same essential difficulty in solving the relevant Einstein equations. But the asymptotic or boundary conditions are quite different for the description of the relevant problems.

The equations are much simplified when the two Killing fields are hypersurface orthogonal. The former case then corresponds to the absence of rotation and the latter to the collision of plane gravitational waves with collinear polarizations. In vacuum regions, where there exists only the gravitational field, the former case is described by the well-known Weyl solutions and the latter was studied by Szekeres.[1] So, rotation in stationary axisymmetric space-times corresponds to different polarizations for space-times with two spacelike Killing fields. And this correspondence seems to go even deeper between the mostly interesting solutions in the two cases. It was recently found[2,3] that the black hole solutions in the former case correspond, in a well-defined mathematical manner, to the solutions describing the interaction region of colliding impulsive gravitational waves.[4,5]

There exists, however, an essential difference between stationary axisymmetry space-times and space-times with two spacelike Killing fields: while in the former case the space-time metric is time independent, in the latter case the metric is time dependent and as such, the basic problem is that of the time evolution of the prescribed initial data. The main objective of the present paper is to show how to solve this Cauchy problem for space-times with two hypersurface orthogonal spacelike commuting Killing vectors when the gravitational field is coupled with a perfect fluid satisfying the extremely relativistic equation of state energy density $\epsilon =$ pressure p. We shall show that this problem reduces to two quite similar linear partial differential equations for which the underlying Cauchy problem can be explicitly solved. Szekeres[1] has solved the same problem for the vacuum equations.

II. THE EQUATIONS

We have studied,[6] *ab initio*, space-times with two spacelike commuting Killing fields coupled to perfect fluids with $\epsilon = p$ the equation of state. We summarize here the results of the reduction of the coupled Einstein and hydrodynamic equations.

The metric can be written in the form

$$ds^2 = e^{2\nu + f}(dx^0)^2 - e^{2\mu_3 + f}(dx^3)^2 - e^{2\psi}(dx^1 - q_2\, dx^2)^2 - e^{2\mu_2}(dx^2)^2, \tag{2.1}$$

where $\partial/\partial x^1$ and $\partial/\partial x^2$ are the two Killing fields. We can impose the same gauge conditions as in the vacuum case,[2] namely the conditions

$$e^{\mu_3 - \nu} = \sqrt{\Delta}, \quad e^{\beta} = e^{\psi + \mu_2} = \sqrt{\Delta\delta}, \tag{2.2}$$

where

$$\Delta = 1 - \eta^2, \quad \delta = 1 - \mu^2. \tag{2.3}$$

With these gauge conditions and the notation

$$\chi = e^{\mu_2 - \psi}, \tag{2.4}$$

the metric (1) becomes

$$ds^2 = e^{\nu + \mu_3 + f}\sqrt{\Delta}[\Delta^{-1}(d\eta)^2 - \delta^{-1}(d\mu)^2] - \sqrt{\Delta\delta}[\chi(dx^2)^2 + \chi^{-1}(dx^1 - q_2\, dx^2)^2]. \tag{2.5}$$

The complex combination

$$\chi + iq_2 = Z = (1+E)/(1-E) \tag{2.6}$$

satisfies the Ernst equation

$$(Z + Z^*)[(\Delta Z_{,\eta})_{,\eta} - (\delta Z_{,\mu})_{,\mu}] = 2(\Delta Z_{,\eta}^2 - \delta Z_{,\mu}^2) \tag{2.7}$$

or

$$(1 - EE^*)[(\Delta E_{,\eta})_{,\eta} - (\delta E_{,\mu})_{,\mu}] = -2E^*(\Delta E_{,\eta}^2 - \delta E_{,\mu}^2), \tag{2.8}$$

 0022-2488/86/082129-08$02.50

while $\nu + \mu_3$ is determined (up to an irrelevant additive constant) by

$$(\mu/\delta)(\nu+\mu_3)_{,\eta} + (\eta/\Delta)(\nu+\mu_3)_{,\mu}$$
$$= -(1/\chi^2)(\chi_{,\eta}\chi_{,\mu} + q_{2,\eta}q_{2,\mu})$$
$$= -\frac{2(E_{,\eta}E^*_{,\mu} + E^*_{,\eta}E_{,\mu})}{(1-EE^*)^2}, \qquad (2.9a)$$

$$2\eta(\nu+\mu_3)_{,\eta} + 2\mu(\nu+\mu_3)_{,\mu}$$
$$= \frac{3}{\Delta} + \frac{1}{\delta} - \frac{1}{\chi^2}[\Delta(\chi^2_{,\eta} + q^2_{2,\eta}) + \delta(\chi^2_{,\mu} + q^2_{2,\mu})]$$
$$= \frac{3}{\Delta} + \frac{1}{\delta} - \frac{4}{(1-EE^*)^2}(\Delta E_{,\eta}E^*_{,\eta} + \delta E_{,\mu}E^*_{,\mu}). \qquad (2.9b)$$

The fluid is described by a stream potential $\phi = \phi(\eta,\mu)$ subject to the equation

$$\Delta\phi_{,\eta\eta} - \delta\phi_{,\mu\mu} = 0. \qquad (2.10)$$

Then the scalar f in the metric (2.1), which owes its existence entirely to the presence of the fluid, is determined from ϕ (again up to an additive constant) by

$$\frac{\mu}{\delta}f_{,\eta} + \frac{\eta}{\Delta}f_{,\mu} = \frac{8}{\Delta\delta}\phi_{,\eta}\phi_{,\mu}, \qquad (2.11a)$$

$$\eta f_{,\eta} + \mu f_{,\mu} = \frac{4}{\Delta\delta}(\Delta\phi^2_{,\eta} + \delta\phi^2_{,\mu}). \qquad (2.11b)$$

The energy density of the fluid is

$$\epsilon = -(\Delta\delta)^{-1}e^{-2\mu_3 - f}(\Delta\phi^2_{,\eta} - \delta\phi^2_{,\mu}) \qquad (2.12)$$

and the tetrad components $u_{(a)}$ of the fluids for velocity, related to the tensor components u_a by

$$u^{(0)} = u_{(0)} = e^{\nu}u^0 = e^{-\nu}u_0,$$
$$u^{(3)} = -u_{(3)} = e^{\mu_3}u^3 = -e^{-\mu_3}u_3, \qquad (2.13)$$

are given by

$$u_{(0)}\sqrt{\epsilon} = \Delta^{-1/2}e^{-\mu_3}\phi_{,\mu}, \quad u_{(3)}\sqrt{\epsilon} = -\delta^{-1/2}e^{-\mu_3}\phi_{,\eta}. \qquad (2.14)$$

The difficult part is to solve Eq. (2.7) or Eq. (2.8) for the gravitational field and Eq. (2.10) for the fluid; then $\nu + \mu_3$ and f are determined by straightforward quadratures. When $\phi = f = 0$, the problem reduces to the vacuum Einstein equations with two spacelike commuting Killing fields.

III. THE SOLUTION OF THE INITIAL VALUE PROBLEM

We shall be concerned, from now on, with space-times with hypersurface orthogonal Killing fields, i.e., with space-times for which $q_2 = 0$. In this case the Ernst potential Z is real and Eq. (2.7) reads

$$\chi[(\Delta\chi_{,\eta})_{,\eta} - (\delta\chi_{,\mu})_{,\mu}] = \Delta\chi^2_{,\eta} - \delta\chi^2_{,\mu}, \qquad (3.1)$$

which, with the introduction of

$$\mathbf{X} = \ln\chi, \qquad (3.2)$$

reduces to

$$(\Delta\mathbf{X}_{,\eta})_{,\eta} - (\delta\mathbf{X}_{,\mu})_{,\mu} = 0. \qquad (3.3)$$

Therefore, the collision of two plane gravitational and hydrodynamic waves with collinear polarizations is described by the two linear equations (3.3) for the gravitational field and (2.10) for the fluid. The two equations are quite similar, having the same second-order derivative terms, but not identical. And their similarity becomes more apparent when we write them in the null coordinates (u,v) given by

$$\eta = u\sqrt{1-v^2} + v\sqrt{1-u^2}, \quad \mu = u\sqrt{1-v^2} - v\sqrt{1-u^2}. \qquad (3.4)$$

They become

$$\mathbf{X}_{,uv} - [1/(1-u^2-v^2)](v\mathbf{X}_{,u} + u\mathbf{X}_{,v}) = 0, \qquad (3.5)$$

and

$$\phi_{,uv} + [1/(1-u^2-v^2)](v\phi_{,u} + u\phi_{,v}) = 0, \qquad (3.6)$$

respectively. In fact, we shall be able to treat both equations (3.5) and (3.6) simultaneously. To do that, we shall introduce the two new null independent variables

$$x = 1 - 2u^2, \quad y = 1 - 2v^2. \qquad (3.7)$$

Then the two equations reduce to

$$\mathscr{L}(U) = U_{,xy} + [n/(x+y)](U_{,x} + U_{,y}) = 0, \quad (3.8)$$

where ($U = X$, $n = \frac{1}{2}$) is Eq. (3.5) for the gravitational and ($U = \phi$, $n = -\frac{1}{2}$) is Eq. (3.6) for the fluid field. Equation (3.8) can be further reduced to the self-adjoint equation

$$\widetilde{U}_{,xy} + n(1-n)(x+y)^{-2}\widetilde{U}(x,y) = 0, \qquad (3.9)$$

by the substitution

$$U = (x+y)^{-n}\widetilde{U}, \qquad (3.10)$$

but we shall not be using this self-adjoint form of the equation because, since $n(1-n)$ equals $\frac{1}{4}$ for $n = \frac{1}{2}$ and $-\frac{3}{4}$ for $n = -\frac{1}{2}$, we shall have to consider two different equations.

The initial value problem that we wish to solve is to determine $U(x,y)$ in the domain of dependence of the initial data (Fig. 1). The initial data are U and its first derivatives, specified on the curve PQ, which does not intersect any line parallel to the characteristics of Eq. (3.8) more than once. It is well known that the solution of the initial value problem that we have formulated can be obtained by Riemann's method, for a very readable account of which we recommend the book of Copson.[7]

A successful application of Riemann's method depends on finding explicitly a solution $V(x,y;x_0,y_0)$—the Riemann function!—of the adjoint of the equation that one studies, which for the particular equation (3.8) we are concerned with is

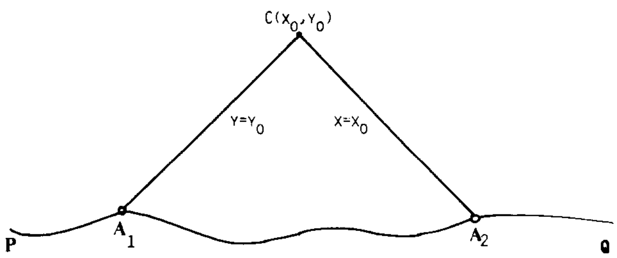


FIG. 1. The lines CA_1 and CA_2 are the characteristics of the equation through the (arbitrary) point C at which we would like to determine the solution. The initial data are given on the curve PQ, which does not intersect any line parallel to the characteristics more than once. Due to the hyperbolic type of the equation, the solution at C depends only on the initial data on A_1A_2.

$$\mathscr{L}^*(V) = V_{,xy} - \frac{n}{(x+y)}(V_{,x} + V_{,y}) + \frac{2nV}{(x+y)^2} = 0. \tag{3.11}$$

Besides solving Eq. (3.11), the Riemann function also satisfies the conditions

$$V_{,x} = [n/(x+y)]V, \quad \text{when } y = y_0,$$
$$V_{,y} = [n/(x+y)]V, \quad \text{when } x = x_0, \tag{3.12}$$

on the characteristics $x = x_0$ and $y = y_0$ of Eq. (3.8) through the point (x_0, y_0) that we are interested in to evaluate the solution. For normalization, Riemann's function also satisfies the condition

$$V(x_0,y_0;x_0,y_0) = 1. \tag{3.13}$$

Fortunately, Riemann's function for Eq. (3.8) is known (Copson,[7] §5.7). It is

$$V(x,y;x_0,y_0) = \frac{(x+y)^n}{(x_0+y_0)^n} P_{-n}\left(1 + \frac{2(x-x_0)(y-y_0)}{(x+y)(x_0+y_0)}\right), \tag{3.14}$$

where P_{-n} is Legendre's ring function with (not necessarily an integer) index $-n$. In terms of $V(x,y;x_0,y_0)$ the value of U at the arbitrary point (x_0,y_0) is determined by the values it takes along the curve A_1A_2, where A_1 and A_2 are the intersections of the initial data curve PQ with the characteristics through (x_0,y_0).

Following Copson, we write

$$V\mathscr{L}(U) - U\mathscr{L}^*(V) = \frac{\partial H}{\partial x} + \frac{\partial K}{\partial y}, \tag{3.15}$$

where

$$H = \frac{nUV}{x+y} + \frac{1}{2}(VU_{,y} - UV_{,y}),$$
$$K = \frac{nUV}{x+y} + \frac{1}{2}(VU_{,x} - UV_{,x}). \tag{3.16}$$

Then integrating in the "triangle" CA_1A_2 and using that $\mathscr{L}(U) = \mathscr{L}^*(V) = 0$, we get

$$\begin{aligned} 0 &= \int\int [V\mathscr{L}(U) - U\mathscr{L}^*(V)]dx\,dy = \int\int \left(\frac{\partial H}{\partial x} + \frac{\partial K}{\partial y}\right)dx\,dy = \oint (H\,dy - K\,dx) \\ &= \int_{A_1}^{A_2} (H\,dy - K\,dx) - \int_C^{A_2} H(x_0)dy - \int_C^{A_1} K(y_0)dx = (UV)_{(C)} - \frac{1}{2}[(UV)_{(A_1)} + (UV)_{(A_2)}] \\ &\quad + \int_{A_1}^{A_2} \left\{\left(\frac{nUV}{x+y} + \frac{1}{2}VU_{,y} - \frac{1}{2}UV_{,y}\right)dy - \left(\frac{nUV}{x+y} + \frac{1}{2}VU_{,x} - \frac{1}{2}UV_{,x}\right)dx\right\} = 0, \end{aligned} \tag{3.17}$$

from which we can express $U(x_0,y_0)$ as an integral over the initial data

$$\begin{aligned} U(x_0,y_0) &= \tfrac{1}{2}[U(A_1)V(A_1;x_0,y_0) + U(A_2)V(A_2;x_0,y_0)] \\ &\quad + \int_{A_1}^{A_2} \left\{\left(\frac{nUV}{x+y} + \frac{1}{2}VU_{,x} - \frac{1}{2}UV_{,x}\right)dx - \left(\frac{nUV}{x+y} + \frac{1}{2}VU_{,y} - \frac{1}{2}UV_{,y}\right)dy\right\}. \end{aligned} \tag{3.18}$$

IV. THE CHARACTERISTIC PROBLEM

We are mainly interested in applying the previous initial value problem to the description of the collision of plane gravitational and hydrodynamic waves. In this case the space-time consists of four distinct regions (Fig. 2). Region IV is the part of the space-time before the arrival of neither of the waves and regions II and III are the parts of the space-time after the arrival of only one or the other of the waves, depending on the position of the observer. Finally, region I, which is to the future of the collision (occurring at 0), is the part of the space-time where the interaction of the waves occurs. In this problem, therefore, the data are given on the null boundaries $u = 0$ and $v = 0$ separating regions I and II and I and III, which coincide with the characteristics of Eqs. (3.5) and (3.6). The problem is to determine the evolution of the characteristic data, i.e., to determine X and ϕ at the arbitrary point C of the interaction region, from the values of X and ϕ on the null boundaries $u = 0$, $0 \leqslant v \leqslant 1$ and $v = 0$, $0 \leqslant u \leqslant 1$, which are specified from the particular incoming gravitational and sound waves. Since the lines $u = u_0$ and $v = v_0$ through C, the characteristics of Eqs. (3.5) and (3.6), intersect each of the null boundaries only once, there is no problem in applying the previously developed theory in the present setup.

In the coordinates x,y [see Eq. (3.7)] the situation of Fig. 2 is mapped into that of Fig. 3. The characteristic data are given on OA and OB and we would like to determine their evolution in the arbitrary point (x_0,y_0) inside the triangle ABO on which region I of Fig. 2 is mapped by the transformation (3.7). After some elementary manipulations on the expression (3.18) we find that the solution of Eq. (3.8) is

$$U(x_0,y_0) = U(1,1)V(1,1;x_0,y_0) - \int_{x_0}^1 \left[\left(U_{,x} + \frac{nU}{x+y}\right)V\right]_{(y=1)} dx - \int_{y_0}^1 \left[\left(U_{,y} + \frac{nU}{x+y}\right)V\right]_{(x=1)} dy. \tag{4.1}$$

Moreover, by substituting the Riemann function (3.14) and changing the arbitrary point of evaluation of the solution from (x_0,y_0) to (x,y) we get

$$\begin{aligned} U(x,y) &= \frac{2^nU(1,1)}{(x+y)^n} P_{-n}\left(\frac{1+xy}{x+y}\right) - \frac{1}{(x+y)^n}\int_x^1 \left[\left(U_{,x} + \frac{nU}{1+\xi}\right)_{(\xi,1)}\right](1+\xi)^n P_{-n}\left(1 + \frac{2(\xi-x)(1-y)}{(1+\xi)(x+y)}\right)d\xi \\ &\quad - \frac{1}{(x+y)^n}\int_y^1 \left[\left(U_{,y} + \frac{nU}{1+\xi}\right)_{(1,\xi)}\right](1+\xi)^n P_{-n}\left(1 + \frac{2(\xi-y)(1-x)}{(1+\xi)(x+y)}\right)d\xi. \end{aligned} \tag{4.2}$$

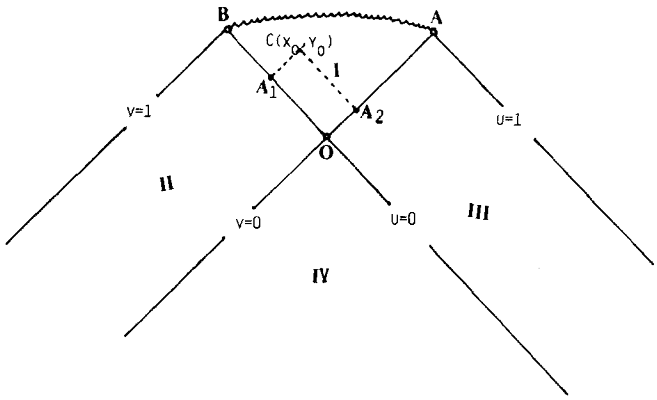


FIG. 2. The space-time diagram for two colliding plane gravitational waves. The gravitational waves are propagated along the null directions u and v. The plane of the wave fronts (on which the geometry is invariant) is orthogonal to the plane of the diagram. The instance of the collision is the point O ($u=0$, $v=0$). Region IV, which is flat, is the portion of the space-time prior to the arrival of neither wave. Regions II and III are the portions of space-time after the passage of one or the other of the waves. They are flat for impulsive waves and vacuum solutions for arbitrary plane waves. Region I, which is to the future of the collision, is the interaction region. The problem is to determine the solution in region I from initial data given on the null lines OA and OB. The curve AB is where the focusing occurs and the singularity develops.

The expression (45) provides the general solution for ($U=X, n=\frac{1}{2}$) for the metric function and for ($U=\phi$, $n=-\frac{1}{2}$) for the stream potential of the fluid.

ACKNOWLEDGMENTS

I would like to thank Professor S. Chandrasekhar of the University of Chicago, who introduced me to the study of the collision of gravitational waves, and with whom initial parts of the present work were done in collaboration. I also would like to thank Professor S. Pichorides of the University of Crete for many useful discussions concerning Riemann's method.

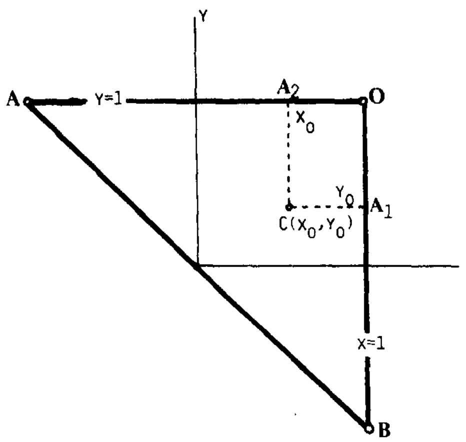


FIG. 3. The interaction region I of Fig. 2 is mapped to the triangle OAB by the transformation (3.7). OB and OA are the boundaries between regions I and II, and I and III, respectively. We would like to determine the solution at C from initial data given on OA and OB. A singularity is expected to develop at AB.

APPENDIX A: THE INTRODUCTION OF NULL COORDINATES

For the description of the collision of gravitational waves we also need to describe the space-time in null coordinates (u,v), related to η and μ by

$$\eta = u\sqrt{1-v^2} + v\sqrt{1-u^2}, \quad \mu = u\sqrt{1-v^2} - v\sqrt{1-u^2}. \tag{A1}$$

Then the Ernst equation reads

$$(Z+Z^*)\left[Z_{,uv} - \frac{uZ_{,v} + vZ_{,u}}{1+-u^2-v^2}\right] = 2Z_{,u}Z_{,v} \tag{A2}$$

or

$$(1-EE^*)\left[E_{,uv} - \frac{uE_{,v} + vE_{,u}}{1-u^2-v^2}\right] = -2E^*E_{,u}E_{,v}. \tag{A3}$$

To integrate equations (2.9) for $\nu+\mu_3$, is most convenient to introduce Φ by

$$e^{\nu+\mu_3} = [(\eta^2-\mu^2)/(1-\mu^2)^{1/4}(1-\eta^2)^{3/4}]e^{\Phi}. \tag{A4}$$

Then Φ is determined from E via

$$\Phi_{,u} = -\frac{(1-u^2-v^2)}{u}\frac{E_{,u}E^*_{,u}}{(1-EE^*)^2}, \tag{A5a}$$

$$\Phi_{,v} = -\frac{(1-u^2-v^2)}{v}\frac{E_{,v}E^*_{,v}}{(1-EE^*)^2}, \tag{A5b}$$

where, in obtaining Eqs. (A5), we made use of the identities

$$\begin{aligned} &E_{,\eta}E^*_{,\mu} + E^*_{,\eta}E_{,\mu} \\ &\quad = [1/2(1-u^2-v^2)][(1-u^2)E_{,u}E^*_{,u} \\ &\qquad - (1-v^2)E_{,v}E^*_{,v}] \end{aligned} \tag{A6}$$

and

$$\begin{aligned} &(1-\eta^2)E_{,\eta}E^*_{,\eta} + (1-\mu^2)E_{,\mu}E^*_{,\mu} \\ &\quad = \tfrac{1}{2}[(1-u^2)E_{,u}E^*_{,u} + (1-v^2)E_{,v}E^*_{v}]. \end{aligned} \tag{A7}$$

The stream-potential equation reads

$$\phi_{,uv} + (u\phi_{,v} + v\phi_{,u})/(1-u^2-v^2) = 0, \tag{A8}$$

and f is obtained from ϕ by

$$f_{,u} = \frac{2\phi^2_{,u}}{u(1-u^2-v^2)}, \quad f_{,v} = \frac{2\phi^2_{,v}}{v(1-u^2-v^2)}, \tag{A9}$$

while the expression for the energy density reads

$$\epsilon = -\frac{(1-u^2)^{1/2}(1-v^2)^{1/2}}{(1-u^2-v^2)^2}e^{-2\mu_3-f}\phi_{,u}\phi_{,v}. \tag{A10}$$

Note also the relationships

$$\Delta\delta = (1-u^2-v^2)^2 \tag{A11}$$

and

$$\frac{(d\eta)^2}{\Delta} - \frac{(d\mu)^2}{\delta} = \frac{4(du)(dv)}{(1-u^2)^{1/2}(1-v^2)^{1/2}}, \tag{A12}$$

which are needed to express the metric (2.5) in the null coordinates.

An intermediate step in the transition from the (η,μ) to the (u,v) coordinates is the introduction of the angular coordinates (ψ,ϑ):

$$\eta = \cos\psi, \quad \mu = \cos\vartheta. \tag{A13}$$

In fact it was found that to describe the space-time of the collision of plane gravitational waves in a Newman–Penrose formalism and to evaluate the corresponding Weyl and Ricci scalars it is most convenient to work in the angular variables. The null coordinates are then obtained from the angular ones by

$$u = \cos((\psi + \vartheta)/2), \quad v = \sin((\vartheta - \psi)/2). \tag{A14}$$

In the angular coordinates the Ernst equation reads

$$(Z + Z^*)(Z_{,\psi\psi} - Z_{,\vartheta\vartheta} + \cot\psi Z_{,\psi} - \cot\vartheta Z_{,\vartheta}) = 2(Z_{,\psi}^2 - Z_{,\vartheta}^2) \tag{A15}$$

or

$$(1 - EE^*)(E_{,\psi\psi} - E_{,\vartheta\vartheta} + \cot\psi E_{,\psi} - \cot\vartheta E_{,\vartheta}) = -2E^*(E_{,\psi}^2 - E_{,\vartheta}^2). \tag{A16}$$

From every solution of the previous equations the function Φ of Eq. (A4) is then determined from E by

$$\Phi_{,\psi} = \frac{\sin\vartheta\sin\psi}{(1 - EE^*)^2}\left[\frac{|E_{,\vartheta} + E_{,\psi}|^2}{\sin(\vartheta + \psi)} + \frac{|E_{,\vartheta} - E_{,\psi}|^2}{\sin(\vartheta - \psi)}\right], \tag{A17a}$$

$$\Phi_{,\vartheta} = \frac{\sin\vartheta\sin\psi}{(1 - EE^*)^2}\left[\frac{|E_{,\vartheta} + E_{,\psi}|^2}{\sin(\vartheta + \psi)} - \frac{|E_{,\vartheta} - E_{,\psi}|^2}{\sin(\vartheta - \psi)}\right]. \tag{A17b}$$

The equation satisfied by the stream potential becomes

$$\phi_{,\psi\psi} - \phi_{,\vartheta\vartheta} - \phi_{,\psi}\cot\psi + \phi_{,\vartheta}\cot\vartheta = 0, \tag{A18}$$

while f is determined from ϕ by

$$f_{,\psi} = -\frac{2}{\sin\psi\sin\vartheta}\left[\frac{(\phi_{,\vartheta} + \phi_{,\psi})^2}{\sin(\vartheta + \psi)} + \frac{(\phi_{,\vartheta} - \phi_{,\psi})^2}{\sin(\vartheta - \psi)}\right], \tag{A19a}$$

$$f_{,\vartheta} = -\frac{2}{\sin\psi\sin\vartheta}\left[\frac{(\phi_{,\vartheta} + \phi_{,\psi})^2}{\sin(\vartheta + \psi)} - \frac{(\phi_{,\vartheta} - \phi_{,\psi})^2}{\sin(\vartheta - \psi)}\right]. \tag{A19b}$$

APPENDIX B: INTERPLAY BETWEEN SOLUTIONS FOR X AND ϕ

Equations (3.3) and (2.10) or (3.5) and (3.6), or (3.8) for $n = \frac{1}{2}$ and $n = -\frac{1}{2}$ are remarkably similar but not identical. How are, therefore, their solutions related? The answer is given by the following three theorems, expressing the transformations of the solutions of these equations in the three different coordinate systems that we are using in the paper.

Theorem 1: Let $\mathbf{X}(\eta,\mu)$ be any solution of the equation

$$[(1 - \eta^2)\mathbf{X}_{,\eta}]_{,\eta} - [(1 - \mu^2)\mathbf{X}_{,\mu}]_{,\mu} = 0. \tag{B1}$$

Then

$$\phi(\eta,\mu) = (1 - \eta^2)(1 - \mu^2)\mathbf{X}_{,\eta\mu} \tag{B2}$$

is a solution of the equation

$$(1 - \eta^2)\phi_{,\eta\eta} - (1 - \mu^2)\phi_{,\mu\mu} = 0. \tag{B3}$$

Conversely, if $\phi(\eta,\mu)$ is a solution of Eq. (B3) then

$$\mathbf{X}(\eta,\mu) = \phi_{,\eta\mu} \tag{B4}$$

is a solution of Eq. (B1).

By transforming to the (u,v) coordinates (3.4) we get the following theorem.

Theorem 2: Let $\mathbf{X}(u,v)$ be any solution of the equation

$$(1 - u^2 - v^2)\mathbf{X}_{,uv} - (v\mathbf{X}_{,u} + u\mathbf{X}_{,v}) = 0. \tag{B5}$$

Then

$$\phi(u,v) = (1 - u^2 - v^2)[(1 - u^2)\mathbf{X}_{,uu} - (1 - v^2)\mathbf{X}_{,vv} - u\mathbf{X}_{,u} + v\mathbf{X}_{,v}] \tag{B6}$$

is a solution of the equation

$$(1 - u^2 - v^2)\phi_{,uv} + v\phi_{,u} + u\phi_{,v} = 0. \tag{B7}$$

Conversely, if $\phi(u,v)$ is a solution of Eq. (B7), then

$$\mathbf{X}(u,v) = (1 - u^2 - v^2)^{-1}[(1 - u^2)\phi_{,uu} - (1 - v^2)\phi_{,vv} - u\phi_{,u} + v\phi_{,v}] \tag{B8}$$

is a solution of Eq. (B5).

Finally, in the (x,y) coordinates (3.7) the transformation is described by the following theorem.

Theorem 3: Let $\mathbf{X}(x,y)$ be any solution of the equation

$$2(x + y)\mathbf{X}_{,xy} + \mathbf{X}_{,x} + \mathbf{X}_{,y} = 0. \tag{B9}$$

Then

$$\phi(x,y) = (x + y)[(1 - x^2)\mathbf{X}_{,xx} - (1 - y^2)\mathbf{X}_{,yy} - x\mathbf{X}_{,x} + y\mathbf{X}_{,y}] \tag{B10}$$

is a solution of the equation

$$2(x + y)\phi_{,xy} - \phi_{,x} - \phi_{,y} = 0. \tag{B11}$$

Conversely, if $\phi(x,y)$ is a solution of Eq. (B11), then

$$\mathbf{X}(x,y) = (x + y)^{-1}[(1 - x^2)\phi_{,xx} - (1 - y^2)\phi_{,yy} - x\phi_{,x + y}\phi_{,y}] \tag{B12}$$

is a solution of Eq. (B9).

Theorem 1 was obtained by noting that Eqs. (B1) and (B3) admit separable solutions in the (η,μ) variables, which are expressible in terms of Legendre functions. These solutions are

$$\mathbf{X}(\eta,\mu) = Y_m(\eta)Y_m(\mu) \tag{B13}$$

and

$$\phi(\eta,\mu) = (1 - \eta^2)(1 - \mu^2)\dot{Y}_m(\eta)\dot{Y}_m(\mu) \tag{B14}$$

(and any superposition of them) where the Y_m's are solutions of Legendre's equation of order m and the dots denote differentiations with respect to the corresponding arguments.

The transformations of Theorem 1 map the solutions (B13) and (B14) one onto the other by Legendre's equation. The transformations of Theorems 2 and 3 express that of Theorem 1 in the systems of null coordinates (u,v) and (x,y). The three theorems can be proved directly as well, by substituting into the relevant equation to get a fourth-order expression and then reducing it by virtue of the other second-order equation.

APPENDIX C: SEPARABLE SOLUTIONS IN TERMS OF BESSEL FUNCTIONS

In Ref. 6, Sec. 7, we found that Eq. (3.6) admits separable solutions expressible in terms of Bessel functions, for which Eqs. (A9) were explicitly integrated. We shall here show that Eq. (3.5) for X also admits separable solutions expressible interms of Bessel functions.

The variables in which Eq. (3.5) separates are

$$r = u^2 - v^2, \quad s = 1 - u^2 - v^2, \tag{C1}$$

i.e., the same as for Eq. (3.6). Then Eq. (3.5) becomes

$$\mathbf{X}_{,rr} - (\mathbf{X}_{,ss} + (1/s)\mathbf{X}_{,s}) = 0, \tag{C2}$$

while Eqs. (A5) expressed in terms of

$$\mathbf{X} = \ln \chi = \ln[(1+E)/(1-E)] \tag{C3}$$

and in the (r,s) variables become

$$\Phi_{,r} + \Phi_{,s} = (s/2)(\mathbf{X}_{,r} + \mathbf{X}_{,s})^2, \tag{C4a}$$

$$\Phi_{,r} - \Phi_{,s} = -(s/2)(\mathbf{X}_{,r} - \mathbf{X}_{,s})^2. \tag{C4b}$$

Hence

$$\Phi_{,r} = s\mathbf{X}_{,r}\mathbf{X}_{,s} \quad \text{and} \quad \Phi_{,s} = (s/2)(\mathbf{X}_{,r}^2 + X_{,s}^2). \tag{C5}$$

The fundamental solutions of Eq. (C2) are

$$\mathbf{X} = e^{\pm \alpha r}\ell_0(i\alpha s) \quad \text{and} \quad \mathbf{X} = e^{\pm i\alpha r}\ell_0(\alpha s), \tag{C6}$$

where $\pm\alpha^2$ is the separation constant and $\ell_0(i\alpha s)$ [which can be either $K_0(\alpha s)$ or $I_0(\alpha s)$] and $\ell_0(\alpha s)$ [which can be either $J_0(\alpha s)$ or $Y_0(\alpha s)$]) denote Bessel functions of order zero for an imaginary or a real argument, respectively. The general solution for X is expressible as an arbitrary linear superposition (of sums or integrals over the parameters) of the fundamental solutions with different separation constants. It remains to be investigated whether all of these fundamental solutions are permissible.

There is no difficulty in solving for the Φ corresponding to each of the fundamental solutions. By taking, for instance,

$$\mathbf{X} = e^{\pm \alpha r}K_0(\alpha s), \tag{C7}$$

we have

$$\mathbf{X}_{,r} = \pm \alpha e^{\pm \alpha r}K_0(\alpha s) \quad \text{and} \quad \mathbf{X}_{,s} = -\alpha e^{\pm \alpha r}K_1(\alpha s). \tag{C8}$$

It can be readily verified that the corresponding solution for Φ is

$$\Phi = -\tfrac{1}{2} e^{\pm 2\alpha r}\alpha s K_0(\alpha s)K_1(\alpha s). \tag{C9}$$

APPENDIX D: IDENTITIES AMONG LEGENDRE FUNCTIONS

Several mathematical identities involving Legendre's "ring" functions can be obtained from the solution of the characteristic problem outlined in Sec. IV. They arise by using known solutions of Eq. (3.8) and substituting their characteristic data in the expression (4.2). We shall only give some examples.

Equation (3.8) admits the solution $U=$ const throughout the triangle AOB, which has characteristic data $U=$ const, $U_{,x} = U_{,y} = 0$. Substitution gives the identity

$$\int_1^x (1+\xi)^{n-1}P_{-n}\left(1 + \frac{2(\xi - x)(1-y)}{(1+\xi)(x+y)}\right)d\xi + \int_1^y (1+\xi)^{n-1}P_{-n}\left(1 + \frac{2(\xi - y)(1-x)}{(1+\xi)(x+y)}\right)d\xi = \frac{(x+y)^n}{n} - \frac{2^n}{n}P_{-n}\left(\frac{1+xy}{x+y}\right), \tag{D1}$$

for every index n, not necessarily an integer. Special cases like $n = \pm 1$ or arbitrary n but $x=1$ or $y=1$ can be easily verified. But we do not know how to prove the identity generally, except by the indirect method based on the theory of Sec. IV.

By using that

$$P_n(z) = \left(\frac{1+z}{2}\right)^n F\left(-n, -n;1;\frac{z-1}{z+1}\right), \quad \operatorname{Re}(z) > 0, \tag{D2}$$

we obtain from Eq. (D1) the identity

$$(1+y)^n \int_1^x \frac{(1+\xi)^{2n-1}}{(\xi+y)^n} F\left(n,n;1;\frac{(\xi-x)(1-y)}{(\xi+y)(1+x)}\right)d\xi + (1+x)^n \times \int_1^y \frac{(1+\xi)^{2n-1}}{(\xi+x)^n} F\left(n,n;1;\frac{(\xi-y)(1-x)}{(\xi+x)(1+y)}\right)d\xi = \frac{1}{n}(1+x)^n(1+y)^n - \frac{1}{n}2^{2n}F\left(n,n;1;\frac{(1-x)(1-y)}{(1+x)(1+y)}\right), \tag{D3}$$

involving hypergeometric functions.

More involved identities, involving integrals of products of Legendre functions, can be obtained by considering the solutions (B13) and (B14) of Eq. (3.8) for $n = \pm\frac{1}{2}$.

For instance, for $n = -\frac{1}{2}$

$$U = (1-\eta^2)(1-\mu^2)\dot{Y}_m(\eta)\dot{Y}_m(\mu) \tag{D4}$$

is a solution for every index m, where the Y_m's are solutions of Legendre's equation of order m, and the dots denote differentiations. For simplicity we shall assume m to be a positive integer and take Legendre polynomials P_m's for the Y_m's. Then by using that

$$x = 1 \Leftrightarrow u = 0 \Leftrightarrow (\eta = v, \mu = -v),$$

$$\left(\frac{\partial\eta}{\partial v}\right)_{u=0} = -\left(\frac{\partial\mu}{\partial v}\right)_{u=0} = 1, \tag{D5}$$

$$(U_{,v})_{u=0} = (U_{,\eta} - U_{,\mu})_{u=0} = 2m(m+1)(1-v^2)(-1)^m P_m(v)\dot{P}_m(v),$$

we find that

$$\left(U_{,y}-\frac{U}{2(1+\xi)}\right)_{(1,\xi)}=\tfrac{1}{8}(-1)^{m-1}(1+\xi)\left\{\frac{\dot{P}_m(v)}{v}\,[2m(m+1)P_m(v)-v\,\dot{P}_m(v)]\right\}_{v=\sqrt{(1-\xi)/2}}. \tag{D6}$$

Similarly, by using that

$$\begin{aligned}&y=1\Leftrightarrow v=0\Leftrightarrow\eta=\mu=u,\\&\left(\frac{\partial\eta}{\partial u}\right)_{v=0}=\left(\frac{\partial\mu}{\partial u}\right)_{v=0}=1,\\&(U_{,u})_{v=0}=(U_{,\eta}+U_{,\mu})_{v=0}=-2m(m+1)(1-u^2)P_m(u)\dot{P}_m(u),\end{aligned} \tag{D7}$$

we obtain

$$\left(U_{,x}-\frac{U}{2(1+\xi)}\right)_{(\xi,1)}=\tfrac{1}{8}(1+\xi)\left\{\frac{\dot{P}_m(u)}{u}\,[2m(m+1)P_m(u)-u\,\dot{P}_m(u)]\right\}_{u=\sqrt{(1-\xi)/2}}. \tag{D8}$$

Moreover

$$(1-\eta^2)(1-\mu^2)=\tfrac{1}{4}(x+y)^2 \tag{D9}$$

and

$$U(x=1,y=1)=U(\eta=0,\mu=0)=\begin{cases}\dot{P}^2_m(0), & \text{for } m=\text{odd},\\ 0, & \text{for } m=\text{even}.\end{cases} \tag{D10}$$

Substitution into Eq. (4.2) gives the identity in x and y, for every positive integer m,

$$\begin{aligned}&\int_x^1\sqrt{1+\xi}\,G_m(\xi)P_{1/2}\left(1+\frac{2(\xi-x)(1-y)}{(1+\xi)(x+y)}\right)d\xi+(-1)^{m-1}\\&\times\int_y^1\sqrt{1+\xi}\,G_m(\xi)P_{1/2}\left(1+\frac{2(\xi-y)(1-x)}{(1+\xi)(x+y)}\right)d\xi=4\sqrt{2}\dot{P}^2_m(0)P_{1/2}\left(\frac{1+xy}{x+y}\right)-2(x+y)^{3/2}\dot{P}_m(\eta)\dot{P}_m(\mu),\end{aligned} \tag{D11}$$

where

$$G_m(\xi)=\left\{\frac{\dot{P}_m(t)}{t}\,[2m(m+1)P_m(t)-t\dot{P}_m(t)]\right\}_{(t=\sqrt{(1-\xi)/2})} \tag{D12}$$

and η and μ are given by Eqs. (3.4) and (3.7). Note that the product $\dot{P}_m(\eta)\,\dot{P}_m(\mu)$ is rational in x and y. For instance, we have calculated that

$$\dot{P}_m(\eta)\dot{P}_m(\mu)=\begin{cases}\frac{9}{2}(y-x), & \text{for } m=2,\\ \frac{9}{16}(25x^2+25y^2-30xy-16), & \text{for } m=3,\\ \frac{25}{32}(y-x)(49x^2-14xy+49y^2-48), & \text{for } m=4.\end{cases} \tag{D13}$$

Similarly, for $n=\frac{1}{2}$ we know that

$$U=Y_m(\eta)Y_m(\mu) \tag{D14}$$

is a solution of Eq. (3.8), where, again, the Y_m's are Legendre's functions of index m. And as before, we shall assume for simplicity that m is a positive integer and the Y_m's are the corresponding Legèndre polynomials. We now find that

$$(U_{,v})_{u=0}=2(-1)^mP_m(v)\dot{P}_m(v),\quad (U_{,u})_{v=0}=2P_m(u)\dot{P}_m(u), \tag{D15}$$

and therefore that

$$\left(U_{,y}+\frac{U}{2(1+\xi)}\right)_{(1,\xi)}=\frac{(-1)^{m-1}}{2}\left[\frac{P_m(v)\dot{P}_m(v)}{v}-\frac{P^2_m(v)}{1+\xi}\right]_{(v=\sqrt{(1-\xi)/2})}, \tag{D16}$$

$$\left(U_{,x}+\frac{U}{2(1+\xi)}\right)_{(\xi,1)}=-\frac{1}{2}\left[\frac{P_m(u)\dot{P}_m(u)}{u}-\frac{P^2_m(u)}{1+\xi}\right]_{(u=\sqrt{(1-\xi)/2})}. \tag{D17}$$

Substitution into Eq. (4.2) gives the identity

$$\int_x^1 G_m(\xi)\sqrt{1+\xi}\,P_{-1/2}\left(1+\frac{2(\xi-x)(1-y)}{(1+\xi)(x+y)}\right)d\xi + (-1)^m\int_y^1 G_m(\xi)\sqrt{1+\xi}$$

$$\times P_{-1/2}\left(1+\frac{2(\xi-y)(1-x)}{(1+\xi)(x+y)}\right)d\xi = 2\sqrt{x+y}\,P_m(\eta)P_m(\mu) - 2\sqrt{2}\,\dot P_m^2(0)P_{-1/2}\left(\frac{1+xy}{x+y}\right), \tag{D18}$$

where

$$G_m(\xi) = \left[\frac{P_m(t)\dot P_m(t)}{t} - \frac{P_m^2(t)}{1+\xi}\right]_{(t=\sqrt{(1-\xi)/2})}. \tag{D19}$$

The identity (B18) is valid for every real x and y and any positive integer m. The product $P_m(\eta)\,P_m(\mu)$ is rational in x and y as well. For instance,

$$P_m(\eta)P_m(\mu) = \begin{cases} \frac{1}{2}(y-x), & \text{for } m=1, \\ \frac{1}{16}(9x^2+9y^2-6xy-8), & \text{for } m=2, \\ \frac{1}{32}(y-x)(25x^2+25y^2+10xy-24), & \text{for } m=3. \end{cases} \tag{D20}$$

Note also that the expressions $G_m(\xi)$ of Eqs. (D12) and (D19) are rational functions in ξ.

[1]P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[2]S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).
[3]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985).
[4]K. Khan and R. Penrose, Nature **229**, 185 (1971).
[5]Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).
[6]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **402**, 37 (1985).
[7]E. T. Copson, *Partial Differential Equations* (Cambridge U. P., Cambridge, 1975), Chap. 5.

Three-dimensional gauge theory in Dirac formalism

Kiyoshi Kamimura
Department of Physics, Toho University, Funabashi, Japan 274

(Received 3 February 1986; accepted for publication 9 April 1986)

The Hagen model [C. R. Hagen, Ann. Phys. (NY) **157**, 342 (1984); Phys. Rev. D **31**, 331 (1985)] is studied using the method of constrained Hamiltonian formalism developed by Dirac [P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); *Lectures on Quantum Mechanics* (Yeshiva U. P., New York, 1964)]. The results recently obtained by Burnel and Van Der Rest-Jaspers [A. Burnel and M. Van Der Rest-Jaspers, J. Math. Phys. **26**, 3155 (1985)] are reexamined and modified. There appear two second-class constraints and their choice is not crucial. The equivalence of different gauges is proved without referring to the current conservation law.

I. INTRODUCTION

Recently Burnel and Van Der Rest-Jaspers[1] examined the three-dimensional gauge theory proposed by Hagen[2] by applying the Dirac formalism of the constrained Hamiltonian systems.[3] They have concluded from some pathological aspects that there appear three second-class constraints and the structure is quite different from that of the usual gauge theories. However, as it seems there is no exceptional structure in the gauge transformation, the straightfoward application of the Dirac formalism must be possible. In this paper we repeat the study in some detail since the model itself is also interesting as the prototype of the four-dimensional gauge theories. We found that there is no pathological property. There are three primary constraints and one secondary constraint. Two combinations of them are the first class and generate the gauge transformation. The remaining two are the second class though the choice is not unique. The different selections of second-class constraints give the same result. The corresponding variables appearing in the different selections coincide with each other up to additional first-class constraints. The proof of equivalence of the different gauges is refined in the Hamiltonian formalism without using the current conservation law in Abelian theory. In the case of the non-Abelian theory the same difficulty appears as in the Yang–Mills theories in four dimensions.

II. HAMILTONIAN FORMALISM

The Lagrangian of the model given by Hagen[2] is

$$\mathscr{L} = \tfrac{1}{2}\phi^\mu \epsilon_{\mu\nu\rho}\, \partial^\rho \phi^\nu + \phi^\mu j_\mu, \tag{1}$$

where we use the notation of Ref. 1 ($\eta_{\mu\nu} = (+\ -\ -)$, $\epsilon_{012} = \epsilon_{12} = 1,\ldots$). The action is invariant under the Abelian gauge transformation

$$\phi^\mu \to \phi^\mu + \partial^\mu \lambda. \tag{2}$$

Since it depends on $\dot\lambda$ and λ we expect one primary and one secondary first-class constraint in the Hamiltonian formalism.

The momenta conjugate to ϕ^μ are

$$\begin{aligned}\pi_0 &= \frac{\partial \mathscr{L}}{\partial \dot\phi^0} = 0,\\ \pi_j &= \frac{\partial \mathscr{L}}{\partial \dot\phi^j} = -\frac{1}{2}\epsilon_{jk}\phi^k \quad (j=1,2)\end{aligned} \tag{3}$$

and there are three primary constraints

$$K_0 \equiv \pi_0 = 0, \quad K_j = \pi_j + \tfrac{1}{2}\epsilon_{jk}\phi^k = 0. \tag{4}$$

The Hamiltonian is defined without using (4) (see Ref. 4) as

$$\begin{aligned}\mathscr{H} &= \int d^2\mathbf{x}(\pi_\mu \dot\phi^\mu - \mathscr{L})\\ &= \int d^2\mathbf{x}(\pi_0\dot\phi^0 + (\pi_j + \tfrac{1}{2}\epsilon_{jk}\phi^k)\dot\phi^j\\ &\quad - \phi^0\{\partial^2\phi^1 - \partial^1\phi^2 + j_0\} - \phi^j j_j).\end{aligned} \tag{5}$$

The first two terms are linear combinations of primary constraints and their coefficients are set to be undetermined functions in the Dirac Hamiltonian. The consistency condition that the primary constraints are preserved gives

$$\begin{aligned}\dot K_0 &= \partial^2\phi^1 - \partial^1\phi^2 + j_0 \equiv K_3 = 0,\\ \dot K_j &= \epsilon_{jk}(\dot\phi^k - \partial^k\phi^0) + j_k = 0.\end{aligned} \tag{6}$$

The second equation determines $\dot\phi^j$,

$$\dot\phi^j = \partial^j\phi^0 + \epsilon^{jk} j_k, \tag{7}$$

and the first one is the secondary constraint. It is preserved using the current conservation $\partial^\mu j_\mu = 0$, which is the result of the equations of motion of the matter fields. No more constraints appear and $\dot\phi^0$ remains undetermined. Using (7), the Hamiltonian (5) becomes

$$\begin{aligned}\mathscr{H} = \int d\mathbf{x}[\pi_0\dot\phi^0 - \phi^0\{\partial^k\pi_k + \tfrac{1}{2}(\partial^2\phi^1 - \partial^1\phi^2) + j_0\}\\ - (\pi_j - \tfrac{1}{2}\epsilon_{jk}\phi^k) j_k].\end{aligned} \tag{8}$$

The coefficient of ϕ^0 is a combination of the constraints

$$\partial^k\pi_k + \tfrac{1}{2}(\partial^2\phi^1 - \partial^1\phi^2) + j_0 = \partial^k K_k + K_3 \equiv K_4 = 0, \tag{9}$$

and it commutes with all constraints, therefore it is the first-class constraint. It is the Gauss law constraint in the usual gauge theories. Actually the gauge transformation (2) is generated by two first-class constraints K_0 and K_4,

$$G = \int d\mathbf{x}[\dot\lambda K_0 - \lambda K_4]. \tag{10}$$

Since K_4 in (9) is the first-class constraint all of the three K_j's ($j = 1,2,3$) cannot be second-class ones. In fact the rank of $\det\{K_j, K_l\}$ is 2. The second-class constraints are two independent linear combinations of K_1, K_2, and K_3. Two

 0022-2488/86/082137-03$02.50

canonical variables are eliminated by using them after defining the Dirac bracket. Taking K_1 and K_2 as the second-class constraints, the Dirac bracket is defined as

$$\{A,B\}^* = \{A,B\} + \{A,K_1\}\{K_2,B\} - \{A,K_2\}\{K_1,B\}. \tag{11}$$

Eliminating π_j ($j = 1,2$), the values of the Dirac bracket for the remaining variables ϕ^0, ϕ^j, and π_0 are

$$\{\phi^0(x),\pi_0(x)\}^* = -\{\phi^1(x),\phi^2(y)\}^* = \delta^2(\mathbf{x}-\mathbf{y}), \tag{12}$$

and the others are zero. Now the Hamiltonian, in terms of the reduced variables, becomes

$$\mathscr{H}^* = \int d\mathbf{x}[\pi_0\dot{\phi}^0 - \phi^0(\partial^2\phi^1 - \partial^1\phi^2 + j_0) - \phi^k j_k] \tag{13}$$

and the two first-class constraints are

$$K_0^* = \pi_0 = 0, \quad K_4^* = \partial^2\phi^1 - \partial^1\phi^2 + j_0 = 0. \tag{14}$$

The choices of the second-class constraints and the elimination variables are a matter of convenience of the description and irrelevant to the physics. In fact if we choose K_2 and K_3 as the second-class set and eliminate ϕ^2 and π_2, the Dirac bracket, constraints, and the Hamiltonian are

$$\begin{aligned} &\{\tilde{\phi}^1(x),\tilde{\pi}_1(y)\}^* = \delta(\mathbf{x}-\mathbf{y}),\\ &\{\tilde{\pi}_1(\mathbf{x}),\widetilde{\text{matter}}\}^* = \left\{\frac{1}{2\partial^1}j^0(\mathbf{x}),\text{matter}\right\},\\ &\tilde{K}_0^* = \tilde{\pi}_0 = 0, \quad \tilde{K}_4^* = \partial^1\tilde{K}_1 = \partial^1\tilde{\pi}_1 + \tfrac{1}{2}(\partial^2\tilde{\phi}^1 + j_0) = 0, \end{aligned} \tag{15}$$

$$\begin{aligned} \tilde{\mathscr{H}}^* = \int d\mathbf{x}[&\tilde{\pi}_0\dot{\tilde{\phi}}^0 - \tilde{\phi}_0\tilde{K}_4^*\\ &- \tilde{\phi}^1\tilde{j}_1 + (\tilde{\pi}_1 - (1/2\partial^1)(\partial^2\tilde{\phi}^1 + \tilde{j}_0))\tilde{j}_2]. \end{aligned}$$

They look different from (11)–(14) written in terms of ϕ_1 and π_1

$$\begin{aligned} &\{\phi^1(x),\pi_1(y)\}^* = \tfrac{1}{2}\delta(\mathbf{x}-\mathbf{y}), \quad \{\pi_1(\mathbf{x}),\text{matter}\}^* = 0,\\ &K_0^* = \pi_0 = 0, \quad K_4^* = K_3 = 2\partial^1\pi_1 + (\partial^2\phi^1 + j_0) = 0, \end{aligned} \tag{16}$$

$$\mathscr{H}^* = \int d\mathbf{x}[\pi_0\dot{\phi}^0 - \phi^0 K_4^* - \phi^1 j_1 - 2\pi_1 j_2].$$

The difference comes from the ambiguity of the first-class constraints in various quantities. To take K_3 as the second-class constraint means to take $K_1 - (1/\partial^1)K_4$ in place of K_1, since

$$K_3 = -\partial^1(K_1 - (1/\partial^1)K_4) - \partial^2 K_2. \tag{17}$$

This change shifts the definition of the variables by a certain amount of an additional first-class constraint. In this case the quantities with tildas in (15) are related to the ones without tildas in (16) by

$$A = \tilde{A} + \{A,K_2\}(1/\partial^1)\tilde{K}_4, \tag{18}$$

or, more explicitly,

$$\begin{aligned} &\phi^1 = \tilde{\phi}^1, \quad (\text{matter}) = (\widetilde{\text{matter}}),\\ &\pi_1 = \tilde{\pi}_1 - (1/2\partial^1)\tilde{K}_4^*\\ &\quad = \tfrac{1}{2}\tilde{\pi}_1 - (1/4\partial^1)(\partial^2\tilde{\phi}^1 + \tilde{j}_0). \end{aligned} \tag{19}$$

We can show that (15) is derived from (16) and vice versa. The origin of (18) is understood from the fact that to define the Dirac bracket is equivalent to replacing A by its starred variable,[5]

$$A^* = A - \{A,K_\alpha\}C_{\alpha\beta}^{-1}K_\beta, \tag{20}$$

where the K's are the second class constraints and $C_{\alpha\beta} = \{K_\alpha,K_\beta\}$. The difference of A and $\tilde{A}$ in the expression (18) is that of the starred quantities defined in (20). Since the difference is the first-class constraint, they are physically equivalent. In fact, when the gauge is fixed and the constraints are eliminated the difference will disappear completely.

III. AXIAL GAUGE

The gauge freedom is fixed by setting two gauge conditions; it makes two first-class constraints, generators of the gauge transformation, to be second-class ones. The axial gauge is defined by

$$\chi_1 \equiv n_1\phi^1 + n_2\phi^2 = 0, \quad (n_1^2 + n_2^2 = 1). \tag{21}$$

The condition that χ_1 is conserved in time further requires

$$\dot{\chi}_1 = (n_j\partial^j)\phi^0 + (n_1 j_2 - n_2 j_1) = 0. \tag{22}$$

It is the constraint on ϕ^0 [by assuming the inverse of $(n\cdot\partial)$ as in Refs. 1 and 2];

$$\chi_2 = \phi^0 + [1/(n\cdot\partial)](n_1 j_2 - n_2 j_1) = 0. \tag{23}$$

The consistency determines $\dot{\phi}^0$ in the Hamiltonian. Now that all the four constraints are in the second class and all components of the gauge field are eliminated, we have (23) and

$$\pi_0 = 0, \quad \phi^1 = \frac{-n_2}{(n\cdot\partial)}j_0, \quad \phi^2 = \frac{n_1}{(n\cdot\partial)}j_0. \tag{24}$$

In this choice of gauge, the values of the Dirac brackets coincide with those of Poisson brackets for the matter fields. The Hamiltonian in this gauge becomes

$$\mathscr{H}^{**} = \int d\mathbf{x}\left[(n_2 j_1 - n_1 j_2)\frac{1}{(n\cdot\partial)}j_0\right]. \tag{25}$$

The different choice of n_j gives a different form of the Hamiltonian. In Ref. 1, Burnel and Van Der Rest-Jaspers proved the equivalence of the different gauges by a formal discussion using the current conservation $\partial^\mu j_\mu = 0$. We give an alternative proof in the Hamiltonian formalism without referring to current conservation. By a small change of $n_1 = \cos\theta$ and $n_2 = \sin\theta$,

$$\delta\mathscr{H}^{**} = \int d\mathbf{x}\left[-(\partial^1 j_1 + \partial^2 j_2)\frac{1}{(n\cdot\partial)^2}j_0\right]\delta\theta. \tag{26}$$

It never vanishes and $\mathscr{H}^{**}$ itself is actually gauge dependent. The point is that the total Hamiltonian, the sum of $\mathscr{H}^{**}$ and the matter Hamiltonian, is gauge independent. To see the gauge dependence of the matter fields we again use the starred variables introduced in (20). For the matter field ψ it is

$$\psi^* = \psi - \{\psi,j_0\}[-1/(n\cdot\partial)](n_1\phi^1 + n_2\phi^2). \tag{27}$$

Under the change of n_j, it is varied as

$$\delta\psi^* = ie\delta\theta\left(\frac{1}{(n\cdot\partial)^2}(-\partial^2\phi^1 + \partial^1\phi^2)\right)\psi$$

$$= ie\delta\theta\left(\frac{1}{(n\cdot\partial)^2}j_0\right)\psi. \tag{28}$$

The conjugate momentum of ψ is varied correspondingly. The change of the matter Hamiltonian is

$$\delta\mathscr{H}_M = \int d^2\mathbf{x}\,\frac{\partial\mathscr{H}}{\partial\partial^k\psi}\,\partial^k\left(ie\delta\theta\frac{1}{(n\cdot\partial)^2}j_0\right)\psi$$

$$= \int d^2\mathbf{x}(\partial^k j_k)\,\frac{1}{(n\cdot\partial)^2}j_0\delta\theta, \tag{29}$$

which cancels exactly with (26). Then the total Hamiltonian is gauge independent. In other words, the change of $\mathscr{H}^{**}$ is absorbed into the matter Hamiltonian by the redefinition of phase of the matter fields. The finite transformation is obtained by integrating the infinitesimal ones. For example, consider the following transformation:

$$\psi = e^{ie(1/\Delta)\partial^j\phi^j}\,\tilde{\psi}. \tag{30}$$

In terms of $\tilde{\psi}$, the matter Hamiltonian is written as

$$\mathscr{H}_M(\psi) = \mathscr{H}_M(\tilde{\psi}) + \int d\mathbf{x}(\partial^k j_k)\,\frac{1}{\Delta}\,\frac{\partial^1 n_2 - \partial^2 n_1}{(n\cdot\partial)}j_0, \tag{31}$$

and the total Hamiltonian becomes, in the Coulomb gauge,

$$\mathscr{H}_{\text{tot}} = \mathscr{H}_M(\tilde{\psi}) + \int d\mathbf{x}\left[\tilde{j}_1\frac{\partial^2}{\Delta}\tilde{j}_0 - \tilde{j}_2\frac{\partial^1}{\Delta}\tilde{j}_0\right]. \tag{32}$$

It must be noticed that we are still in the axial gauge fixed by the constraints (20) and (21), though we have the form of a Coulomb gauge Hamiltonian.

IV. SUMMARY AND DISCUSSIONS

In this paper we have formulated Hagen model using the Dirac formalism. Two second class constraints appeared. Although their choice has an ambiguity of additive first-class constraints, it causes no physical inequivalence. We showed it explicitly for two cases examined in Ref. 1 using starred variables. We also have given the proof of the equivalence of different gauge conditions. The proof of Ref. 1 uses the current conservation and is not within the canonical formalism. In fact, the Hamiltonian $\mathscr{H}^{**}$ is not invariant by itself but the matter field Hamiltonian must be considered.

The non-Abelian theory is more interesting. Hagen[2] has reported the Lorentz noninvariance of the axial gauges. He showed it in the Lagrangian formalism and the perturbation theory. The formal extension of the canonical formalism to non-Abelian theory is straightforward. Especially in the axial gauge, the non-Abelian nature of the constraint K_4^* disappears,

$$K_4^* = \partial^2\phi_a^1 - \partial^1\phi_a^2 + gf_{abc}\phi_b^2\phi_c^1 + j_{0a}$$

$$= \partial^2\phi_a^1 - \partial^1\phi_a^2 + j_{0a}, \tag{33}$$

and all components of the gauge field are solved explicitly as in (23) and (24):

$$\phi_a^0 = [-1/(n\cdot\partial)]\,(n_1 j_{2a} - n_2 j_{1a}),$$

$$\phi_a^1 = [-n_2/(n\cdot\partial)]\,j_{0a},\quad \phi_a^2 = [n_1/(n\cdot\partial)]\,j_{0a}. \tag{34}$$

The Hamiltonian takes the same form as Abelian theory in (25):

$$\mathscr{H}^{**} = \int d^2\mathbf{x}\left[(n_2 j_{1a} - n_1 j_{2a})\frac{1}{(n\cdot\partial)}j_{0a}\right]. \tag{35}$$

The non-Abelian nature appears only in the current commutators.

The formal proofs that the different choices of the second-class constraints and the gauge-fixing constraints give an equivalent result, shown in the previous sections, seem valid for the infinitesimal transformations for the non-Abelian theory. For the finite transformations, for example, from axial gauge to Coulomb gauge, the equivalence is not evident due to the noncommutativity of currents. However these statements must be reexamined from the topological discussions of the non-Abelian gauge theory.[6] So far we have assumed the existence of $(n\cdot\partial)^{-1}$ and its partial integration. It is not allowed, however, for the non-Abelian theory generally. Even the pure gauge configuration does not mean the vanishing of the fields at infinity. It was shown[7] that no gauge fixing, Coulomb as well as axial gauges, can be allowed for the non-Abelian theories from the topological arguments. It will fail the previous formal proofs of equivalence and further investigations are required.

[1] A. Burnel and M. Van Der Rest-Jaspers, J. Math. Phys. **26**, 3155 (1985).
[2] C. R. Hagen, Ann. Phys. (NY) **157**, 342 (1984); Phys. Rev. D **31**, 331 (1985).
[3] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); *Lectures on Quantum Mechanics* (Yeshiva U. P., New York, 1964).
[4] K. Kamimura, Nuovo Cimento B **68**, 33 (1982).
[5] A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionaledei Lincei, Rome, 1976).
[6] V. N. Gribov, Nucl. Phys. B **139**, 1 (1978).
[7] I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).

Infinite hierarchies of *t*-independent and *t*-dependent conserved functionals of the Federbush model

P. H. M. Kersten
Department of Applied Mathematics, Twente University of Technology, P. O. Box 217, 7500 AE Enschede, The Netherlands

H. M. M. Ten Eikelder
Department of Mathematics and Computing Science, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 6 December 1985; accepted for publication 6 March 1986)

The construction of four infinite hierarchies of t-independent and t-dependent conserved functionals for the Federbush model is given. A formal proof of the existence of these infinite hierarchies is given in Appendix B.

I. INTRODUCTION

In a recent paper[1] one of the authors constructed four infinite hierarchies of Lie–Bäcklund transformations of the Federbush model.[2,3] Moreover he computed four creating and annihilating local (x,t)-dependent Lie–Bäcklund transformations that lead to these hierarchies. In this paper we show that to these four creating Lie–Bäcklund transformations, we can associate four t-dependent conserved functionals. By consequence the attempt to construct recursion[4,5] operators from these creating Lie–Bäcklund transformations failed since they are Hamiltonian vector fields. By recursive action of the Poisson bracket with these functionals we construct infinite hierarchies of conserved functionals associated to the (x,t)-independent Lie–Bäcklund transformations. This will be done in Sec. II. In Sec. III we construct four new (x,t)-dependent Lie–Bäcklund transformations from which we shall prove the existence of four infinite hierarchies of t-dependent conserved functionals, and consequently hierarchies of (x,t)-dependent Lie–Bäcklund transformations of the Federbush model. A formal proof is given in Appendix B, while a survey of the already known vector fields is given in Appendix A.

We want to stress the fact that all computations have been worked out on a DEC-system 20 computer using REDUCE[6] and a software package[7,8] to do these calculations.

Lie–Bäcklund transformations are vector fields V defined on the infinite jet bundle[9] of M,N, $J^\infty(M,N)$, where M is the space of independent variables and N the space of the dependent variables. A Lie–Bäcklund transformation of a differential equation is a vector field V defined on $J^\infty(M,N)$ satisfying the condition

$$\mathscr{L}_V(D^\infty I)\subset D^\infty I, \tag{1.1}$$

where I denotes a differential ideal associated to the differential equation at hand, while $D^\infty I$ denotes its infinite prolongation to $J^\infty(M,N)$; $\mathscr{L}_V$ is the Lie derivative with respect to the vector field V (Ref. 9). Since the vector fields V are supposed to depend only on a finite number of variables, condition (1.1) reduces to

$$\mathscr{L}_V I\subset D^r I \quad \text{for some } r. \tag{1.2}$$

Using this method we computed Lie–Bäcklund transformation of the Federbush model.[1]

It can be shown that the Lie–Bäcklund transformations in this setting are just symmetries in the works of Magri,[4] Ten Eikelder,[4,5] and Fuchssteiner and Fokas,[10] where (generators of) symmetries of partial differential equations of evolutionary type are described as transformations on special types of infinite dimensional spaces. Suppose that

$$\frac{du}{dt}=\Omega^{-1}\,dH \tag{1.3}$$

is an infinite dimensional Hamiltonian system, where Ω is the symplectic operator, H the Hamiltonian, dH is the Fréchet derivative of H. Then to each Hamiltonian symmetry (also called canonical symmetry) Y, there corresponds by definition a Hamiltonian $F(Y)$ such that

$$Y=\Omega^{-1}\,dF(Y), \tag{1.4}$$

and the Poisson bracket of F and H vanishes.[4,5] Suppose that Y_1, Y_2 are two Hamiltonian symmetries, then $[Y_1,Y_2]$ is a Hamiltonian symmetry and

$$F([Y_2,Y_1])=\{F(Y_1),F(Y_2)\}, \tag{1.5}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket defined by

$$\{F(Y_1),F(Y_2)\}=\langle dF(Y_1),Y_2\rangle, \tag{1.6}$$

where $\langle\cdot,\cdot\rangle$ denotes the contraction of a one-form and a vector field. These notions shall be used throughout Sec. II and III.

II. CONSERVED FUNCTIONALS FOR THE FEDERBUSH MODEL

We shall discuss conserved functionals for the Federbush model. This model is described by

$$\begin{pmatrix} i(\partial_t+\partial_x) & -m(s)\\ -m(s) & i(\partial_t-\partial_x)\end{pmatrix}\begin{pmatrix}\psi_{s,1}\\ \psi_{s,2}\end{pmatrix} = 4s\pi\lambda\begin{pmatrix}|\psi_{-s,2}|^2 & \psi_{s,1}\\ |\psi_{-s,1}|^2 & \psi_{s,2}\end{pmatrix}\quad (s=\pm 1), \tag{2.1}$$

where $\psi_s(x,t)$ are two-component complex-valued functions.[3] Suppressing the factor 4π ($\lambda'=4\pi\lambda$) and introducing the eight real variables $u_1,v_1,u_2,v_2,u_3,v_3,u_4,v_4$ by

$$\begin{aligned}&\psi_{1,1}=u_1+iv_i,\quad \psi_{-1,1}=u_3+iv_3,\quad m(+1)=m_1,\\ &\psi_{1,2}=u_2+iv_2,\quad \psi_{-1,2}=u_4+iv_4,\quad m(-1)=m_2,\end{aligned} \tag{2.2}$$

 0022-2488/86/082140-06$02.50

Eq. (2.1) is rewritten as a system of eight nonlinear partial differential equations for the functions $u_1,...,v_4$; i.e.,

$$
\begin{aligned}
&u_{1t} + u_{1x} - m_1 v_2 = \lambda R_4 v_1,\\
&-v_{1t} - v_{1x} - m_1 u_2 = \lambda R_4 u_1,\\
&u_{2t} - u_{2x} - m_1 v_1 = -\lambda R_3 v_2,\\
&-v_{2t} + v_{2x} - m_1 u_1 = -\lambda R_3 u_2,\\
&u_{3t} + u_{3x} - m_2 v_4 = -\lambda R_2 v_3,\\
&-v_{3t} - v_{3x} - m_2 u_4 = -\lambda R_2 u_3,\\
&u_{4t} - u_{4x} - m_2 v_3 = \lambda R_1 v_4,\\
&-v_{4t} + v_{4x} - m_2 u_3 = \lambda R_1 u_4,
\end{aligned}
\tag{2.3}
$$

where, in (2.3),

$$
R_1 = u_1^2 + v_1^2, \quad R_2 = u_2^2 + v_2^2,
$$
$$
R_3 = u_3^2 + v_3^2, \quad R_4 = u_4^2 + v_4^3.
$$

Equation (2.3) can be written as a Hamiltonian system[4,5]

$$
\frac{du}{dt} = \Omega^{-1}\, dH, \tag{2.4}
$$

whereas in (2.4), $u = (u_1,v_1,...,u_4,v_4)$,

$$
\Omega = \begin{pmatrix} J & & & 0 \\ & J & & \\ & & J & \\ 0 & & & J \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.5a}
$$

and

$$
\begin{aligned}
H = \int_{-\infty}^{\infty} & \tfrac{1}{2}\{u_{1x}v_1 - u_1 v_{1x} - u_{2x}v_2 + u_2 v_{2x}\\
&+ u_{3x}v_3 - u_3 v_{3x} - u_{4x}v_4 + u_4 v_{4x}\}\\
&- m_1(u_1u_2 + v_1v_2) - m_2(u_3u_4 + v_3v_4)\\
&- (\lambda/2)R_1R_4 + (\lambda/2)R_2R_3
\end{aligned}
\tag{2.5b}
$$

(by $\int_{-\infty}^{\infty}$ we mean integration of the integrand with respect to x). In (2.4), dH is the Fréchet derivative of H defined by

$$
\frac{d}{d\epsilon} H(x+\epsilon y)\Big|_{\epsilon=0} = \langle dH, y\rangle. \tag{2.6}
$$

In a previous paper[1] we constructed four first-order Lie–Bäcklund transformations Y_1^+, Y_{-1}^+, Y_1^-, Y_{-1}^- (Appendix A) that are Hamiltonian[4,5] vector fields; the associated Hamiltonian densities are given by

$$
\begin{aligned}
\widetilde{F}(Y_1^+) &= -\tfrac{1}{2}(u_{2x}v_2 - u_2v_{2x}) + (\lambda/4)R_{34}R_2\\
&\quad - \tfrac{1}{2}m_1(u_1u_2 + v_1v_2),\\
\widetilde{F}(Y_{-1}^+) &= -\tfrac{1}{2}(u_{1x}v_1 - u_1v_{1x}) + (\lambda/4)R_{34}R_1\\
&\quad + \tfrac{1}{2}m_1(u_1u_2 + v_1v_2),\\
\widetilde{F}(Y_1^-) &= -\tfrac{1}{2}(u_{4x}v_4 - u_4v_{4x}) - (\lambda/4)R_{12}R_4\\
&\quad - \tfrac{1}{2}m_2(u_3u_4 + v_3v_4),\\
\widetilde{F}(Y_{-1}^-) &= -\tfrac{1}{2}(u_{3x}v_3 - u_3v_{3x}) - (\lambda/4)R_{12}R_3\\
&\quad + \tfrac{1}{2}m_2(u_3u_4 + v_3v_4),
\end{aligned}
\tag{2.7a}
$$

while the Hamiltonian densities associated to the gauge transformations Y_0^+, Y_0^- (see Ref. 1 and Appendix A) are given by

$$
\widetilde{F}(Y_0^+) = \tfrac{1}{2}(R_1 + R_2), \quad \widetilde{F}(Y_0^-) = \tfrac{1}{2}(R_3 + R_4). \tag{2.7b}
$$

In (2.7a), R_{12}, R_{34} are defined by

$$
R_{12} = R_1 + R_2, \quad R_{34} = R_3 + R_4. \tag{2.8}
$$

(Note that we use $\widetilde{F}$ for the density of the conserved functional F, so $F = \int_{-\infty}^{\infty} \widetilde{F}$.) The associated Lie–Bäcklund transformations can be derived from (2.7a) by the formula

$$
Y = \Omega^{-1}\, dF(Y), \tag{2.9}
$$

and for reasons of completeness they are surveyed in Appendix A at the end of this paper. The Hamiltonian densities associated with the second-order Lie–Bäcklund transformations Y_2^+, Y_{-2}^+, Y_2^-, Y_{-2}^- (see Ref. 1 and Appendix A) are computed, yielding

$$
\begin{aligned}
\widetilde{F}(Y_2^+) &= -\tfrac{1}{2}(u_{2x}^2 + v_{2x}^2) + (\lambda/2)R_{34}(u_{2x}v_2 - u_2v_{2x})\\
&\quad - \tfrac{1}{2}m_1(u_{2x}v_1 - u_1v_{2x}) - \tfrac{1}{8}\lambda^2R_{34}^2R_2\\
&\quad + \tfrac{1}{4}m_1\lambda R_{34}(u_1u_2 + v_1v_2) - \tfrac{1}{8}m_1^2R_{12},\\
\widetilde{F}(Y_{-2}^+) &= -\tfrac{1}{2}(u_{1x}^2 + v_{1x}^2) + (\lambda/2)R_{34}(u_{1x}v_1 - u_1v_{1x})\\
&\quad + \tfrac{1}{2}m_1(u_{1x}v_2 - u_2v_{1x}) - \tfrac{1}{8}\lambda^2R_{34}^2R_1\\
&\quad - \tfrac{1}{4}m_1\lambda R_{34}(u_1u_2 + v_1v_2) - \tfrac{1}{8}m_1^2R_{12},\\
\widetilde{F}(Y_2^-) &= -\tfrac{1}{2}(u_{4x}^2 + v_{4x}^2) - (\lambda/2)R_{12}(u_{4x}v_4 - u_4v_{4x})\\
&\quad - \tfrac{1}{2}m_2(u_{4x}v_3 - u_3v_{4x}) - \tfrac{1}{8}\lambda^2R_{12}^2R_4\\
&\quad - \tfrac{1}{4}m_2\lambda R_{12}(u_3u_4 + v_3v_4) - \tfrac{1}{8}m_2^2R_{34},\\
\widetilde{F}(Y_{-2}^-) &= -\tfrac{1}{2}(u_{3x}^2 + v_{3x}^2) - (\lambda/2)R_{12}(u_{3x}v_3 - u_3v_{3x})\\
&\quad + \tfrac{1}{2}m_2(u_{3x}v_4 - u_4v_{3x}) - \tfrac{1}{8}\lambda^2R_{12}^2R_3\\
&\quad + \tfrac{1}{4}m_2\lambda R_{12}(u_3u_4 - v_3v_4) - \tfrac{1}{8}m_2^2R_{34}.
\end{aligned}
\tag{2.10}
$$

The Hamiltonian densities associated to the vector fields Y_3^+, Y_{-3}^+ (see Ref. 1) are computed to be

$$
\begin{aligned}
\widetilde{F}(Y_3^+) = &-(u_{2xx}v_{2x} - u_{2x}v_{2xx}) - \lambda R_{34}(u_{2xx}u_2 + v_{2xx}v_2) + (\lambda/2)R_{34}(u_{2x}^2 + v_{2x}^2) - m_1(u_{1x}u_{2x} + v_{1x}v_{2x})\\
&- \tfrac{3}{4}\lambda^2R_{34}^2(u_{2x}v_2 - u_2v_{2x}) + \tfrac{1}{2}m_1\lambda R_{34}(u_{1x}v_2 - u_1v_{2x} + u_{2x}v_1 - u_2v_{1x}) - \tfrac{1}{4}m_1^2(u_{1x}v_1 - u_1v_{1x})\\
&- \tfrac{1}{2}m_1^2(u_{2x}v_2 - u_2v_{2x}) - \tfrac{1}{4}m_1^3(u_1u_2 + v_1v_2) + \tfrac{1}{8}\lambda^3R_{34}^3R_2 - \tfrac{1}{4}m_1\lambda^2R_{34}^2(u_1u_2 + v_1v_2) + \tfrac{1}{8}m_1^2\lambda R_{34}(R_1 + 2R_2)
\end{aligned}
\tag{2.11a}
$$

and

$$\widetilde{F}(Y^{+}_{-3}) = u_{1xx}v_{1x} - u_{1x}v_{1xx} + \lambda R_{34}(u_{1xx}u_1 + v_{1xx}v_1) + (\lambda/2)R_{34}(u_{1x}^2 + v_{1x}^2) - m_1(u_{1x}u_{2x} + v_{1x}v_{2x})$$
$$+ \tfrac{3}{4}\lambda^2 R^2_{34}(u_{1x}v_1 - u_1v_{1x}) + \tfrac{1}{2}m_1\lambda R_{34}(u_{1x}v_2 - u_1v_{2x} + u_{2x}v_1 - u_2v_{1x}) + \tfrac{1}{2}m_1^2(u_{1x}v_1 - u_1v_{1x})$$
$$+ \tfrac{1}{4}m_1^2(u_{2x}v_2 - u_2v_{2x}) - \tfrac{1}{4}m_1^3(u_1u_2 + v_1v_2) - \tfrac{1}{8}\lambda^3R^3_{34}R_1 - \tfrac{1}{4}m_1\lambda^2R^2_{34}(u_1u_2 + v_1v_2) - \tfrac{1}{8}m_1^2\lambda R_{34}(2R_1 + R_2). \qquad (2.11b)$$

Similar results are obtained for the Hamiltonians associated to the Lie–Bäcklund transformations Y_3^-, Y^{-}_{-3}. The vector fields Z_0^+, Z_0^- (see Ref. 1 and Appendix A) are Hamiltonian vector fields also, and the associated Hamiltonian densities are

$$\widetilde{F}(Z_0^+) = x(\widetilde{F}(Y_1^+) - \widetilde{F}(Y^{+}_{-1})) + t(\widetilde{F}(Y_1^+) + \widetilde{F}(Y^{+}_{-1})),$$
$$\widetilde{F}(Z_0^-) = x(\widetilde{F}(Y_1^-) - \widetilde{F}(Y^{-}_{-1})) + t(\widetilde{F}(Y^{-}_{-1}) + \widetilde{F}(Y^{-}_{-1})). \qquad (2.12)$$

Now we arrive at the remarkable fact that the creating and annihilating Lie–Bäcklund transformations Z_1^+, Z^{+}_{-1}, Z_1^-, Z^{-}_{-1}, (see Ref. 1 and Appendix A) turn out to be Hamiltonian vector fields. The corresponding Hamiltonian densities are

$$\widetilde{F}(Z_1^+) = x\{\widetilde{F}(Y_2^+) - \tfrac{1}{4}m_1^2\widetilde{F}(Y_0^+)\} + t\{\widetilde{F}(Y_2^+) + \tfrac{1}{4}m_1^2\widetilde{F}(Y_0^+)\},$$
$$\widetilde{F}(Z^{+}_{-1}) = x\{-\widetilde{F}(Y^{+}_{-2}) + \tfrac{1}{4}m_1^2\widetilde{F}(Y_0^+)\} + t\{\widetilde{F}(Y^{+}_{-2}) + \tfrac{1}{4}m_1^2\widetilde{F}(Y_0^+)\},$$
$$\widetilde{F}(Z_1^-) = x\{\widetilde{F}(Y_1^-) - \tfrac{1}{4}m_2^2\widetilde{F}(Y_0^-)\} + t\{\widetilde{F}(Y_2^-) + \tfrac{1}{4}m_2^2\widetilde{F}(Y_0^-)\},$$
$$\widetilde{F}(Z^{-}_{-1}) = x\{-\widetilde{F}(Y^{-}_{-2}) + \tfrac{1}{4}m_2^2\widetilde{F}(Y_0^-)\} + t\{\widetilde{F}(Y^{-}_{-2}) + \tfrac{1}{4}m_2^2\widetilde{F}(Y_0^-)\}. \qquad (2.13)$$

The Hamiltonians $F(Z_1^+),\ldots, F(Z^{-}_{-1})$ act as creating and annihilating operators on the t-independent Hamiltonians $F(Y^{+}_{-3}),\ldots, F(Y_3^+)$, $F(Y^{-}_{-3}),\ldots, F(Y_3^-)$ by the action of the Poisson bracket (1.6), for example,

$$\{F(Z_1^+), F(Y_0^+)\} = 0,$$
$$\{F(Z_1^+), F(Y^{+}_{-1})\} = \tfrac{1}{4}m_1^2\{\tfrac{1}{2}R_1 + \tfrac{1}{2}R_2\} = \tfrac{1}{4}m_1^2\, F(Y_0^+), \qquad (2.14)$$
$$\{F(Z_1^+), F(Y_1^+)\} = -F(Y_2^+),$$

and similar results for $F(Z^{+}_{-1})$, $F(Z_1^-)$, $F(Z^{-}_{-1})$. So the Hamiltonians $F(Z_1^+),\ldots,F(Z^{-}_{-1})$ generate four hierarchies of (probably commuting t-independent) Hamiltonians

$$F(Y^{\pm}_{\pm i}) \quad (i = 0,1,\ldots). \qquad (2.15)$$

Note that due to results described in Sec. III, we are more likely to consider

$$\ldots,F(Y^{+}_{-3}),\ldots,F(Y_0^+),\ldots,F(Y_3^+),\ldots \qquad (2.16a)$$

and

$$\ldots,F(Y^{-}_{-3}),\ldots,F(Y_0^-),\ldots,F(Y_3^-),\ldots \qquad (2.16b)$$

as two hierarchies instead of four.

III. INFINITE HIERARCHIES OF (x,t)-DEPENDENT LIE–BÄCKLUND TRANSFORMATIONS AND THEIR ASSOCIATED HAMILTONIANS

In this section we shall prove by construction the existence of infinite hierarchies of (x,t)-dependent Lie–Bäcklund transformations

$$Z_0^+, Z_1^+, Z_2^+, Z_3^+ = [Z_1^+, Z_2^+],\ldots, \quad Z_k^+ = [Z_1^+, Z_{k-1}^+],\ldots,$$
$$Z_0^+, Z^{+}_{-1}, Z^{+}_{-2}, Z^{+}_{-3} = [Z^{+}_{-1}, Z^{+}_{-2}],\ldots, \quad Z^{+}_{-k} = [Z^{+}_{-1}, Z^{+}_{-k+1}],\ldots. \qquad (3.1)$$

Since the Lie algebra of Lie–Bäcklund transformations is a direct sum of two Lie algebras,[1] we shall restrict our considerations from now on to the " + " part. First of all we construct the vector fields Z_2^+, Z^{+}_{-2} (cf. Table I). Second, we prove that $[Z_1^+, Z_2^+]$ is independent of Z_0^+, Z_1^+, Z_2^+, and by an induction argument we obtain an infinite hierarchy. The same arguments apply to the other hierarchies. Moreover we shall prove that the vector fields $Z^{+}_{\pm i}$ are Hamiltonian vector fields, and the associated Hamiltonian densities are given.

Motivated by the result of Z_0^+, Z_1^+, Z^{+}_{-1} (Ref. 1) we search for a local (x,t)-dependent Lie–Bäcklund transformation, linear in x,t and of degree 4. The structure of such a Lie–Bäcklund transformation has to be

$$x\Big(\sum_{i=-3}^{3}\alpha_i m_1^{3-|i|}Y_i^+\Big) + t\Big(\sum_{i=-3}^{3}\beta_i m_1^{3-|i|}Y_i^+\Big) + C, \qquad (3.2)$$

where, in (3.2), α_i, β_i $(i = -3,\ldots,3)$ are constants and C is (x,t) independent of degree 4. Eventually, after a huge computation, we obtained two Lie–Bäcklund transformations

$$Z_2^+ = x(Y_3^+ + \tfrac{1}{2}m_1^2Y_1^+) + t(Y_3^+ - \tfrac{1}{2}m_1^2Y_1^+) + C_2^+,$$
$$Z^{+}_{-2} = x(-Y^{+}_{-3} + \tfrac{1}{2}m_1^2Y^{+}_{-1}) + t(Y^{+}_{-3} + \tfrac{1}{2}m_1^2Y^{+}_{-1}) + C^{+}_{-2}, \qquad (3.3)$$

where, in (3.3),

TABLE I. The Lie-algebraic picture of the Federbush model.

	Y_3^+	Y_3^-		deg = 6
Z_2^+	Y_2^+	Y_2^-	Z_2^-	deg = 4
Z_1^+	Y_1^+	Y_1^-	Z_1^-	deg = 2
Z_0^+	Y_0^+	Y_0^-	Z_0^-	deg = 0
Z^{+}_{-1}	Y^{+}_{-1}	Y^{-}_{-1}	Z^{-}_{-1}	deg = 2
Z^{+}_{-2}	Y^{+}_{-2}	Y^{-}_{-2}	Z^{-}_{-2}	deg = 4
	Y^{+}_{-3}	Y^{-}_{-3}		deg = 6

$$C_2^{+,u_1} = \tfrac{1}{4}m_1(-2v_{2x} - \lambda R_{34}u_2 + m_1u_1),$$
$$C_2^{+,v_1} = \tfrac{1}{4}m_1(-2u_{2x} - \lambda R_{34}v_2 + m_1v_1),$$
$$C_2^{+,u_2} = \tfrac{3}{4}(-4u_{2xx} + 4\lambda R_{34}v_{2x} + 2\lambda(R_{34})_x v_2 - 2m_1v_{1x} + \lambda^2R_{34}^2u_2 - m_1\lambda R_{34}u_1 + m_1^2u_2), \tag{3.4a}$$
$$C_2^{+,v_2} = \tfrac{3}{4}(-4v_{2xx} - 4\lambda R_{34}u_{2x} - 2\lambda(R_{34})_x u_2 + 2m_1u_{1x} + \lambda^2R_{34}^2v_2 - m_1\lambda R_{34}v_1 + m_1^2v_2),$$
$$C_2^{+,u_3} = (\lambda/3)v_3L_2^+, \quad C_2^{+,v_3} = -(\lambda/2)u_3L_2^+,$$
$$C_2^{+,u_4} = (\lambda/2)v_4L_2^+, \quad C_2^{+,v_4} = -(\lambda/2)u_4L_2^+,$$
and
$$C_{-2}^{+,u_1} = \tfrac{3}{4}(-4u_{1xx} + 4\lambda R_{34}v_{1x} + 2\lambda(R_{34})_x v_1 + 2m_1v_{2x} + \lambda^2R_{34}^2u_1 + m_1\lambda R_{34}u_2 + m_1^2u_1),$$
$$C_{-2}^{+,v_1} = \tfrac{3}{4}(-4v_{1xx} - 4\lambda R_{34}u_{1x} - 2\lambda(R_{34})_x u_1 - 2m_1u_{2x} + \lambda^2R_{34}^2v_1 + m_1\lambda R_{34}v_2 + m_1^2v_1), \tag{3.4b}$$
$$C_{-2}^{+,u_2} = \tfrac{1}{4}m_1(2v_{1x} + \lambda R_{34}u_1 + m_1u_2),$$
$$C_{-2}^{+,v_2} = \tfrac{1}{4}m_1(-2u_{1x} + \lambda R_{34}v_1 + m_1v_2),$$
$$C_{-2}^{+,u_3} = (\lambda/2)v_3L_{-2}^+, \quad C_{-2}^{+,v_3} = -\frac{\lambda}{2}u_3L_{-2}^+,$$
$$C_{-2}^{+,u_4} = (\lambda/2)v_4L_{-2}^+, \quad C_{-2}^{+,v_4} = -(\lambda/2)u_4L_{-2}^+,$$
while
$$L_2^+ = 2u_{2x}u_2 + 2v_{2x}v_2 - m_1(u_1v_2 - u_2v_1),$$
$$L_{-2}^+ = 2u_{1x}u_1 + 2v_{1x}v_1 - m_1(u_1v_2 - u_2v_1). \tag{3.4c}$$
Remarkably, the vector fields Z_2^+, Z_{-2}^+ are again Hamiltonian vector fields, and the associated Hamiltonian densities are computed to be
$$F(Z_2^+) = x(\tilde{F}(Y_3^+) + \tfrac{1}{2}m_1^2\tilde{F}(Y_1^+)) + t(\tilde{F}(Y_3^+) - \tfrac{1}{2}m_1^2\tilde{F}(Y_1^+)) - (\lambda/2)R_{34}(u_2u_{2x} + v_2v_{2x}) + (\lambda/4)m_1R_{34}(u_1v_2 - u_2v_1) - \tfrac{1}{2}m_1(u_1u_{2x} + v_1v_{2x}) \tag{3.5a}$$
and
$$\tilde{F}(Z_{-2}^+) = x(-\tilde{F}(Y_{-3}^+) + \tfrac{1}{2}m_1^2\tilde{F}(Y_{-1}^+)) + t(\tilde{F}(Y_{-3}^+) + \tfrac{1}{2}m_1^2\tilde{F}(Y_{-1}^+)) - (\lambda/2)R_{34}(u_1u_{1x} + v_1v_{1x}) + (\lambda/4)m_1R_{34}(u_1v_2 - u_2v_1) - \tfrac{1}{2}m_1(u_1u_{2x} + v_1v_{2x}). \tag{3.5b}$$

Obviously, similar results will hold for vector fields Z_2^-, Z_{-2}^- and their associated Hamiltonian densities. A formal proof of the existence of infinite hierarchies of t-dependent Hamiltonians and corresponding Lie–Bäcklund transformations is given in Appendix B by application of Lemma 1.

Finally we computed the action of the vector fields Z_2^+ on the hierarchy $(Y_i^+)_{i\in\mathbb{Z}}$ by a calculation of the Poisson bracket of the associated Hamiltonians, which resulted in
$$\begin{aligned}
&\{F(Z_2^+), F(Y_{-2}^+)\} = -\tfrac{1}{4}m_1^4F(Y_0^+),\\
&\{F(Z_{-2}^+), F(Y_2^+)\} = -\tfrac{1}{4}m_1^4F(Y_0^+),\\
&\{F(Z_2^+), F(Y_{-1}^+)\} = -\tfrac{1}{2}m_1^2F(Y_1^+),\\
&\{F(Z_{-2}^+), F(Y_1^+)\} = -\tfrac{1}{2}m_1^2F(Y_{-1}^+),\\
&\{F(Z_2^+), F(Y_0^+)\} = 0, \quad \{F(Z_{-2}^+), F(Y_0^+)\} = 0,
\end{aligned} \tag{3.6}$$
while the action on the $F(Z_i^+)_{i\in\mathbb{Z}}$ hierarchy is
$$\begin{aligned}
&\{F(Z_2^+), F(Z_{-1}^+)\} = -\tfrac{3}{2}m_1^2F(Z_1^+),\\
&\{F(Z_{-2}^+), F(Z_{+1}^+)\} = -\tfrac{3}{2}m_1^2F(Z_{-1}^+),\\
&\{F(Z_2^+), F(Z_{-2}^+)\} = -m_1^4F(Z_0^+),
\end{aligned} \tag{3.7}$$
a result which is twice the action of $Z_{\pm1}^+$, being similar to the result obtained by Ten Eikelder[11] for the massive Thirring model.

IV. CONCLUSION

We obtained four infinite hierarchies of (x,t)-independent Lie–Bäcklund transformations and four infinite hierarchies of (x,t)-dependent Lie–Bäcklund transformations, which are all Hamiltonian vector fields. The corresponding densities are given.

ACKNOWLEDGMENTS

The authors wish to thank Professor R. Martini and Professor J. de Graaf for stimulating this joint research.

APPENDIX A: LIE–BÄCKLUND TRANSFORMATIONS OF THE FEDERBUSH MODEL

We summarize the Lie–Bäcklund transformations obtained in Ref. 1, only giving the " + " part, Y_0^+, Y_1^+, $Y_{\pm2}^+$, Z_0^+, $Z_{\pm+}^+$, i.e.,

$$Y_0^+ = -v_1\,\partial_{u_1} + u_1\,\partial_{v_1} - v_2\,\partial_{u_2} + u_2\,\partial_{v_2},$$
$$Y_1^+ = \tfrac{1}{2}m_1v_2\,\partial_{u_1} - \tfrac{1}{2}m_1u_2\,\partial_{v_1} + \tfrac{1}{2}(2u_{2x} + m_1v_1 - \lambda v_2(R_{34}))\partial_{u_2} + \tfrac{1}{2}(2v_{2x} - m_1u_1 + \lambda u_2(R_{34}))\partial_{v_2} - (\lambda/2)v_3R_2\,\partial_{u_3} + (\lambda/2)u_3R_2\,\partial_{v_3} - (\lambda/2)v_4R_2\,\partial_{u_4} + (\lambda/2)u_4R_2\,\partial_{v_4},$$
$$Y_{-1}^+ = \tfrac{1}{2}(2u_{1x} - m_1v_2 - \lambda v_1(R_{34}))\partial_{u_1} + \tfrac{1}{2}(2v_{1x} + m_1u_2 + \lambda u_1(R_{34}))\partial_{v_1} - \tfrac{1}{2}m_1v_1\,\partial_{u_2} + \tfrac{1}{2}m_1u_1\,\partial_{v_2} - (\lambda/2)v_3R_1\,\partial_{u_3} + (\lambda/2)u_3R_1\,\partial_{v_3} - (\lambda/2)v_4R_1\,\partial_{u_4} + (\lambda/2)u_4R_1\,\partial_{v_4},$$
$$Y_2^{+,u_1} = \tfrac{1}{4}m_1\{+2u_{2x} - \lambda v_2R_{34} + m_1v_1\}, \quad Y_2^{+,v_1} = \tfrac{1}{4}m_1\{+2v_{2x} + \lambda u_2R_{34} - m_1u_1\},$$
$$Y_2^{+,u_2} = \tfrac{1}{4}\{-4v_{2xx} - 2\lambda u_2(R_{34})_x - 4\lambda u_{2x}R_{34} + 2m_1u_{1x} - \lambda m_1v_1R_{34} + \lambda^2v_2R_{34}^2 + m_1^2v_2\},$$
$$Y_2^{+,v_2} = \tfrac{1}{4}\{+4u_{2xx} - 2\lambda v_2(R_{34})_x - 4\lambda v_{2x}R_{34} + 2m_1v_{1x} + \lambda m_1u_1R_{34} - \lambda^2u_2R_{34}^2 + m_1^2u_2\},$$
$$Y_2^{+,u_3} = (\lambda/2)v_3K_2^+, \quad Y_2^{+,v_3} = -(\lambda/2)u_3K_2^+, \quad Y_2^{+,u_4} = (\lambda/2)v_4K_2^+, \quad Y_2^{+,v_4} = -(\lambda/2)u_1K_2^+,$$
where

$$K_2^{+} = -2u_{1x}v_1 + 2u_1v_{1x} + m_1(u_1u_2 + v_1v_2) + \lambda R_1R_{34},$$

$$Y^{+;u_1}_{-2} = \tfrac{1}{4}\{-4v_{1xx} - 2\lambda u_1(R_{34})_x - 4\lambda u_{1x}R_{34} - 2m_1u_{2x} + \lambda m_1v_2R_{34} + \lambda^2 v_1R_{34}^2 + m_1^2v_1\},$$

$$Y^{+;v_1}_{-2} = \tfrac{1}{4}\{+4u_{1xx} - 2\lambda v_1(R_{34})_x - 4\lambda v_{1x}R_{34} - 2m_1v_{2x} - \lambda m_1u_2R_{34} - \lambda^2 u_1R_{34}^2 - m_1^2u_1\},$$

$$Y^{+;u_2}_{-2} = \tfrac{1}{4}m_1\{-2u_{1x} + \lambda v_1R_{34} + m_1v_2\}, \quad Y^{+;v_2}_{-2} = \tfrac{1}{4}m_1\{-2v_{1x} - \lambda u_1R_{34} - m_1u_2\},$$

$$Y^{+;u_3}_{-2} = (\lambda/2)v_3K^{+}_{-2}, \quad Y^{+;v_3}_{-2} = -(\lambda/2)u_3K^{+}_{-2}, \quad Y^{+;u_4}_{-2} = (\lambda/2)v_4K^{+}_{-2}, \quad Y^{+;v_4}_{-2} = -(\lambda/2)u_4K^{+}_{-2},$$

where

$$K^{+}_{-2} = -2u_{1x}v_1 + 2u_1v_{1x} + m_1(u_1u_2 + v_1v_2) + \lambda R_1R_{34},$$

while the (x,t)-dependent Lie–Bäcklund transformations are given by

$$Z_0^{+} = x(Y_1^{+} - Y^{+}_{-1}) + t(Y_1^{+} + Y^{+}_{-1}) + \tfrac{1}{2}(-u_1\,\partial_{u_1} - v_1\,\partial_{v_1} + u_2\,\partial_{u_2} + v_2\,\partial_{v_2}),$$

$$Z_1^{+} = x(+Y_2^{+} - \tfrac{1}{4}m_1^2Y_0^{+}) + t(+Y_2^{+} + \tfrac{1}{4}m_1^2Y_0^{+}) + \tfrac{1}{2}(-2v_{2x} + m_1u_1 - \lambda u_2R_{34})\partial_{u_2}$$
$$+ \tfrac{1}{2}(+2u_{2x} + m_1v_1 - \lambda v_2R_{34})\partial_{v_2},$$

$$Z^{+}_{-1} = x(-Y^{+}_{-2} - \tfrac{1}{4}m_1^2Y_0^{+}) + t(+Y^{+}_{-2} + \tfrac{1}{4}m_1^2Y_0^{+}) + \tfrac{1}{2}(+2v_{1x} + m_1u_2 + \lambda u_1R_{34})\partial_{u_1}$$
$$+ \tfrac{1}{2}(-2u_{1x} + m_1v_2 - \lambda v_1R_{34})\partial_{v_1},$$

$$Y_3^{+} = [Z_1^{+}, Y_2^{+}], \quad Y^{+}_{-3} = [Z^{+}_{-1}, Y^{+}_{-2}].$$

Similar results have been obtained for the " $-$ " part.[1]

APPENDIX B: THE INFINITY OF THE HIERARCHIES

We shall prove a lemma from which the existence of infinite hierarchies of Hamiltonians

$$\begin{aligned}
&F(Y_0^{+}), F(Y_1^{+}), F(Y_2^{+}), \ldots,\\
&F(Y_0^{+}), F(Y^{+}_{-1}), F(Y^{+}_{-2}), \ldots,\\
&F(Z_0^{+}), F(Z_1^{+}), F(Z_2^{+}), \ldots,\\
&F(Z_0^{+}), F(Z^{+}_{-1}), F(Z^{+}_{-2}), \ldots,
\end{aligned} \tag{B1}$$

and their associated Lie–Bäcklund transformations

$$Y_0^{+}, Y_1^{+}, Y^{+}_{\pm 2}, \ldots, \quad Z_0^{+}, Z^{+}_{\pm 1}, Z_2^{+}, \ldots, \tag{B2}$$

immediately follow. In this lemma the lower indices of u, v refer to partial derivatives with respect to x (i.e., $u_1 = u_x$, $u_2 = u_{xx}, \ldots$).

Lemma: Let $H_n(u,v)$, $K_n(u,v)$, $\overline{H}_n(u,v)$, and $\overline{K}_n(u,v)$ be defined by

$$\begin{aligned}
H_n(u,v) &= \int_{-\infty}^{\infty} (u_n^2 + v_n^2),\\
K_n(u,v) &= \int_{-\infty}^{\infty} (u_{n+1}v_n - v_{n+1}u_n),\\
\overline{H}_n(u,v) &= \int_{-\infty}^{\infty} x(u_n^2 + v_n^2),\\
\overline{K}(u,v) &= \int_{-\infty}^{\infty} x(u_{n+1}v_n - v_{n+1}u_n),
\end{aligned} \tag{B3}$$

and define the Poisson bracket of F and L $\{F,L\}$ by

$$\{F,L\} = \int_{-\infty}^{\infty} \left(+ \frac{\delta F}{\delta v}\frac{\delta L}{\delta u} - \frac{\delta F}{\delta u}\frac{\delta L}{\delta v} \right), \tag{B4}$$

then the following results hold

$$\{\overline{H}_1, H_n\} = +4nK_n, \tag{B5a}$$

$$\{\overline{H}_1, K_n\} = +2(2n+1)H_{n+1}, \tag{B5b}$$

$$\{\overline{H}_1, \overline{H}_n\} = +4(n-1)\overline{K}_n, \tag{B5c}$$

$$\{\overline{H}_1, \overline{K}_n\} = +2(2n-1)\overline{H}_{n+1}. \tag{B5d}$$

Proof: We shall prove relations (B5a) and (B5c) (the other proofs run along the same lines):

$$\frac{\delta H_n}{\delta u} = (-1)^n 2u_{2n}, \quad \frac{\delta H_n}{\delta v} = (-1)^n 2v_{2n}, \tag{B6a}$$

$$\frac{\delta \overline{H}_n}{\delta u} = (-1)^n 2(xu_n)^{(n)}, \quad \frac{\delta \overline{H}_n}{\delta v} = (-1)^{(n)} 2(xv_n)^{(n)}. \tag{B6b}$$

Substitution of (B6a) and (B6b) into (B4) yields

$$\begin{aligned}
\{\overline{H}_1, H_n\} &= -\int_{-\infty}^{\infty} 4(-1)^n u_{2n}(xv_1)^{(1)} - 4(-1)^n v_{2n}(xu_1)^{(1)}\\
&= 4(-1)^{2n}\int_{-\infty}^{\infty} (xv_1)^{(n)}u_{n+1} - (xu_1)^{(n)}v_{n+1}\\
&= +4n\int_{-\infty}^{\infty} v_nu_{n+1} - u_nv_{n+1} = +4nK_n,
\end{aligned}$$

which proves relation (B5a). Substitution of (B6b) into (2.4) yields

$$\begin{aligned}
\{\overline{H}_1, \overline{H}_n\} &= -\int_{-\infty}^{\infty} 4(-1)^n(xv_1)^{(1)}(xu_n)^{(n)}\\
&\quad - 4(-1)^n(xu_1)^{(1)}(xv_n)^{(n)}\\
&= 4(-1)^n(-1)^n\int_{-\infty}^{\infty} (xv_1)^{(n)}(xu_n)^{(1)}\\
&\quad - (xu_1)^{(n)}(xv_n)^{(1)}\\
&= +4(n-1)\int_{-\infty}^{\infty} x(u_{n+1}v_n - u_nv_{n+1})\\
&= +4(n-1)\overline{K}_n,
\end{aligned}$$

which proves relation (B5c). This existence of infinite hierarchies $H(Y^{+}_{\pm i})$ now follows from the explicit structure of $H(Z^{+}_{\pm 1})$ [Eq. (2.12)] and $H(Y^{+}_{\pm 1})$ [Eq. (2.6)] by considering the (λ, m_1, m_2)-independent parts and application of part a and b of this Lemma. The existence of the infinite hierarchies $H(Z^{\pm}_{\pm n})$ follows from a similar argument using $\overline{H}_m(u,v), \overline{K}_n(u,v)$.

[1]P. H. M. Kersten, J. Math. Phys. **27**, 1139 (1986).
[2]P. Federbush, Phys. Rev. **121**, 1247 (1961).
[3]S. N. M. Ruijsenaars, Commun. Math. Phys. **87**, 181 (1982).
[4]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[5]H. M. M. Ten Eikelder, "Symmetries for dynamical and Hamiltonian system," CWI Tract 17, Centre for Mathematics and Computer Science, Amsterdam, 1985.
[6]A. C. Hearn, *REDUCE User's Manual* (Version 3.0) (Rand Corp., Santa Monica, 1983).
[7]P. K. H. Gragert, Ph.D. thesis, Twente University of Technology, Enschede, The Netherlands, 1981.
[8]P. H. M. Kersten, Ph.D. thesis, Twente University of Technology, Enschede, The Netherlands, 1985.
[9]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, "Local jet bundle formulations of Bäcklund transformations," in *Mathematical Physics Studies*, Vol. 1 (Reidel, Boston, 1979).
[10]B. Fuchssteiner and A. S. Fokas, Phys. D **4**, (1981).
[11]H. M. M. Ten Eikelder, J. Math. Phys. **27**, 1404 (1986).

"Twist" for simply laced groups

Jae Kwan Kim, Jeong Yong Kim, and I. G. Koh
Physics Department, Korea Advanced Institute of Science and Technology, P. O. Box 150, Cheongryang, Seoul, Korea

(Received 2 January 1986; accepted for publication 2 April 1986)

"Twist," which is necessary to remove awkward signs in the commutator of vertex operators, is explicitly constructed with string momenta for simply laced groups. For different embeddings of roots into Euclidean spaces as well as for roots in nonorthogonal space, existence and construction of "twist" are shown in detail.

I. INTRODUCTION

Recently, the superstring theories[1] have received much attention for cancellation of gravitational and gauge anomalies[2] and for good quantum properties.[3] The fundamental object in string theories is the vertex operator in the emission of tachyon, Yang–Mills particles, and so on.

The vertex operators are known to satisfy the affine Kac–Moody algebra if the momenta in the vertex operators lie in the root spaces of simply laced groups (A_n, D_n, E_6, E_7, and E_8).[4–6] Such momenta can be realized if the string is compactified to an appropriate torus.

In construction of affine Kac–Moody algebras by using vertex operators, there appears an awkward sign in commutators. Frenkel and Kac[4] have shown that the introduction of "twist" for generators of affine Kac–Moody algebras resolves this sign problem.

Two different constructions of "twist" are available. One is to use the Dirac γ-algebra. Such "twist" has been constructed in Ref. 6. However, the γ-algebra is foreign to the bosonic string theories. The other is to use the center-of-mass momentum operator of string theories. Such "twists" have been constructed in Ref. 7 for the A_n, D_n, and E_8 groups.

In this paper, we study the general form and properties of "twist" using the center-of-mass momentum operator of string theories, and construct "twists" for all simply laced groups, including E_6 and E_7. For different embeddings of roots into Euclidean spaces as well as for roots in nonorthogonal space, the existence and the construction of "twist" are shown in detail.

II. A SHORT REVIEW OF THE VERTEX OPERATOR[1,6]

The vertex operator in string theories is

$$V(\mathbf{r},\theta) = :\exp[i\mathbf{r}\cdot\mathbf{Q}(\theta)]: = \exp[i\mathbf{r}\cdot\mathbf{Q}_<(\theta)]\exp[i\mathbf{r}\cdot(\mathbf{q}+\mathbf{p}\theta)] \times\exp[i\mathbf{r}\cdot\mathbf{Q}_>(\theta)], \tag{1}$$

where $\mathbf{r}$ and $\theta(=\tau+\sigma$ or $\tau-\sigma)$ are the momentum and the light-cone coordinate of a two-dimensional world sheet, respectively. The string coordinates $Q^\mu(\theta)$ $(\mu = 1,2,...,d)$ are expanded as

$$Q^\mu(\theta) = q^\mu + p^\mu\theta + i\sum_{n=1}^{\infty}\frac{1}{n}(\alpha_n^\mu e^{-in\theta} - \alpha_n^{\mu+}e^{in\theta}), \tag{2}$$

where α_n^μ, $\alpha_n^{\mu+}$, q^μ, and p^μ satisfy the following Hermitian condition and commutation relations:

$$\alpha_n^{\mu+} = \alpha_{-n}^\mu, \tag{3}$$

$$[\alpha_m^\mu,\alpha_n^\nu] = m\delta_{m+n,0}g^{\mu\nu}, \tag{4a}$$

$$[q^\mu,p^\nu] = ig^{\mu\nu}, \tag{4b}$$

with Euclidean metric $g^{\mu\nu}\sim(1,1,...,1)$. The $Q_>^\mu(\theta)$ and $Q_<^\mu(\theta)$ contain destruction and creation oscillators, respectively, as

$$Q_>^\mu(\theta) = i\sum_{n=1}^{\infty}\frac{1}{n}\alpha_n^\mu e^{-in\theta}, \tag{5a}$$

$$Q_<^\mu(\theta) = -i\sum_{n=1}^{\infty}\frac{1}{n}\alpha_{-n}^\mu e^{in\theta}. \tag{5b}$$

The momentum $\mathbf{r}$ in Eq. (1) is arbitrary for tachyonic emission vertex in the Veneziano model. However, when the d-dimensional string is compactified into a torus,[8] the allowed momenta are restricted. In this paper, we concentrate on compactifications where the allowed momenta satisfy

$$\mathbf{r}^2 = 2, \tag{6}$$

$$\mathbf{r}\cdot\mathbf{s} = -1, \quad \text{if } \mathbf{r}+\mathbf{s} \text{ is a root.} \tag{7}$$

We deal with cases where these momenta form a root space of a simply laced group.

The product of two vertex operators with momenta $\mathbf{r}$ and $\mathbf{s}$ is rewritten as a normal ordered product:

$$V(\mathbf{r},\theta)V(\mathbf{s},\theta') = [(e^{(i/2)(\theta'-\theta)})/(1-e^{i(\theta'-\theta)})]^{-\mathbf{r}\cdot\mathbf{s}} \times:\exp[i\{\mathbf{r}\cdot\mathbf{Q}(\theta)+\mathbf{s}\cdot\mathbf{Q}(\theta')\}]:. \tag{8}$$

The c-number coefficient in the right-hand side of Eq. (8) is computed by using

$$e^Ae^B = e^Be^Ae^{[A,B]}, \tag{9a}$$

$$e^Ae^B = e^{A+B}e^{(1/2)[A,B]}, \tag{9b}$$

where $[A,B]$ is a c-number. Note that product $V(\mathbf{s},\theta')V(\mathbf{r},\theta)$ of two vertex operators in reverse order is

$$V(\mathbf{s},\theta')V(\mathbf{r},\theta) = (-1)^{\mathbf{r}\cdot\mathbf{s}}V(\mathbf{r},\theta)V(\mathbf{s},\theta'). \tag{10}$$

The (anti)commutator of two vertex operators for distinct points θ and θ' in the world sheet should vanish. This motivates an additional sign $(-1)^{\mathbf{r}\cdot\mathbf{s}}$ in the definition of an (anti)commutator as

 0022-2488/86/082146-05$02.50

$$\{V(\mathbf{r},\theta),V(\mathbf{s},\theta)\}$$
$$= V(\mathbf{r},\theta)V(\mathbf{s},\theta') - (-1)^{\mathbf{r}\cdot\mathbf{s}}V(\mathbf{s},\theta')V(\mathbf{r},\theta). \quad (11)$$

The coefficient in Eq. (8) blows up for $\theta' = \theta$ for $\mathbf{r}\cdot\mathbf{s}<0$. To handle such singularities, we take limit $\theta' \to \theta$ for $\mathbf{r}\cdot\mathbf{s}<0$ in obtaining the (anti)commutator (11). The relevant (anti)commutators for simply laced groups are

$$\mathbf{r}\cdot\mathbf{s}\geqslant 0, \quad [V(\mathbf{r},\theta),V(\mathbf{s},\theta')\} = 0, \quad (12a)$$

$$\mathbf{r}\cdot\mathbf{s} = -1,$$
$$\{V(\mathbf{r},\theta),V(\mathbf{s},\theta')\} = 2\pi\delta(\theta-\theta')V(\mathbf{r}+\mathbf{s},\theta), \quad (12b)$$

$$\mathbf{r}\cdot\mathbf{s} = -2,$$
$$[V(\mathbf{r},\theta),V(\mathbf{s},\theta')] = 2\pi i\delta'(\theta-\theta') + 2\pi\delta(\theta-\theta')\mathbf{r}\cdot\mathbf{p}. \quad (12c)$$

The (anti)commutator in Eq. (11) is dependent on θ. To obtain a θ-independent algebra, one takes the Laurent expansion. It is more convenient to rewrite the vertex operator in $z = e^{i\theta}$ as

$$V(\mathbf{r},z) = \exp[i\mathbf{r}\cdot\mathbf{Q}_<(z)]\exp[i\mathbf{r}\cdot\mathbf{q} + \mathbf{r}\cdot\mathbf{p}\ln z] \times\exp[i\mathbf{r}\cdot\mathbf{Q}_>(z)], \quad (13)$$

which is singular at $z=0$. The coefficient in the Laurent expansion is

$$V_n(r) = \frac{1}{2\pi i}\oint_0 \frac{dz}{z} z^n V(\mathbf{r},z), \quad (14)$$

where the contour is taken around $z=0$. The $V_n(\mathbf{r})$ forms an affine Kac–Moody algebra with a $(-1)^{\mathbf{r}\cdot\mathbf{s}}$ factor as in Eq. (11) (see Refs. 4–6).

III. GENERAL PROPERTIES OF TWIST[4,8]

To obtain the commutator in standard form, one introduces "twist" $C(\mathbf{r})$ such that

$$[C(\mathbf{r})V(\mathbf{r},z),C(\mathbf{s})V(\mathbf{s},z')]$$
$$= \epsilon(\mathbf{r},\mathbf{s})C(\mathbf{r}+\mathbf{s})(V(\mathbf{r},z)V(\mathbf{s},z') - (-1)^{\mathbf{r}\cdot\mathbf{s}}V(\mathbf{s},z')V(\mathbf{r},z)), \quad (15)$$

where

$$C(\mathbf{r})V(\mathbf{r},z)C(\mathbf{s})V(\mathbf{s},z') = \epsilon(\mathbf{r},\mathbf{s})C(\mathbf{r}+\mathbf{s})V(\mathbf{r},z)V(\mathbf{s},z'). \quad (16)$$

From the associative law for Eq. (16), ϵ satisfies the following two-cocycle condition:

$$\epsilon(\mathbf{r},\mathbf{s})\epsilon(\mathbf{r}+\mathbf{s},\mathbf{t}) = \epsilon(\mathbf{r},\mathbf{s}+\mathbf{t})\epsilon(\mathbf{s},\mathbf{t}). \quad (17)$$

A "twist" $C(\mathbf{r})$ can be constructed by using the γ-matrix with the property[6]

$$C(\mathbf{r})C(\mathbf{s}) = (-1)^{\mathbf{r}\cdot\mathbf{s}+r^2s^2}C(\mathbf{s})C(\mathbf{r}). \quad (18)$$

The $C(\mathbf{r})$ formed the γ-matrix commutes with the vertex operator. However, the γ-matrix is not present in the bosonic string theories and should be introduced from outside. In this paper, we take an alternative method, where $C(\mathbf{r})$ does not commute with vertex operator. *A priori*, $C(\mathbf{r})$ can contain p, α_n, and α_{-n}. But the phase factor obtained from "twists" with α_n and α_{-n} will necessarily include a z-dependent term, which does not agree with a z-independent $(-1)^{\mathbf{r}\cdot\mathbf{s}}$. Thus $C(\mathbf{r})$ contains only the p-operator.

Equation (15) gives

$$\epsilon(\mathbf{s},\mathbf{r}) = (-1)^{\mathbf{r}\cdot\mathbf{s}}\epsilon(\mathbf{r},\mathbf{s}), \quad (19)$$

$$\epsilon(\mathbf{r},0) = \epsilon(0,\mathbf{r}) = C(0). \quad (20)$$

In our construction, we require that two-cocycle $\epsilon(\mathbf{r},\mathbf{s})$ be a c-number. This requirement gives the linear property to both "twist" and two-cocycle:

$$C(\mathbf{r}+\mathbf{s}) = C(\mathbf{r})C(\mathbf{s}), \quad (21)$$

$$\epsilon(\mathbf{r},\mathbf{s}+\mathbf{t}) = \epsilon(\mathbf{r},\mathbf{s})\epsilon(\mathbf{r},\mathbf{t}), \quad (22a)$$

$$\epsilon(\mathbf{r}+\mathbf{s},\mathbf{t}) = \epsilon(\mathbf{r},\mathbf{t})\epsilon(\mathbf{s},\mathbf{t}). \quad (22b)$$

We normalize "twist" and two-cocycle as

$$C(0) = 1, \quad (23)$$

$$\epsilon(\mathbf{r},0) = 1, \quad (24)$$

$$\epsilon(\mathbf{r},-\mathbf{r}) = 1. \quad (25)$$

From these properties, we obtain the following form for "twist" and two-cocycle:

$$C(\mathbf{r}) = \exp[i\pi(r^T\Lambda p)], \quad (26)$$

$$\epsilon(\mathbf{r},\mathbf{s}) = \exp[-i\pi(s^T\Lambda r)], \quad (27)$$

with antisymmetric matrix Λ. Then Eq. (19) becomes

$$\sum_{i,j} r_i s_j(\Lambda^{ji} - \Lambda^{ij}) = \mathbf{r}\cdot\mathbf{s}\ (\mathrm{mod}\ 2). \quad (28)$$

The final equation (28) gives an even-integer lattice condition for roots

$$\mathbf{r}^2\in 2Z, \quad (29)$$

$$\mathbf{r}\cdot\mathbf{s}\in Z. \quad (30)$$

This even-integer lattice condition is satisfied for simply laced groups with Eqs. (6) and (7). For convenience, we define a phase K as

$$K(\mathbf{r},\mathbf{s};\Lambda) = \sum_{i,j} r_i s_j(\Lambda^{ji} - \Lambda^{ij}), \quad (31)$$

with antisymmetric matrix Λ.

The root system of rank L can be embedded in N-dimensional Euclidean space ($N\geqslant L$). For construction of the phase $K(\mathbf{r},\mathbf{s};\Lambda)$ in Eq. (28), the relevant quantity is $(s_i r_j - r_i s_j)\Lambda^{ij}$. While one can take ${}_LC_2$ equations for ${}_LC_2$ independent choices of $\mathbf{r}$ and $\mathbf{s}$ in Eq. (28), there are ${}_NC_2$ independent matrix elements for antisymmetric Λ. Since N is greater than or equal to L, Λ always can be constructed regardless of particular embeddings of roots in N-dimensional Euclidean space.

IV. EXPLICIT CONSTRUCTION OF "TWISTS" FOR SIMPLY LACED GROUPS

The explicit form of $C(\mathbf{r})$ for a simply laced group depends on the representation of roots. First, let us consider an oblique system. As bases, one can take simple roots that are nonorthogonal. For simply laced groups, indices are raised or lowered with Cartan matrix A^{ij} or its inverse, respectively. The explicit solutions of Λ^{ij} for Eq. (28) are given as follows[9]:

for A_n,

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & -1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & -1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & -1 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \tag{32}$$

for D_n,

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix}, \tag{33}$$

for E_6,

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \tag{34}$$

for E_7,

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{35}$$

and for E_8,

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{36}$$

Second, let us deal with embedding of roots into Euclidean space. If all the coefficients in expansion of roots in Euclidean bases are integers, we can obtain easily a form of Λ by putting

$$\Lambda^{ij} - \Lambda^{ji} = 1 \pmod 2, \quad 1 \leqslant i \neq j \leqslant N. \tag{37}$$

The proof of this is

$$\begin{aligned} K(\mathbf{r},\mathbf{s};\Lambda) &= \sum_{i,j} r_i s_j (\Lambda^{ji} - \Lambda^{ij}) \\ &= \sum_{i \neq j} r_i s_j \pmod 2 \\ &= \Big(\sum_i r_i\Big)\Big(\sum_j s_j\Big) - \mathbf{r}\cdot\mathbf{s} \\ &= \Big(\sum_i r_i^2\Big)\Big(\sum_j s_j^2\Big) + \mathbf{r}\cdot\mathbf{s} \pmod 2 \\ &= \mathbf{r}\cdot\mathbf{s} \pmod 2, \end{aligned}$$

where $\mathbf{r}^2 = \mathbf{s}^2 = 2$ is used for the last equation.

For A_n and D_n, we can have the following representation of roots in Euclidean bases with integer coefficients:

$$A_n{:}\ \mathbf{r} = \mathbf{e}_i - \mathbf{e}_j, \quad 1 \leqslant i \neq j \leqslant n+1,$$

$$D_n{:}\ \mathbf{r} = \pm \mathbf{e}_i \pm \mathbf{e}_j, \quad 1 \leqslant i \neq j \leqslant n,$$

where $\mathbf{e}_i$ is an orthogonal unit vector. For this representation of roots, we choose a form of Λ as

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & -1 & -1 & \cdots \\ 1 & 0 & -1 & -1 & \cdots \\ 1 & 1 & 0 & -1 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{38}$$

Then the explicit form of twists $C(\mathbf{e}_i)$ for an unit vector $\mathbf{e}_i$ is

$$C(\mathbf{e}_i) = \exp[i(\pi/2)\{(p_1 + p_2 + \cdots + p_{i-1}) - (p_{i+1} + \cdots + p_N)\}]. \tag{39}$$

In the cases of E_6, E_7, and E_8, we cannot have integer coefficients for all roots in expansion by Euclidean bases. The following representations for simple roots are one of the convenient choices:

$$E_6{:}\ \mathbf{r} = \mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,\mathbf{r}_8, \tag{40a}$$

$$E_7{:}\ \mathbf{r} = \mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,\mathbf{r}_6,\mathbf{r}_8, \tag{40b}$$

$$E_8{:}\ \mathbf{r} = \mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,\mathbf{r}_6,\mathbf{r}_7,\mathbf{r}_8, \tag{40c}$$

where

$$\mathbf{r}_i = \mathbf{e}_i - \mathbf{e}_{i+1}, \quad i = 1,2,\ldots,7 \tag{41}$$

and

$$\mathbf{r}_8 = \tfrac{1}{2}(-\mathbf{e}_1 - \mathbf{e}_2 - \mathbf{e}_3 + \mathbf{e}_4 + \mathbf{e}_5 + \mathbf{e}_6 + \mathbf{e}_7 + \mathbf{e}_8). \tag{42}$$

The following form is one of the convenient choices of Λ satisfying Eq. (28) for E_6, E_7, and E_8:

$$\Lambda = \frac{1}{2}\begin{bmatrix} 0 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{43}$$

Another convenient representation of roots for E_6, E_7, and E_8 is given as

$$\mathbf{r}_8 = \tfrac{1}{3}(-2\mathbf{e}_1 - 2\mathbf{e}_2 - 2\mathbf{e}_3 + \mathbf{e}_4 + \mathbf{e}_5 + \mathbf{e}_6 + \mathbf{e}_7 + \mathbf{e}_8 + \mathbf{e}_9), \tag{44}$$

with $\mathbf{r}_i$ in the same form as in Eq. (41). For the choice (44),

$$\Lambda^{ij} - \Lambda^{ji} = 9 \pmod{18}, \quad 1 \leqslant i \neq j \leqslant 9, \tag{45}$$

satisfies Eq. (28). More generally if the coefficients of roots $\mathbf{r}$ and $\mathbf{s}$ in Euclidean bases are given as

$$r_i = N_i/n \tag{46a}$$

and

$$s_i = M_i/m, \tag{46b}$$

with N,M integers and n,m odd integers,

$$\Lambda^{ij} - \Lambda^{ji} = nm \pmod{2nm}, \quad i \neq j, \tag{47}$$

gives

$$\begin{aligned} K(\mathbf{r},\mathbf{s};\Lambda) &= \sum_{i,j} r_i s_j (\Lambda^{ji} - \Lambda^{ij}) \\ &= \sum_{i \neq j} N_i M_j \pmod 2 \\ &= \Big(\sum_i N_i\Big)\Big(\sum_j M_j\Big) - \mathbf{N}\cdot\mathbf{M} \\ &= \Big(\sum_i N_i^2\Big)\Big(\sum_j M_j^2\Big) + \mathbf{N}\cdot\mathbf{M} \pmod 2 \\ &= n^2\mathbf{r}^2 m^2 \mathbf{s}^2 + nm\mathbf{r}\cdot\mathbf{s} \pmod 2 \\ &= \mathbf{r}^2\mathbf{s}^2 + \mathbf{r}\cdot\mathbf{s} \pmod 2 \\ &= \mathbf{r}\cdot\mathbf{s} \pmod 2. \end{aligned}$$

This completes the proof for the solution (45).

ACKNOWLEDGMENTS

We are grateful to T. Eguchi, P. Goddard, A. Kent, and D. Olive for helpful introductions to these subjects.

This research is supported in part by the Research Institute for Basic Sciences, Hanyang University as a part of the basic science research institute program, Ministry of Education, 1985.

We would like to dedicate this paper to Professor Byung Ha Cho for his sixtieth birthday.

[1] J. H. Schwarz, Phys. Rev. **89**, 223 (1982), and references therein.
[2] M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984).
[3] M. B. Green and J. H. Schwarz, Phys. Lett. B **151**, 21 (1985).
[4] I. B. Frenkel and V. G. Kac, Inv. Math. **62**, 23 (1980).
[5] G. Segal, Commun. Math. Phys. **80**, 301 (1981).
[6] P. Goddard and D. Olive, in *Proceedings of the Conference on Vertex Operators in Mathematics and Physics*, edited by J. Lepowsky, S. Mandelstam, and I. M. Singer (Springer, Berlin, 1984).
[7] T. Eguchi and K. Higashijima, talk presented at the XIVth International Colloquim on Group Theoretical Methods in Physics, Seoul, 1985.
[8] D. G. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Nucl. Phys. B **256**, 253 (1985).
[9] We follow conventions given in R. Slansky, Phys. Rep. **79**, 1 (1981).

Non-Hamiltonian symmetries of a Boussinesq equation

H. M. M. Ten Eikelder
Department of Mathematics and Computing Science, Eindhoven University of Technology, P. O. Box 513, Eindhoven, The Netherlands

L. J. F. Broer
Department of Physics, Eindhoven University of Technology, P. O. Box 513, Eindhoven, The Netherlands

(Received 12 March 1986; accepted for publication 16 April 1986)

For a class of Hamiltonian systems there exist infinite series of non-Hamiltonian symmetries. Some properties of these series are illustrated using a Boussinesq equation. It is shown that the recursion operators generated by these non-Hamiltonian symmetries are powers of the original recursion operator. A class of recursion formulas for the constants of the motion (not for the corresponding symmetries!) is given.

I. INTRODUCTION

For a certain class of Hamiltonian systems there exist so-called recursion operators for symmetries. Repeated application of such a recursion operator yields a series of symmetries. Often it is possible to construct in this way infinite series of Hamiltonian symmetries (corresponding to constants of the motion) and infinite series of non-Hamiltonian symmetries. The most well-known example is the Korteweg–de Vries equation, where the Lénard operator generates an infinite series of Hamiltonian symmetries and an infinite series of non-Hamiltonian symmetries. In this paper we use a Boussinesq equation to illustrate some properties of these series, in particular the series of non-Hamiltonian symmetries. Similar results can be obtained for various other equations, see Ten Eikelder.[1,2] In this paper we work within the framework of differential geometry. For definitions of various concepts (symmetry, recursion operator for symmetries, etc.) see, for instance, Ref. 2, where also notations and conventions are given.

II. SYMMETRIES OF A BOUSSINESQ EQUATION

We study a Boussinesq equation of the form

$$\begin{aligned} v_t &= w_x, \\ w_t &= vv_x + \lambda v_{xxx}, \end{aligned} \qquad -\infty < x < \infty, \quad t > 0. \tag{1}$$

We consider (1) as an evolution equation in a topological vector space $\mathscr{W}$ of pairs of smooth functions (v,w), which decay, together with their x derivatives, sufficiently fast for $|x| \to \infty$. The spaces $\mathscr{W}$ and $\mathscr{W}^*$ are constructed such that their duality map $\langle \cdot,\cdot \rangle$ is the L_2 inner product. A possible choice is $\mathscr{W} = \mathscr{S}_p \times \mathscr{S}_p$ and $\mathscr{W}^* = \mathscr{U}_p \times \mathscr{U}_p$, where the function spaces $\mathscr{S}_p$ and $\mathscr{U}_p$ are described in Ref. 1. In terms of $u = (v,w) \in \mathscr{W}$ we can write (1) as

$$\dot u = X(u) \quad \left(\dot u(t) = \frac{d}{dt} u(t)\right).$$

A Hamiltonian form of (1) is well known. Let the function (functional) F_0 on $\mathscr{W}$ be given by

$$F_0 = \int_{-\infty}^{\infty} \left(\frac{1}{6} v^3 - \frac{1}{2} \lambda v_x^2 + \frac{1}{2} w^2\right) dx$$

and let the symplectic form Ω on $\mathscr{W}$ be (represented by the linear mapping Ω: $\mathscr{W} \to \mathscr{W}^*$) given by

$$\Omega = \begin{pmatrix} 0 & \partial^{-1} \\ \partial^{-1} & 0 \end{pmatrix}.$$

Then the vector field X can be written as $X = \Omega^{-1} dF_0$ (d = exterior derivative), so (1) is a Hamiltonian system.

The invariance of (1) for translations along the t and x axis and for a scale transformation yields the following elementary symmetries:

$$\begin{aligned} X_0 = X &= \begin{pmatrix} w_x \\ vv_x + \lambda v_{xxx} \end{pmatrix}, \\ Y_0 = \begin{pmatrix} v_x \\ w_x \end{pmatrix}, &\quad Z_0 = \begin{pmatrix} 2v + xv_x \\ 3w + xw_x \end{pmatrix} + 2tX_0. \end{aligned} \tag{2}$$

A recursion operator for symmetries of (1), written in terms of the "coordinates" of a modified Boussinesq equation, has been given by Fordy and Gibbons.[3] In terms of the "original coordinates" v and w this operator reads

$$\Lambda = \begin{pmatrix} \partial(2w\,\partial^{-1} + \partial^{-1}w) & \partial(v\,\partial^{-1} + \partial^{-1}v) + 8\lambda\,\partial^2 \\ 2v\,\partial v\,\partial^{-1} + 2\lambda\,\partial(\partial^2 v\,\partial^{-1} + \partial^{-1} v\,\partial^2) + 3\lambda\,\partial(\partial v + v\,\partial) + 8\lambda^2\partial^4 & \partial(w\,\partial^{-1} + 2\,\partial^{-1}w) \end{pmatrix}.$$

Three infinite series of symmetries now can be defined by

$$X_k = \Lambda^k X_0, \quad Y_k = \Lambda^k Y_0, \quad Z_k = \Lambda^k Z_0, \quad k = 0,1,2,\dots .$$

It is shown by Fokas and Anderson[4] that the Nijenhuis tensor of Λ vanishes (in their terminology, Λ is a hereditary symmetry). So for all vector fields A we have $\mathscr{L}_{\Lambda A}\Lambda = \Lambda \mathscr{L}_A \Lambda$ ($\mathscr{L}_A$ = Lie derivative in direction of A). This also can be verified by a straightforward computation.

Let A and B be vector fields on $\mathscr{W}$ such that $\mathscr{L}_A \Lambda = a\Lambda$ and $\mathscr{L}_B \Lambda = b\Lambda$ for $a,b \in \mathbb{R}$. Define $A_k = \Lambda^k A$ and $B_k = \Lambda^k B$, for $k = 0,1,2,\dots$. Using the fact that the Nijenhuis tensor of Λ vanishes, it is easily shown (see, for instance, Ref.

 0022-2488/86/082151-03$02.50

2) that the Lie bracket $[A_k, B_l]$ is given by

$$[A_k, B_l] = laB_{k+l} - kbA_{k+l} + \Lambda^{k+l}[A,B]. \quad (3)$$

A simple computation shows that

$$\mathscr{L}_{X_0}\Lambda = 0, \quad \mathscr{L}_{Y_0}\Lambda = 0, \quad \mathscr{L}_{Z_0}\Lambda = 3\Lambda, \quad (4)$$
$$[X_0,Y_0] = 0, \quad [Z_0,X_0] = 2X_0, \quad [Z_0,Y_0] = Y_0.$$

Substitution in (3) yields that the only nonvanishing Lie brackets between the elements of the series X_k, Y_k, and Z_k are given by

$$[Z_k, X_l] = (3l+2)X_{k+l}, \quad [Z_k,Y_l] = (3l+1)Y_{k+l},$$
$$[Z_k,Z_l] = 3(l-k)Z_{k+l}. \quad (5)$$

Since the Nijenhuis tensor of Λ vanishes, it immediately follows that

$$\mathscr{L}_{X_k}\Lambda = 0, \quad \mathscr{L}_{Y_k}\Lambda = 0, \quad \mathscr{L}_{Z_k}\Lambda = 3\Lambda^{k+1},$$
$$k = 0,1,2,\dots . \quad (6)$$

The first relation corresponds to the well-known fact that Λ is also a recursion operator for symmetries of the "higher-order Boussinesq equations" $\dot{u} = X_k$. The second relation shows that Λ is also a recursion operator for the equations $\dot{u} = Y_k$.

Next we discuss some properties of the series of symmetries Z_k. For every non-Hamiltonian symmetry Z a nonvanishing recursion operator for symmetries is given by $\Omega^{-1}\mathscr{L}_Z\Omega$. If Z is a Hamiltonian symmetry this expression yields 0 [because $\mathscr{L}_Z\Omega = d(\Omega Z) = 0$]. Note that the recursion operators obtained in this way are always the product of a canonical operator Ω^{-1} (also called Hamiltonian operator or implectic operator) and a closed operator $\mathscr{L}_Z(\Omega)$ (also called symplectic operator). Most interesting recursion operators have such a factorization, see, for instance, Magri,[5] Fuchssteiner and Fokas,[6] or Gel'fand and Dorfman.[7] In Ref. 2, we computed recursion operators for the massive Thirring model by this method.

The symmetries Z_0 and Z_1 turn out to be non-Hamiltonian. The corresponding recursion operators are found to be

$$\Omega^{-1}\mathscr{L}_{Z_0}\Omega = 3I \quad (I = \text{identity mapping: } \mathscr{W}\to\mathscr{W}),$$
$$\Omega^{-1}\mathscr{L}_{Z_1}\Omega = 6\Lambda. \quad (7)$$

So the recursion operator Λ can be reconstructed from the symmetry Z_1. From (6) and (7) it is easily shown by induction that

$$\mathscr{L}^k_{Z_1}\Omega = 3^k(k+1)!\Omega\Lambda^k, \quad k = 0,1,2,\dots . \quad (8)$$

Since the Lie derivatives and the exterior derivative commute, this relation yields a very simple proof of the well-known fact that all the two-forms $\Omega\Lambda^k$ are closed. This property implies that

$$L_{Z_k}\Omega = d(\Omega Z_k) = d(\Omega\Lambda^k Z_0) = \mathscr{L}_{Z_0}(\Omega\Lambda^k)$$
$$= (\mathscr{L}_{Z_0}\Omega)\Lambda^k + \Omega\mathscr{L}_{Z_0}(\Lambda^k) = (3k+3)\Omega\Lambda^k \neq 0,$$
$$k = 0,1,2,\dots . \quad (9)$$

Thus we have proved that all the symmetries Z_k are non-Hamiltonian and that the corresponding recursion operators are powers of Λ (up to a multiplicative constant).

Because X_0 is a Hamiltonian symmetry $\mathscr{L}_{X_0}\Omega = 0$. A simple computation shows that $Y_0 = \Omega^{-1} dG_0$ with $G_0 = \int_{-\infty}^{\infty} vw\, dx$, so Y_0 is also a Hamiltonian symmetry. A computation similar to (9) shows that all the symmetries X_k, Y_k $(k = 0,1,2,\dots)$ are Hamiltonian vector fields, i.e., there exist two series of constants of the motion F_k and G_k such that

$$X_k = \Omega^{-1} dF_k, \quad Y_k = \Omega^{-1} dG_k, \quad k = 0,1,2,\dots . \quad (10)$$

The corresponding symmetries commute, so all these constants of the motion are in involution. The existence of the series F_k is a standard property in this case, see, for instance, Ref. 6. It follows from (10) that

$$\Omega\Lambda^k X = dF_k,$$

which can be considered as "pre-Hamiltonian" forms for $X = X_0$. The original Hamiltonian form is obtained for $k=0$, while formally $k = -1$ with $F_{-1} = \int_{-\infty}^{\infty} \frac{1}{2}w\, dx$ yields the second Hamiltonian form of the Boussinesq equation.

We now give a class of recursion formulas for the constants of the motion F_k and G_k. The Hamiltonian vector field corresponding to the function $\mathscr{L}_{Z_l}F_k$ on $\mathscr{W}$ is

$$\begin{aligned}\Omega^{-1} d\mathscr{L}_{Z_l}F_k &= \Omega^{-1}\mathscr{L}_{Z_l} dF_k \\ &= \mathscr{L}_{Z_l}(\Omega^{-1} dF_k) - (\mathscr{L}_{Z_l}\Omega^{-1})dF_k \\ &= [Z_l,X_k] + \Omega^{-1}(\mathscr{L}_{Z_l}\Omega)\Omega^{-1} dF_k \\ &= (3k+2)X_{k+l} + (3l+3)\Lambda^l X_k \\ &= (3k+3l+5)\Omega^{-1} dF_{k+l},\end{aligned}$$

where we used (5) and (9). This yields the recursion formulas

$$F_{k+l} = \frac{1}{3k+3l+5}\mathscr{L}_{Z_l}F_k \equiv \frac{1}{3k+3l+5}\langle dF_k,Z_l\rangle. \quad (11)$$

In a similar way we get

$$G_{k+l} = \frac{1}{3k+3l+4}\mathscr{L}_{Z_l}G_k \equiv \frac{1}{3k+3l+4}\langle dG_k,Z_l\rangle. \quad (12)$$

Note that in these recursion formulas it is not necessary to reconstruct a functional from its derivatives. The part of Z_l with "coefficient" t is $2X_l$ [see (2)], so this term can be omitted in (11) and (12).

The symmetry Z_1 is given by

$$Z_1 = \begin{pmatrix} Z_{1,1} \\ Z_{1,2} \end{pmatrix} + 2tX_1,$$

where

$$\begin{aligned}Z_{1,1} &= 12vw + 2w_x\,\partial^{-1}v + 2v_x\,\partial^{-1}w \\ &\quad + 40\lambda w_{xx} + x(4(vw)_x + 8\lambda w_{xxx}), \\ Z_{1,2} &= 4v^3 + 2vv_x\,\partial^{-1}v + 2\lambda v_{xxx}\,\partial^{-1}v + 58\lambda vv_{xx} \\ &\quad + 45\lambda v_x^2 + 48\lambda^2 v_{xxxx} + 9w^2 + 2w_x\,\partial^{-1}w \\ &\quad + x(4v^2v_x + 12\lambda vv_{xxx} \\ &\quad + 24\lambda v_xv_{xx} + 8\lambda^2 v_{xxxxx} + 4ww_x).\end{aligned}$$

The part of Z_1 with coefficient x turns out to be Y_1. So $Z_1 = C_1 + xY_1 + 2tX_1$, where C_1 contains (also nonlocal) terms not depending explicitly on x and t. Similar relations turn out to hold for the other symmetries Z_k. For $l = 1$ we obtain from (11) and (12) the recursion formulas

$$F_{k+1} = \frac{1}{3k+8} \int_{-\infty}^{\infty} \left(\frac{\delta F_k}{\delta v} Z_{1,1} + \frac{\delta F_k}{\delta w} Z_{1,2} \right) dx, \qquad (13)$$

$$G_{k+1} = \frac{1}{3k+7} \int_{-\infty}^{\infty} \left(\frac{\delta G_k}{\delta v} Z_{1,1} + \frac{\delta G_k}{\delta w} Z_{1,2} \right) dx. \qquad (14)$$

Starting with F_0 and G_0 these relations enable us to generate the series F_k and G_k. In fact it is also possible to begin with F_{-1} and $G_{-1} = \int_{-\infty}^{\infty} v\, dx$.

A constant of the motion that depends explicitly on t is $J = \int_{-\infty}^{\infty} (xv + tw) dx$. Constants of the motion of this type always exist if a conserved density (in this case v) has a flux that is also conserved, see Broer and Backerra.[8] The Hamiltonian symmetry corresponding to J is formally given by

$$Z_{-1} = \Omega^{-1}\, dJ = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

It can be shown that (11) and (12) also hold for $l = -1$ and $k \geqslant 0$. This yields the relations

$$F_{k-1} = \frac{1}{3k+2} \mathscr{L}_{Z_{-1}} F_k \equiv \frac{1}{3k+2} \int_{-\infty}^{\infty} \frac{\delta F_k}{\delta w} dx,$$

$$G_{k-1} = \frac{1}{3k+1} \mathscr{L}_{Z_{-1}} G_k \equiv \frac{1}{3k+1} \int_{-\infty}^{\infty} \frac{\delta G_k}{\delta w} dx,$$

$$k = 0,1,2,\ldots.$$

While (13) and (14) allow us to go upwards in the series of constants of the motion, these two relations allow us to go downwards.

[1]H. M. M. Ten Eikelder, "Symmetries for dynamical and Hamiltonian systems," CWI tract 17 (Centre for Mathematics and Computer Science, Amsterdam, 1985).
[2]H. M. M. Ten Eikelder, J. Math. Phys. **27**, 1404 (1986).
[3]A. P. Fordy and J. Gibbons, J. Math. Phys. **22**, 1170 (1981).
[4]A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).
[5]F. Magri, "A geometrical approach to the nonlinear solvable equations," in *Nonlinear Evolution Equations and Dynamical Systems, Lecture Notes in Physics*, Vol. 120 (Springer, Berlin, 1980).
[6]B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[7]I. M. Gel'fand and I. Ya. Dorfman, Funct. Anal. Appl. **13**, 248 (1979); **14**, 223 (1980).
[8]L. J. F. Broer and S. C. M. Backerra, Appl. Sci. Res. **32**, 495 (1976).

On solution spaces of massless field equations with arbitrary spin

W. Heidenreich
Institut für Theoretische Physik, Technische Universität Clausthal, D-3392 Clausthal-Zellerfeld, Federal Republic of Germany

(Received 6 November 1985; accepted for publication 6 March 1986)

The solution spaces of massless field equations $\Box\Psi = 0$, with Ψ being a tensor, a (multi)spinor, or a Rarita–Schwinger field, are studied. They carry indecomposable representations of the Poincaré group whose invariant subspaces are determined. There exist indefinite invariant scalar products that allow Gupta–Bleuler quantization for all spins. For particular cases like the third-rank tensor with mixed symmetry and the Rarita–Schwinger field for spin-$\frac{3}{2}$, additional field equations are discussed, which project on subspaces.

I. INTRODUCTION

For massive particles it is straightforward to assign field equations to irreducible representations of the Poincaré group. For example, traceless symmetric tensors of rank s describe spin s particles. Yet the belief that fundamental physical theories are theories of massless particles gains ground. While high-energy physics deals with helicities $|\lambda| \leqslant 1$, theories of gravitation require at least $|\lambda| = 2$ in addition for the description of gravitons. Traditionally a symmetric two-tensor is used for this purpose. Clearly this is not the only possibility: in theories with torsion,[1] in locally Lorentz-invariant theories,[2] and in conformal helicity-2 theories[3] appear mixed tensors with rank 3; the Riemann and the Weyl tensors are used to describe gravitational waves.

In spite of many applications even the simple question of which Poincaré representations act on the solution space of massless field equations seems to be answered only for very few special cases. In electrodynamics, for example, the space of positive energy solutions of the field equations $\Box A_\mu = 0$ carries an indecomposable representation, which we label by $(0)\rightarrow(+1,-1)\rightarrow(0)$. It contains an irreducible invariant subspace of gauge modes with helicity $\lambda = 0$. The invariant subspace $(+1,-1)\rightarrow 0$ satisfies in addition the Lorentz condition $\partial^\mu A_\mu = 0$, the physical modes lie in the quotient space over the gauge modes. The "scalar" photons $\Phi \equiv \partial^\mu A_\mu$ form the upper helicity-0 states in the Gupta–Bleuler triplet.[4]

In this example, some of the complications of massless field equations can be seen. We have to deal with indecomposable representations in which the physical states lie in a quotient space with respect to the gauge modes. To define a free quantum field we need in addition "scalar" modes conjugate to the gauge modes. Wigner's classification of the unitary irreducible representations (UIR's) of the Poincaré group[5] is not sufficient to describe the solution space of gauge fields.

There are several classes of indecomposable representations of the Poincaré group discussed in the literature.[6] But they either do not describe the present situation (e.g., solutions of $\Box^2\Phi = 0$), or they treat the case of a vector and symmetric two-tensor only. Some properties of the representations considered here have been discussed by Barut and Raczka.[7]

To tackle the problem of classifying the solution spaces, we first show that it is equivalent to reducing the tensor product of the massless scalar representation with finite representations of the Lorentz group. In Sec. III we use Wigner's little group E(2) to discuss these tensor products in momentum space. To each finite irreducible Lorentz representation we get an indecomposable Poincaré representation, whose invariant subspaces and whose leak structure will be calculated. Using this result we obtain a group-theoretical classification of the solution spaces of $\Box\Psi = 0$ with Ψ being a (multi)spinor, a tensor, or a Rarita–Schwinger field.

In Sec. IV, part of these solution spaces will be supplied with an indefinite scalar product, which is necessary for Gupta–Bleuler quantization of the corresponding fields. In Sec. V, some physically interesting examples will be discussed more closely, specifically additional field equations that project on subspaces. The Appendix supplies some formulas for tensors that are used in Sec. V.

II. CONNECTION TO TENSOR PRODUCT OF INFINITE AND FINITE REPRESENTATION

The positive energy solutions of the scalar wave equation $\Box\Phi = 0$, normalizable with respect to the scalar product

$$i\int \Phi^* \overleftrightarrow{\partial}_0 \Phi \, d^3\vec{x}, \tag{1}$$

form a Hilbert space H_0. The Poincaré group acts as

$$\Phi'(x') = \Phi(x), \quad x'_\mu = \Lambda_\mu{}^\nu x_\nu + a_\mu \tag{2}$$

on the scalar field Φ, with infinitesimal generators

$$P_\mu = -i\,\partial_\mu, \tag{3}$$

$$M_{\mu\nu} = -i(x_\mu\,\partial_\nu - x_\nu\,\partial_\mu). \tag{4}$$

Here Φ carries the UIR with mass 0 and helicity 0, which we denote by (0). [The metric $(-+++)$ will be used.] In the sequel we will consider field equations

$$\Box\Psi(x) = 0, \tag{5}$$

where Ψ carries spinor or tensor indices (or both), i.e., it transforms as

$$\Psi'(x') = Q\Psi(x). \tag{6}$$

The finite representation Q of the Lorentz group acts on a vector space V_N. It is not unitarizable (except in the trivial case); in the case of a tensor it has an indefinite invariant

 0022-2488/86/082154-06$02.50

sesquilinear form, e.g., $A^*_\mu A^\mu$ for a vector. The positive energy solution space of $\Box\Psi=0$ is just the product space

$$H_N = H_0 \times V_N. \tag{7}$$

On this space acts the tensor product representation

$$(0)\otimes D_N \tag{8}$$

of the massless helicity 0 and some finite representation of the Poincaré group (translations act trivially in the finite representation). Due to this tensor-product structure we can deal with the spaces H_N easily.

Finite-dimensional representations of the Lorentz group: We label the finite-dimensional irreducible representations of the Lorentz group $SO(3.1)\sim Sl(2,C)$ as $D(j_1,j_2)$, where the $j_1,j_2=0,\frac{1}{2},1,...$ are the angular momenta of the complex extension $SU(2,C)\times SU(2,C)$. Explicitly we use as the infinitesimal generators of the fundamental representation $D(\frac{1}{2},0)$ the 2×2 matrices

$$S_{\mu\nu} = (4i)^{-1}(\sigma_\mu\bar\sigma_\nu - \sigma_\nu\bar\sigma_\mu), \tag{9}$$

with $\sigma_0 = -\bar\sigma_0 = 1$, and $\sigma_i = \bar\sigma_i$ the Pauli matrices. They act on a two-spinor Ψ_A. The conjugate fundamental representation $D(0,\frac{1}{2})$, whose generators $\bar S_{\mu\nu}$ are obtained by exchanging σ_ν and $\bar\sigma_\nu$, act on a dotted two-spinor $\Psi_{\dot A}$. The traceless multispinor $\Psi_{A_1\cdots A_n\dot A_1\cdots\dot A_m}$, symmetrized in the dotted and in the undotted indices, carries precisely the IR $D(n/2, m/2)$. With these multispinors all finite-dimensional IR's of $Sl(2,C)$ can be realized.

A vector A_ν carries $D(\frac{1}{2},\frac{1}{2})$, the defining representation of $SO(3.1)$; the traceless symmetric tensors $T_{\nu_1\cdots\nu_n}$ carry

$$D(n/2,n/2). \tag{10}$$

A general traceless tensor is obtained by acting with the Young symmetrizer, which has λ_1 boxes in the first row and λ_2 boxes in the second row on the tensor $T_{\nu_1,\ldots,\nu_n}$ with $n=\lambda_2+\lambda_2$ (a completely antisymmetric three-tensor, e.g., is equivalent to a vector). It carries, for $\lambda_2>0$,

$$D(\tfrac{1}{2}(\lambda_1+\lambda_2),\tfrac{1}{2}(\lambda_1-\lambda_2)) \oplus D(\tfrac{1}{2}(\lambda_1-\lambda_2),\tfrac{1}{2}(\lambda_1+\lambda_2)). \tag{11}$$

The constant Rarita–Schwinger field $\Psi_{A\nu_1\cdots\nu_n}$, traceless and symmetric in the tensor indices, carries

$$D(\tfrac{1}{2},0)\otimes D(n/2,n/2) = D(n/2+\tfrac{1}{2},n/2)\oplus D(n/2-\tfrac{1}{2},n/2). \tag{12}$$

III. SOLUTION SPACES

The aim of this section is to reduce the tensor product (8), resulting in an indecomposable representation. For this purpose we will define in momentum space subspaces that are not invariant under Poincaré transformations only because they are not invariant under the translations of the little group E(2). The leaks under action of these generators are used to read off the complete structure of invariant subspaces of the indecomposable representations, which are carried by the positive energy harmonic functions of various fields $\Psi(x)$.

A. Momentum space

To get into contact with Wigner's classification of the UIR's of the Poincaré group[5] we go to momentum space

$$\Psi(x) = (2\pi)^{-2}\int e^{ipx}\tilde\Psi(p)d^4p. \tag{13}$$

The scalar product (1) and the field equation guarantee that $\tilde\Psi(p)$ has support on the positive light cone $p^2=0$, $p^0>0$ only; we have the orbit $0^+_{m=0}$. The Poincaré group acts as

$$d(\Lambda,a)\tilde\Psi(p) = e^{-ipa}Q(\Lambda)\tilde\Psi(\Lambda^{-1}p), \tag{14}$$

where Q is the finite matrix representation from Eq. (6). We first reduce these representations for multispinors. The cases with $\tilde\Psi$ being a tensor or a spinor–tensor will be obtained easily using Eqs. (10)–(12).

B. Projection operators

We define projection operators

$$P_+ = -(\sigma p)/(2p_0), \quad P_- = +(\bar\sigma p)/(2p_0), \tag{15}$$

which can act on each spinor index A_i or $\dot A_i$. With their help we can decompose the vector space V of the multispinor field $\tilde\Psi_{A_1\cdots A_n\dot A_1\cdots\dot A_m}(p)$ into $2^n\cdot 2^m$ linearly independent subspaces $V(\pm,...,\pm;\dot\pm,...,\dot\pm)$, whose direct sum spans V. Here $V(+)$ is, for example, the space of the $P_+\tilde\Psi_A(p)$, $V(\dot -)$ is the space of the $P_-\tilde\Psi_{\dot A}(p)$. Explicit calculation shows that P_+ projects on an invariant subspace when acting on an index A, but not when acting on an index $\dot A$. For P_- the situation is reversed. $V(+,...,+,\dot -,...,\dot -)$ is the only irreducible invariant subspace under the Poincaré group.

For example in the case of a spinor field $\tilde\Psi_A$, $V(+)=\{P_+\tilde\Psi(p)\}=\{\tilde\Psi|P_-\tilde\Psi=0\}$ is invariant. In configuration space these are the solutions of the Weyl equation $(\bar\sigma\partial)\,\Psi_A(x)=0$, while the solution space of the equation $(\sigma\partial)\,\Psi_A(x)=0$ is not invariant. $[\sigma\,\partial\Psi_{\dot A}(x)=0$ is invariant.$]$ Next we want to show that the noninvariance of most spaces $V(\cdots)$ can be traced down to their noninvariance under the translations in the little group E(2).

C. Reduction to E(2)

Following Wigner, we decompose an arbitrary Lorentz transformation in three factors:

$$\Lambda = \alpha(p)\alpha(p)^{-1}\Lambda\alpha(\Lambda^{-1}p)(\alpha(\Lambda^{-1}p))^{-1}. \tag{16}$$

Choosing $\alpha(p)$ such that it maps q to p guarantees that the factor $\alpha(p)^{-1}\,\Lambda\alpha(\Lambda^{-1}p)$ belongs to the stability group E(2) of lightlike q. We take $q^\mu=(1,0,0,1)$ for definiteness. For the $\alpha(p)$ we use first a three-boost $B(p)$, which maps q to $(p,0,0,p)$ and then a rotation $R(p)$, which maps in $(p,\vec p)$. Rotations act identically on dotted and undotted spinors. Therefore $Q(R(p))$ leaves the subspaces $V(\cdots)$ invariant. The same is true for the boosts $Q(B(p))$: by construction we only apply them on functions of $(p,0,0,p)$. For these momenta the projection operators become $P_+=(1+\sigma_3)/2$, $P_-=(1-\sigma_3)/2$, which commute with the generator $S_{03}=(2i)^{-1}\sigma_3$ of the three-boost. So the subspaces $V(\cdots)$ remain invariant under the $Q(\alpha(p))$ defined above. It remains to investigate the action of the little group E(2).

D. Action of the stability group E(2)

The stability group of the lightlike momentum q is an E(2). Applied to the states $\tilde\Psi(q)$ it acts only on the indices.

All we have to do is to restrict the finite $\mathrm{Sl}(2,C)$ representations to the E(2) subgroup.

It consists of rotations around q with generator

$$\Lambda = (\Sigma + \dot{\Sigma})S_{12} = \tfrac{1}{2}(\Sigma + \dot{\Sigma})\sigma_3, \tag{17}$$

and of $E(2)$ translations with the generators

$$\begin{aligned} T_1 &= \Sigma(S_{01} + S_{31}) + \dot{\Sigma}(\bar{S}_{01} + \bar{S}_{31}) \\ &= \tfrac{1}{2}\Sigma(\sigma_2 - i\sigma_1) + \tfrac{1}{2}\dot{\Sigma}(\sigma_2 + i\sigma_1), \\ T_2 &= \Sigma(S_{02} + S_{32}) + \dot{\Sigma}(\bar{S}_{02} + \bar{S}_{32}) \\ &= -(i/2)\Sigma(\sigma_2 - i\sigma_1) + (i/2)\dot{\Sigma}(\sigma_2 + i\sigma_1). \end{aligned} \tag{18}$$

The sum Σ runs over all undotted indices and the sum $\dot{\Sigma}$ runs over all dotted indices. Now we consider a subspace $V(\cdots)$ with n_+ times "$+$," n_- times "$-$," $\dot{n}_+$ times "$\dot{+}$," and $\dot{n}_-$ times "$\dot{-}$." Its basis vectors at momentum q are, e.g.,

$$|+\rangle = |\dot{+}\rangle = (1,0) \quad \text{and} \quad |-\rangle = |\dot{-}\rangle = (0,1).$$

These basis states are eigenvectors of the E(2) rotation Λ with eigenvalues

$$\lambda = \tfrac{1}{2}(n_+ + \dot{n}_+ - n_- - \dot{n}_-). \tag{19}$$

The $E(2)$ translation $T_+ = \frac{1}{2}(iT_1 - T_2)$ maps components $(-)$ into $(+)$, the translation $T_- = -\frac{1}{2}(iT_1 + T_2)$ maps $(\dot{+})$-components into $(\dot{-})$. We have

$$\begin{aligned} &T_+ \colon V(n_+,n_-,\dot{n}_+,\dot{n}_-) \to V(n_+ + 1, n_- - 1, \dot{n}_+, \dot{n}_-), \\ &T_- \colon V(n_+,n_-,\dot{n}_+,\dot{n}_-) \to V(n_+, n_-, \dot{n}_+ - 1, \dot{n}_- + 1). \end{aligned} \tag{20}$$

So the E(2) translations leak between different subspaces $V(\cdots)$. They are responsible for having an indecomposable representation. We can now give the reduction of finite $\mathrm{Sl}(2,C)$ representations $D(j_1,j_2)$ on the E(2) subgroup as

$$\begin{tikzcd}[column sep=tiny]
 & & & & (j_2+j_1) \arrow[drrr, dashed] & & & & \\
 & (j_2-j_1+1) \arrow[urrr, dashed] \arrow[dr] & & & & & & (-j_2+j_1+1) \arrow[dr] & \\
(j_2-j_1) \arrow[ur] \arrow[dr] & & (j_2-j_1) & & & & (-j_2+j_1) \arrow[ur] \arrow[dr] & & (-j_2+j_1), \\
 & (j_2-j_1-1) \arrow[ur] \arrow[drrr, dashed] & & & & & & (-j_2+j_1-1) \arrow[ur] & \\
 & & & & (-j_2-j_1) \arrow[urrr, dashed] & & & &
\end{tikzcd} \tag{21}$$

where (λ) denotes a E(2) representation with $\Lambda(\lambda) = \lambda(\lambda)$.

It is clear that the action of T can change the helicity λ only by $|\Delta\lambda| = 1$. Keeping this information in mind we can simplify the notation and get, e.g., from the vector $D(\frac{1}{2},\frac{1}{2})$ the reduction $(0)\to(+1,-1)\to(0)$, from a symmetric two-tensor $D(1,1)$ the reduction

$$(0)\to(+1,-1)\to(+2,0,-2)\to(+1,-1)\to(0),$$

from a spinor $D(\frac{1}{2},0)$ we get $(-\frac{1}{2})\to(+\frac{1}{2})$, and from a Rarita–Schwinger $D(1,\frac{1}{2})$ we get

$$(-\tfrac{1}{2})\to(-\tfrac{3}{2},+\tfrac{1}{2})\to(-\tfrac{1}{2},+\tfrac{3}{2})\to(+\tfrac{1}{2}).$$

E. Solution spaces of $\Box\Psi = 0$

To each $V(\cdots)$ we can define an isomorphic invariant quotient space

$$\tilde{V}(\cdots) = \frac{V(\cdots) \oplus V'(\cdots)}{V'(\cdots)}, \tag{22}$$

where $V'(\cdots)$ is the vector space sum of all subspaces into which the $E(2)$ translations leak from $V(\cdots)$. On these spaces $\tilde{V}$ act irreducible unitarizable representations of the Poincaré group with helicity λ given by Eq. (19).

Now we are ready to state a main result of the present work: The reduction of the product representation $(0) \otimes D(j_1,j_2)$ of the Poincaré group is an indecomposable representation given by Eq. (21), where (λ) denotes the massless representation with helicity λ of the Poincaré group. According to Sec. II this representation is carried by the positive energy solution space of the field equation $\Box\Psi = 0$ with Ψ a multispinor with j_1 undotted indices and j_2 dotted indices. The positive energy solution spaces of harmonic tensors carry just the product representation $(0) \otimes D_N$, where D_N is given by Eq. (10) for symmetric tensors and Eq. (11) for all tensors. For Rarita–Schwinger fields with symmetric tensor indices, D_N is given by Eq. (12). Explicit examples will be discussed in Sec. V.

IV. INDEFINITE SCALAR PRODUCTS AND GUPTA–BLEULER QUANTIZATION

Indecomposable representations can be used for the description of physical particles in the framework of gauge theories in Gupta–Bleuler quantization. For this we need an invariant (indefinite) scalar product, whose zero norm states are gauge and "scalar" modes. The question of to what extent an indecomposable representation causes a scalar product (and the reverse) has been investigated by Araki.[8] In the case of tensors an invariant scalar product is simply

$$(T,T') = i\int T^*_{\nu_1\cdots\nu_n} \overleftrightarrow{\partial}_0 T'^{\nu_1\cdots\nu_n}\, d^3\vec{x}. \tag{23}$$

In the case of a constant spinor there is an invariant bilinear form $\Psi^T\sigma_2\Psi$, but no invariant sesquilinear form. Yet a sesquilinear form does exist on a constant Dirac spinor $\Psi = (\Psi_A, \Psi_{\dot{A}})$, namely

$$(\Psi,\Psi') = \Psi_A^H\Psi'_{\dot{A}} + \Psi_{\dot{A}}^H\Psi'_A. \tag{24}$$

So an invariant scalar product for a Dirac spinor field $\Psi(x)$ is

$$(\Psi(x),\Psi'(x)) = \int (\Psi_A^H \overleftrightarrow{\partial}_0 \Psi'_A + \Psi_A^H \overleftrightarrow{\partial}_0 \Psi'_A) d^3\vec{x} \quad (25)$$

(we are dealing here with harmonic functions, not with solutions of the Dirac equation, which projects on a subspace). These scalar products can be used for Gupta–Bleuler quantization of the corresponding fields. A closer investigation shows that the full positive energy solution space of harmonic functions contains physical modes (with positive norm) only in the case of symmetric tensors. In the other cases there are only norm-0 gauge and "scalar" modes. It is possible to find scalar products on invariant subspaces. If, for example, a spinor field satisfies the Weyl equation $\bar{\sigma}\,\partial\Psi_A = 0$, then an invariant scalar product is

$$(\Psi(x),\Psi'(x)) = \int \Psi^H \Psi' \, d^3\vec{x}. \quad (26)$$

For Rarita–Schwinger fields with $\bar{\sigma}\partial\Psi_{A\nu_1\cdots\nu_n}(x) = 0$, we have

$$(\Psi,\Psi') = \int \Psi^H_{\nu_1\cdots\nu_n} \Psi'^{\nu_1\cdots\nu_n} \, d^3\vec{x}. \quad (27)$$

On these subspaces we can formulate gauge theories with spin $\geqslant\frac{3}{2}$. No attempt has been made to discuss all Gupta–Bleuler quantum theories on invariant subspaces of the harmonic solutions.

In the next section some particular cases will be treated explicitly.

V. ADDITIONAL FIELD EQUATIONS

We can project on any invariant subspace of the indecomposable modules (21) using the projection operators of Sec. III. Yet the corresponding equations in configuration space, which are of the form $(\sigma\partial)\cdots(\bar{\sigma}\partial)\cdots\Psi = 0$, are in general of higher order. This approach does not seem to be very attractive. Instead I will discuss first- and second-order field equations for some particular examples, especially for spin-$\frac{3}{2}$ and -2.

A. Neutrino

The positive energy harmonic spinors $\Box\Psi_A = 0$ carry $(-\frac{1}{2})\to(\frac{1}{2})$. Imposing the Weyl equation $(\bar{\sigma}\partial)\Psi(x) = 0$ projects on the invariant irreducible subspace $(\frac{1}{2})$, the wave equation becomes superfluous as $(\sigma\partial)(\bar{\sigma}\partial) = \Box$. On this subspace we have the positive definite invariant scalar product (26).

B. Photon

The positive energy solution space of $\Box A_\mu = 0$ carries $(0)\to(+1,-1)\to(0)$. The scalar field $\Phi = \partial^\mu A_\mu$ carries the upper (0) (scalar modes), the Lorentz condition $\partial^\mu A_\mu = 0$ projects on $(+1,-1)\to(0)$. The pure gauge fields $A_\mu = \partial_\mu\Gamma$, $\Box\Gamma = 0$ carry the lower (0). The physical photons lie in the quotient space $\{\partial A = 0\}/\{A = \partial\Gamma\}$. Another way to describe photons irreducibly uses field strengths $F_{\mu\nu}$. The harmonic functions carry $((+1)\to(0)\to(-1)) \oplus ((-1)\to(0)\to(+1))$. Here $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $\partial^\mu F_{\mu\nu} = 0$ project on the lower $(-1)\oplus(+1)$, which describes physical photons. All these statements (besides being well known) can be proved easily by transforming to momentum space and applying the field equations on explicit states at $q^\nu = (1,0,0,1)$ (see the Appendix).

C. Rarita–Schwinger field for spin-$\frac{3}{2}$

The constant spinor–vector $\Psi_{A\mu}$ carries the finite SL$(2,C)$ representation $D(\frac{1}{2},\frac{1}{2}) \otimes D(\frac{1}{2},0) = D(1,\frac{1}{2}) \oplus D(0,\frac{1}{2})$. So the positive energy solution space of $\Box\Psi_{A\mu} = 0$ carries, according to Sec. III,

$$((-\tfrac{1}{2})\to(-\tfrac{3}{2},+\tfrac{1}{2})\to(-\tfrac{1}{2},+\tfrac{3}{2})\to(+\tfrac{1}{2})) \oplus ((+\tfrac{1}{2})\to(-\tfrac{1}{2})). \quad (28)$$

The Rarita–Schwinger field equation $(\bar{\sigma}\partial)\,\Psi = 0$ projects on the subspace

$$(+\tfrac{1}{2})\to(-\tfrac{1}{2},+\tfrac{3}{2})\to(+\tfrac{1}{2}). \quad (29)$$

[The upper $(+\frac{1}{2})$ is a superposition of the corresponding spaces in the two components of the representation (28).]

The indecomposable module (29) has an invariant scalar product

$$\int \Psi_\mu^H \Psi'^\mu \, d^3\vec{x}. \quad (30)$$

The $(+\frac{1}{2})$ modes have norm 0, the $(-\frac{1}{2},+\frac{3}{2})$ modes have positive norm. So we can perform a Gupta–Bleuler quantization with a spin-$\frac{3}{2}$ and a spin-$\frac{1}{2}$ particle in the physical sector. Here $\partial^\mu\Psi_\mu = 0$ projects on $(-\frac{1}{2},+\frac{3}{2})\to(+\frac{1}{2})$; it corresponds to the Lorentz condition. The spinor–trace condition $\bar{\sigma}^\mu\Psi_\mu = 0$ projects on $(+\frac{3}{2})\to(+\frac{1}{2})$, i.e., on the helicity $+\frac{3}{2}$ and the gauge modes. The latter are of the form $\Psi_{A\mu} = \partial_\mu\Phi_A$, $\bar{\sigma}\,\partial\Phi = 0$. The helicity $+\frac{3}{2}$ modes lie irreducibly in the quotient space

$$\{(\bar{\sigma}\Psi) = 0\}/\{\Psi = \partial\Phi\}. \quad (31)$$

D. The symmetric two-tensor

The constant traceless symmetric two-tensor tranforms as a $D(1,1)$ of SO(3.1). So the positive energy solution space of $\Box h_{\mu\nu} = 0$, $h^\mu{}_\mu = 0$ carries the Poincaré representation

$$(0)\to(+1,-1)\to(+2,0,-2)\to(+1,-1)\to(0). \quad (32)$$

The vector $B_\mu = \partial^\nu h_{\nu\mu}$ carries the upper

$$(0)\to(+1,-1)\to(0);$$

the Lorentz condition $\partial^\nu h_{\nu\mu} = 0$ projects on the subspace

$$(+2,-2)\to(+1,-1)\to(0). \quad (33)$$

The pure gauge field $h_{\mu\nu} = \partial_\mu\eta_\nu + \partial_\nu\eta_\mu$, $\partial^\mu\eta_\mu = 0$ carries the lower $(+1,-1)\to(0)$; the physical spin-2 modes lie in the quotient space

$$\{\partial^\nu h_{\nu\mu} = 0\}/\{h_{\mu\nu} = \partial_\mu\eta_\nu + \partial_\nu\eta_\mu\}. \quad (34)$$

The fourth-rank tensor

$$C_{\mu\nu\rho\sigma} = \partial_\rho\,\partial_\sigma h_{\mu\nu} - \partial_\rho\,\partial_\nu h_{\mu\sigma} - \partial_\mu\,\partial_\sigma h_{\nu\rho} + \partial_\mu\,\partial_\nu h_{\rho\sigma} \quad (35)$$

vanishes for pure gauge fields. If we impose the Lorentz condition $\partial^\nu h_{\mu\nu} = 0$ it carries the physical modes irreducibly. It is the field strength of the spin-2 field. In free linearized gravity, h is the traceless part of the metric, $\Box h$ the traceless part of the Ricci tensor, and C the Weyl tensor.

E. Mixed three-tensor

The three-tensor $\Psi_{\mu\nu\rho}$ with mixed symmetry $\Psi_{\nu\mu\rho} = -\Psi_{\mu\nu\rho}$, $\Psi_{\mu\nu\rho} + \Psi_{\rho\mu\nu} + \Psi_{\nu\rho\mu} = 0$, and trace $\Psi_{\mu\nu}{}^{\nu} = 0$ transforms as $D(\frac{3}{2},\frac{1}{2}) \oplus D(\frac{1}{2},\frac{3}{2})$ under SO(3.1). (Its Young diagram has $\lambda_1 = 2$ and $\lambda_2 = 1$.) The positive energy solution space of $\Box\Psi_{\mu\nu\rho} = 0$ carries

$$((+1)\rightarrow(+2,0)\rightarrow(+1,-1)\rightarrow(0,-2)\rightarrow(-1))$$
$$\oplus((-1)\rightarrow(0,-2)\rightarrow(+1,-1)\rightarrow(+2,0)\rightarrow(+1)). \tag{36}$$

So helicity $+2$ and -2 representations appear twice each. The upper spin-2 modes are also carried by the symmetrized divergency

$$H_{\mu\nu} \equiv \partial^{\rho}\Psi_{\rho\mu\nu} + \partial^{\rho}\Psi_{\rho\nu\mu}, \tag{37}$$

which carries the upper

$$(+1,-1)\rightarrow(+2,0,-2)\rightarrow(+1,-1)\rightarrow(0).$$

The lower spin-2 modes lie in the gradient field of a symmetric two-tensor,

$$\Psi_{\mu\nu\rho} = \partial_\mu h_{\rho\nu} - \partial_\nu h_{\rho\mu}, \quad \Box h_{\mu\nu} = 0, \tag{38}$$

which carries

$$(0)\rightarrow(+1,-1)\rightarrow(+2,0,-2)\rightarrow(+1,-1).$$

To get another characterization of the spin-2 modes we introduce a traceless tensor

$$C_{\mu\nu\rho\sigma} = \partial_\mu\Psi_{\rho\nu\sigma} + \partial_\nu\Psi_{\sigma\mu\rho} + \partial_\sigma\Psi_{\nu\rho\mu} + \partial_\rho\Psi_{\mu\sigma\nu}$$
$$+ \tfrac{1}{2}(\eta_{\mu\nu}H_{\rho\sigma} - \eta_{\mu\rho}H_{\sigma\nu} - \eta_{\sigma\nu}H_{\mu\rho} + \eta_{\sigma\rho}H_{\mu\nu}). \tag{39}$$

(Its Young diagram has $\lambda_1 = 2$, $\lambda_2 = 2$.) Here C carries $(+1,-1)\rightarrow(0)\rightarrow(+1,-1)\rightarrow(+2,0,-2)$.

If we impose for the antisymmetric divergence $F_{\mu\nu} \equiv \partial^{\rho}\Psi_{\rho\mu\nu} - \partial^{\rho}\Psi_{\rho\nu\mu} = 0$, then C carries the lower spin-2 modes only.

Next we want to describe the upper spin-2 modes. If we require

$$C_{\mu\nu\rho\sigma} = 0, \tag{40}$$

then $\Psi_{\mu\nu\rho}$ carries

$$(+2,0,-2)\rightarrow(+1,-1)\rightarrow(0,0)\rightarrow(+1,-1),$$

that is, the upper spin-2 modes and the pure gauge field of the form

$$\Psi_{\mu\nu\rho} = 2\,\partial_\rho f_{\mu\nu} + \partial_\nu f_{\mu\rho} - \partial_\mu f_{\nu\rho}, \quad f_{\nu\mu} = -f_{\mu\nu}. \tag{41}$$

The upper spin-2 modes only lie in the quotient space

$$\{C=0\}/\{\Psi = \text{pure gauge}\}. \tag{42}$$

In linear conformal gravity such an equation $C = 0$ projects on the physical (and part of the gauge) modes, i.e., on the conformal gravitons. Therefore the conformal gravitons are just the upper spin-2 modes in the solution space of $\Box\Psi_{\mu\nu\rho} = 0$.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Doebner for encourgement and him and my colleagues in Clausthal for stimulating discussions.

APPENDIX: ACTION OF E(2) ON TENSORS

In this appendix we reduce explicitly tensor representations of SO(3.1) to their E(2) subgroup. The explicit states were used in Sec. V to calculate the action of additional field equations.

The defining vector representation of the Lie algebra of SO(3.1) has a basis

$$(M_{\mu\nu})_{\rho\sigma} = -i(\eta_{\mu\rho}\eta_{\nu\sigma} - \eta_{\nu\rho}\eta_{\mu\sigma}),$$
$$\eta_{\mu\mu} = (-1,+1,+1,+1). \tag{A1}$$

The stability group E(2) of $q^\nu = (1,0,0,1)$ is generated by

$$\Lambda \equiv M_{12}, \quad T_i \equiv M_{3i} + M_{0i}, \quad i = 1,2. \tag{A2}$$

For the vector A^μ we define new basis states

$$|\pm\rangle = 1/2(0,\pm 1,i,0),$$
$$|s\rangle = 1/2(1,0,0,-1), \tag{A3}$$
$$|g\rangle = -1/2(1,0,0,1),$$

which are eigenvectors of Λ with

$$\Lambda|\pm\rangle = \pm|\pm\rangle, \quad \Lambda|s\rangle = 0, \quad \Lambda|g\rangle = 0. \tag{A4}$$

The action of E(2)-translations $T_\pm = 1/2(-T_2 \pm iT_1)$ on this basis is

$$T_\pm|s\rangle = |\pm\rangle, \quad T_\pm|\pm\rangle = 0, \quad T_\pm|\mp\rangle = |g\rangle. \tag{A5}$$

The reduction of the vector representation of SO(3.1) to the E(2) subgroup gives the indecomposable representation $(0)\rightarrow(+1,-1)\rightarrow(0)$, with the states

$$|s\rangle\rightarrow(|+\rangle,|-\rangle)\rightarrow|g\rangle. \tag{A6}$$

For tensors the generators are just a sum of matrices (A1), each acting on one of the indices. As for the vector we introduce a basis $|\mu_1,\ldots,\mu_n\rangle$, where μ_i can take the values $s,+,-,g$. The eigenvalues of the helicity operator are simply $\Lambda|\cdots\rangle = (n_+ - n_-)|\cdots\rangle$, where $n_\pm$ is the number of $(\pm)$ components. The action of $T_\pm$ is given by Eq. (A5) for each component.

So for the antisymmetric two-tensor we get

$$|s+\rangle\rightarrow(|-+\rangle + |sg\rangle)\rightarrow|-g\rangle, \tag{A7}$$

i.e., the E(2) representation $(+1)\rightarrow(0)\rightarrow(-1)$, and

$$|s-\rangle\rightarrow(|+-\rangle + |sg\rangle)\rightarrow|+g\rangle, \tag{A8}$$

i.e., $(-1)\rightarrow(0)\rightarrow(+1)$.

Similarly the states in the E(2) representation $(0)\rightarrow(+1,-1)\rightarrow(+2,0,-2)\rightarrow(+1,-1)\rightarrow(0)$ of the traceless symmetric two-tensor are

$$|ss\rangle\rightarrow(|s+\rangle,|s-\rangle)$$
$$\rightarrow(|++\rangle,|+-\rangle + 2|sg\rangle,|--\rangle)$$
$$\rightarrow(|g+\rangle,|g-\rangle)\rightarrow|gg\rangle. \tag{A9}$$

The trace-0 condition gives $|+-\rangle - |gs\rangle = 0$.

In the case of a three-tensor with mixed symmetry we first have to choose a basis of 20 independent components (out of the 64 of a general three-tensor). The trace-0 condition puts four linear combinations to zero. The action of the E(2) generators on the remaining 16 components can be calculated, giving an indecomposable representation $(+1)\rightarrow(+2,0)\rightarrow(+1,-1)\rightarrow(0,-2)\rightarrow(-1)$ and an inequivalent one with the signs of all helicities reversed. One possible choice of states is

$$\begin{aligned}|+ss\rangle &\to (|s++\rangle, |gss\rangle + |s-+\rangle - 2|s+-\rangle)\\ &\to (|-++\rangle + |sg+\rangle + |s+g\rangle,\\ &\quad |+--\rangle - 2|sg-\rangle + |s-g\rangle)\\ &\to (|g+-\rangle - 2|g-+\rangle + |sgg\rangle,\\ &\quad |g--\rangle) \to |-gg\rangle. \end{aligned} \tag{A10}$$

For the other one, $|+\rangle$ and $|-\rangle$ have to be exchanged. From this calculation it is obvious that the spinor formalism employed in Sec. III is far superior for obtaining the reduction of finite Sl$(2,C)$ representations to E(2).

In the case of a vector–spinor $\Psi_{\mu A}$ we introduce for the vector index the values $(s,+,-,g)$, for the spinor index the values $(+,-)$ of the projection (15). The action of E(2) on the latter is given by Eqs. (19) and (20). The explicit states in the reduction

$$D(\tfrac{3}{2},\tfrac{1}{2})|_{E(2)} = (-\tfrac{1}{2}) \to (-\tfrac{3}{2},\tfrac{1}{2}) \to (-\tfrac{1}{2},\tfrac{3}{2}) \to (\tfrac{1}{2})$$

are

$$\begin{aligned}|s-\rangle &\to (|--\rangle, |+-\rangle + |s+\rangle)\\ &\to (|g-\rangle + |-+\rangle, |++\rangle) \to |g+\rangle, \end{aligned} \tag{A11}$$

and in $D(0,\tfrac{1}{2})|_{E(2)} = (\tfrac{1}{2}) \to (-\tfrac{1}{2})$ they are

$$(|+-\rangle - |s+\rangle) \to (|g-\rangle - |-+\rangle). \tag{A12}$$

The wave functions $\tilde{T}\cdots(q)$ at the stability point $q^\nu = (1,0,0,1)$ of tensors in momentum space are precisely the states given above. Any invariant field equation in momentum space, which is satisfied by one of the states $\tilde{T}(q)$, is also satisfied by the full module that can be reached from this state. The results of Sec. V can be checked by straightforward calculation.

[1]F. W. Hehl, "On the kinematics of the torsion of space-time," Found. Phys. **15**, 451 (1985).

[2]R. Utiyama, "Invariant theoretical interpretation of interaction," Phys. Rev. **101**, 1597 (1956); T. W. B. Kibble, "Lorentz invariance and the gravitational field," J. Math. Phys. **2**, 212 (1961).

[3]B. Binegar, C. Fronsdal, and W. Heidenreich, "Linear conformal quantum gravity," Phys. Rev. D **10**, 2249 (1983).

[4]B. Binegar, C. Fronsdal, and W. Heidenreich, "Conformal QED," J. Math. Phys. **24**, 2828 (1983).

[5]E. Wigner, "On unitary representations of the inhomogeneous Lorentz group," Ann. Math. **40**, 149 (1939).

[6]G. Rideau, "Noncompletely reducible representations of the Poincaré-group associated with the generalized Lorentz gauge," J. Math. Phys. **19**, 1627 (1978); J. Bertrand, "Poincaré covariance and quantization of zero-mass fields," Nuovo Cimento **1**, 1 (1971); W. Heidenreich, "Group theory of the dipole ghost," J. Math. Phys. **25**, 376 (1984); B. Gruber, "Indecomposable spin-0 and spin-$\frac{1}{2}$ representations of the Lorentz algebra so(3.1)," Proc. R. Ir. Acad. Sec. A. **84**, 51 (1984).

[7]A. O. Barut and R. Raczka, "Properties of non-unitary zero mass induced representations of the Poincaré group on the space of tensor-valued functions," Ann. Inst. H. Poincaré **17**, 111 (1972).

[8]H. Araki, "Indecomposable representations with invariant inner product: A theory of the Gupta–Bleuler triplet," Commun. Math. Phys. **97**, 149 (1985).

Geometry of hyperbolic monopoles

C. Nash
Department of Mathematical Physics, St. Patrick's College, Maynooth, Ireland

(Received 2 January 1986; accepted for publication 2 April 1986)

The hyperbolic monopoles of Atiyah [M. F. Atiyah, Commun. Math. Phys. **93**, 471 (1984); "Magnetic monopoles in hyperbolic space," in *Proceedings of the International Colloquium on Vector Bundles* (Tata Institute, Bombay, 1984)] and Chakrabarti [A. Chakrabarti, J. Math. Phys. **27**, 340 (1986)] are introduced and their geometric properties and relations to instantons and ordinary monopoles clarified. A key tool is the use of the ball model of hyperbolic space to construct and examine solutions.

I. HYPERBOLIC MAGNETIC MONOPOLES

Monopoles and instantons are now of central importance to the study of Yang–Mills theories. It is well known that certain topological and geometrical aspects occupy a position of prominence. In this paper we deal with some of these aspects by considering what are called hyperbolic monopoles[1] (these, as we shall see below, are solutions of the Bogomolny equation in a hyperbolic space). One generates thereby immediate links connecting three kinds of classical solutions to field equations: instantons, ordinary monopoles, and hyperbolic monopoles. Furthermore, in most cases this link provides a method of constructing an ordinary monopole from knowledge of the hyperbolic monopole. In this section we describe the problem and give some of the details; Sec. II provides a solution, while the last two sections deal with the geometry and asymptotic properties.

To construct a hyperbolic monopole recall that an ordinary Bogomolny[2] monopole is a Yang–Mills instanton independent of the time coordinate x_0, i.e., it is an instanton invariant under time translations. If one replaces time translational invariance by some other space-time invariance one produces another kind of classical solution.

To be specific let us take the space-time transformation to be rotation about some axis through an angle θ. Now consider the flat line element in R^4:

$$ds^2 = dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2.$$

If we choose the rotations to be in the x_0-x_1 plane and let r, θ be polar coordinates in that plane we have

$$\begin{aligned} ds^2 &= dr^2 + r^2\, d\theta^2 + dx_2^2 + dx_3^2 \\ &= r^2\{d\theta^2 + (dr^2 + dx_2^2 + dx_3^2)/r^2\}. \end{aligned} \tag{1.1}$$

Now because of the conformal invariance in R^4 of the self-dual equations $F = {*F}$, a conformal change of metric leaves these equations invariant. Thus we can take as well

$$ds^2 = d\theta^2 + (dr^2 + dx_2^2 + dx_3^2)/r^2. \tag{1.2}$$

This is the line element for the space $S^1 \times H^3$, where H^3 is the three-dimensional hyperbolic space, $0 < r < \infty$, $-\infty < x_2 < \infty$, $-\infty < x_3 < \infty$, with metric determined by (1.2). It is important to note that the H^3 metric is singular at $r = 0$ and that $r = 0$ is both a plane R^2 in R^4 and the "axis" of the rotation in R^4. Therefore in passing from (1.1) to (1.2) we have made use of the conformal equivalence:

$$R^4 - R^2 \simeq S^1 \times H^3. \tag{1.3}$$

It is worth noting in passing that this conformal equivalence is in turn just a special case of the more general equivalence

$$R^n - R^m \simeq S^{n-m-1} \times H^{m+1}, \quad m < n. \tag{1.4}$$

To derive (1.4) we ask that a connection in R^n be invariant under rotations specified by l angles $l < n$. Then for the line elements ds^2 in R^n we can write ($r = \sqrt{x_1^2 + \cdots x_{l+1}^2}$, Ω_l represents the solid angle in R^n)

$$\begin{aligned} ds^2 &= dx_1^2 + \cdots + dx_{l+1}^2 + dx_{l+2}^2 + \cdots + dx_n^2 \\ &= dr^2 + r^2\, d\Omega_l^2 + dx_{l+2}^2 + \cdots + dx_n^2 \\ &= r^2\{d\Omega_l^2 + (dr^2 + dx_{l+2}^2 + \cdots dx_n^2)/r^2\}. \end{aligned} \tag{1.5}$$

Deleting the conformal factor r^2 in (1.5) amounts to the conformal correspondence $R^n - R^{n-l-1} \simeq S^l \times H^{n-l}$, from which we have (1.4). This correspondence would be relevant if one were to study SO(3)-symmetric instantons (cf., for example, Refs. 3 and 4). One would then choose $n = 4$, $m = 1$, for which we have

$$R^4 - R^1 \simeq S^2 \times H^2. \tag{1.6}$$

For the present we turn to the solution of the Bogomolny equation in the hyperbolic space H^3. The equation is

$$d_A \phi = -{*F}, \tag{1.7}$$

where d_A is the covariant derivative with respect to the connection A acting on the Higgs field ϕ, F is the curvature on H^3 of A, and the $*$ operation is with respect to the hyperbolic metric on H^3 implied by (1.2). We work throughout with gauge group SU(2). Equation (1.7) requires a boundary condition, which we take to be the usual one: $|\phi| = D$ at infinity, D a nonzero constant.[5] Now we can produce a solution to (1.7) either by producing an axially symmetric solution to $F = {*F}$ in R^4, or by solving (1.7) directly on H^3. We choose to do the latter.[6] To this end we make use of the fact that H^3 is a space of constant negative curvature with an alternative representation as the interior of a ball of radius 2, and with H^3 metric defined by

$$ds^2 = \frac{(dx^2 + dy^2 + dz^2)}{(1 - r^2/4)^2}, \quad r = \sqrt{x^2 + y^2 + z^2}. \tag{1.8}$$

The "ball" model of hyperbolic space is in fact isometric to the "upper half-plane" model and the transformation from the former to the latter is given by i, where

 0022-2488/86/082160-05$02.50

$$i=\alpha^{-1}\beta\alpha, \tag{1.9}$$

where α is stereographic projection from R^3 onto $S^3-\{\text{north pole}\}$ and β is a certain $\pi/2$ rotation of S^3 into itself. The precise definitions of α and β are as follows: Let $(x^1,...,x^4)$ be coordinates in $\mathbf{R}^4$. Represent the S^3 and $\mathbf{R}^3$ mentioned above by a sphere of *unit* radius and center $(1,0,0,0)$ and the hyperplane $x^1=0$, respectively. The sphere S^3 is then tangent to this $\mathbf{R}^3$ at its *south pole*, which is the origin of $\mathbf{R}^4$. The map α is then simply a stereographic projection from $\mathbf{R}^3$ onto S^3. To define β, let $\mathbf{x}\in\mathbf{R}^4$, $\mathbf{x}=(x^1,...,x^4)$. Then $\beta\mathbf{x}=(1+x^2,1-x^1,x^3,x^4)$; β is a $\pi/2$ rotation of S^3 about the two-plane $x^1=1, x^2=0$ that passes through its center. More concretely, α maps the ball of radius 2 in $\mathbf{R}^3$ onto the lower hemisphere of S^3, β rotates this lower hemisphere into the side hemisphere $x^1>0$ from which it is then mapped onto the upper half-plane by α^{-1}. For future reference one should also use (1.9) to show that, under i, the boundary $r=2$ of the ball is mapped onto the "axis" $r=0$ in R^4.

II. THE BALL PICTURE

From now on we work with the ball model of H^3. Next we make an ansatz for a solution $A=A^a_\mu(\sigma^a/2i)dx^\mu$ to (1.7): it is

$$A^a_i=[(P-1)/r^3]\epsilon_{iak}x^k,\quad A^a_\theta=\phi^a=Qx^a/r, \tag{2.1}$$

where P and Q are unknown functions depending on r only. In local coordinates we get

$$D_k\phi^a=(-1/2\sqrt{f})\epsilon_{ijk}F^a_{ij}, \tag{2.2}$$

with $f=(1-r^2/4)^{-2}$. This ansatz gives the pair of equations

$$\frac{dP}{dr}=\sqrt{f}PQ,\quad \sqrt{f}r^2\frac{dQ}{dr}=(P^2-1). \tag{2.3}$$

This system of equations can be reduced to the system $[T=\ln P, s=2\tanh^{-1}(r/2)]$

$$\frac{d^2T}{ds^2}=\sinh^{-2}s(e^{2T}-1),\quad Q=\frac{dT}{ds}. \tag{2.4}$$

We now can recognize this system as being the same as that found in Ref. 7, though we have arrived at them from a different starting point and purpose, and thus we have as general solution the expression

$$T=\ln\{(B\sinh s)/[\sinh(Bs+C)]\}, \tag{2.5}$$

with the constants B and C yet to be determined, and the boundary condition yet to be implemented. The constant C is zero in order to have a regular solution and the constant B is determined by the boundary condition $|\phi|=D$ at infinity. The value of B is then related to that of D by the condition

$$D=B-1. \tag{2.6}$$

The positivity of D requires $B>1$; apart from this restriction, D and B can take a continuum of values. Nevertheless, to retain the dual interpretation of our solution, as both a hyperbolic monopole and an axially symmetric instanton, we shall find that D and B are restricted to a discrete set of values. To establish this, let A be an instanton on a principal SU(2)-bundle P and let $e^{i\theta}\in\mathrm{U}(1)$. Then U(1) acts on A leaving it invariant if A is axially symmetric, but U(1) also acts on any other bundle V, say, with the same base R^4 and transition functions but a different fiber. Choose V to be the associated vector bundle with fiber $\mathbf{C}^2$. Then the action of $e^{i\theta}$ produced by rotation through an angle θ in the x_0-x_1 plane also acts on V. However, if we restrict V to that part of R^4 that is left fixed by the rotation, then U(1) just acts on the fibers of this restricted bundle V', say. Here U(1) acts as a set of unitary matrices M_θ and hence we have a two-dimensional unitary representation of U(1). So we write, relative to a suitable basis,

$$M_\theta=\begin{bmatrix} e^{in\theta} & 0\\ 0 & e^{-in\theta}\end{bmatrix}, \tag{2.7}$$

for some (positive) integer n. Thus axially symmetric instantons now have two integer invariants: n and $c_2(P)$ the second Chern class of P. It is also clear from (2.7) that the bundle V' is a sum of two line bundles, in fact

$$V'=L\oplus L^{-1}, \tag{2.8}$$

where L is a line bundle over the axis R^2. This line bundle is completely characterized by $c_1(L)$, its first Chern class. These three integers $c_1(L)$, $c_2(P)$, and n are not independent and are related[8] according to

$$c_2(P)=2nc_1(L). \tag{2.9}$$

Note that this forces $c_2(P)$ to be a point to which we shall return. The integers n and $c_1(L)$ correspond, in the hyperbolic monopole picture, to $|\phi|_{\rm as}$, the asymptotic norm of the Higgs field, and k, the magnetic charge, respectively. To establish this consider first n and $|\phi|_{\rm as}$ and proceed as follows: Take a vector Y and consider the Lie derivative L_XY and the covariant derivative ∇_XY. The difference of the two derivatives for arbitrary Y is a measure of the X-component of the connection A. Next take an S^1 orbit so that the Lie derivative corresponds to $X=\partial/\partial\theta$ and the X-component of A corresponds to the Higgs field ϕ. Now we are interested in $|\phi|_{\rm as}$ and this corresponds in the hyperbolic picture to evaluating $|\phi|$ on the sphere S^2, which is the boundary ∂H^3 of the hyperbolic space H^3, i.e., $|\phi|_{\rm as}=|\phi|_{\partial H^3}$. Further, as we saw above, this $S^2\simeq\partial H^3$ corresponds to the axis of rotation under the transformation i of (1.9). Thus to calculate $|\phi|_{\rm as}$, we restrict the difference $L_XY-\nabla_XY$ to the axis. But since the axis is a fixed point of the group action, the parallel transport leaves the fibers fixed and ∇_XY vanishes. So we select a point p on the axis—any p will do since the axis is fixed under the group action. We know how the group acts on Y and therefore we have the equality [here $\phi_{\rm as}=\phi(p)$ just denotes the Higgs field evaluated on $p\in\partial H^3$]

$$[\phi_{\rm as},Y]=\frac{d}{d\theta}(e^{\theta\phi_{\rm as}}Ye^{-\theta\phi_{\rm as}})\Big|_{\theta=0}$$

$$=\frac{d}{d\theta}(M_\theta YM_\theta^{-1})\Big|_{\theta=0}. \tag{2.10}$$

Thus we can identify M_θ as $e^{\theta\phi_{\rm as}}$. Now we use the identities

$$\phi^{2n}=(-\tfrac{1}{4})^n|\phi|^{2n}I,$$
$$\phi^{2n+1}=(-\tfrac{1}{4})^n|\phi|^{2n}\phi.$$

These give

$$e^{\theta\phi_{as}} = \cos\left(\frac{\theta\,|\phi|_{as}}{2}\right)I + \sin\left(\frac{\theta\,|\phi|_{as}}{2}\right)\frac{2\phi_{as}}{|\phi|_{as}}.$$

But we also have

$$M_\theta = \begin{bmatrix} e^{in\theta} & 0 \\ 0 & e^{-in\theta} \end{bmatrix} = \cos n\theta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sin n\theta \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = e^{n\theta\sigma}, \quad \sigma = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Hence $|\phi|_{as} = 2n$ and we have the relation between $|\phi|_{as}$ and n. [The matrix σ defines the su(2)-direction in which ϕ_{as} points, it looks like σ_3 but will not in general coincide with it because we have loosely assumed that the su(2)-basis in which M_θ is diagonal coincides with that used for ϕ. Actually, if T is the matrix of the change of basis in the underlying space C^2, then $T\sigma T^{-1}$ is the true direction of ϕ_{as}.]

The connection between $c_1(L)$ and the magnetic charge k is obtained by recalling that, for a solution to the Bogomolny equation, the energy $E = 4\pi|\phi|_{as}k$. Also, integrating the action S over θ gives $S = 2\pi E$. But $S = 8\pi^2 c_2$ gives $c_2 = E/4\pi$. These facts, together with Eq. (2.8), give $c_2 = |\phi|_{as}k$, so that $c_1(L) = k$.

We are now in a position to calculate the quantities k, c_2, and $|\phi|_{as}$ for our solution—thereby producing an independent check of the relations between them.

The magnetic charge k is given by (α and β denote angular coordinates on ∂H^3)

$$k = -\frac{1}{2\pi}\int_{\partial H^3} \mathrm{tr}\{\hat\phi\, d\hat\phi \wedge d\hat\phi\}, \quad \hat\phi = \frac{\phi}{|\phi|},$$

$$= \frac{1}{16\pi i}\int_{\partial H^3} \mathrm{tr}(n\, dn \wedge dn), \tag{2.11}$$

$$n = \sin\alpha\cos\beta\sigma_x + \sin\alpha\sin\beta\sigma_y + \cos\alpha\sigma_z,$$

and this integral is easily verified to be unity.

Next we calculate $c_2 = E/4\pi$, but

$$E = -\int_{H^3} \mathrm{tr}\{d_A\phi \wedge {*d_A\phi} + F \wedge {*F}\}$$

$$= -2\int_{H^3} \mathrm{tr}\{d_A\phi \wedge {*d_A\phi}\} = -2\int_{H^3} d\,\mathrm{tr}(\phi^* d\phi)$$

$$= -2\int_{\partial H^3} \mathrm{tr}(\phi^* d\phi) = 4\pi(B-1), \tag{2.12}$$

where we have used the Bogomolny equation, the Bianchi identities, and various properties of covariant derivatives. We obtain, therefore, $c_2 = (B-1)$.

To calculate $|\phi|_{as}$ is to evaluate $|\phi|$ on ∂H^3, where $r \to 2$ or equivalently $s \to \infty$. Explicit calculation yields

$$\lim_{s\to\infty} |\phi| = B - 1 = |\phi|_{as}. \tag{2.13}$$

We see that we do indeed have $c^2 = |\phi|_{as}k$, consistent with $c_2 = 2nk$ and $|\phi|_{as} = 2n$.

III. CONNECTION WITH INSTANTONS AND R^3-MONOPOLES

It is useful here to pursue further the role of the integer invariant n in the construction of axially symmetric instantons. Take the $c_2 = 1$ instanton A, which results if we choose $B = 2$. Recall that then A is given by the simple expression[9]

$$A = \frac{x^2}{x^2 + \lambda^2} g^{-1}\partial_\mu g\, dx^\mu, \quad g(x) = \frac{x_0 \mathrm{I} + i\sigma_j x_j}{\sqrt{x^2}}. \tag{3.1}$$

This form may seem to be independent of the angle θ. However, it is not. This is because g is not invariant under rotation through 2π. To understand this we note that $n = \frac{1}{2}$ and that A, being the one-instanton, is a connection on P, where P is half of the spin bundle corresponding to one of the SU(2)-factors in the decomposition

$$\mathrm{Spin}(4) = \mathrm{Spin}(3) \times \mathrm{Spin}(3) = \mathrm{SU}(2) \times \mathrm{SU}(2).$$

More concretely we can regard $g(x)$ restricted to the equator S^3 of S^4 as being the transition function of the bundle P. Then we can write the x_μ in terms of the Euler angles θ', ϕ', ψ' according to

$$\begin{aligned} x_0 &= \sqrt{x^2}\sin\left(\frac{\theta'}{2}\right)\sin\left(\frac{\psi'-\phi'}{2}\right), \\ x_3 &= \sqrt{x^2}\sin\left(\frac{\theta'}{2}\right)\cos\left(\frac{\psi'-\phi'}{2}\right), \\ x_2 &= \sqrt{x^2}\cos\left(\frac{\theta'}{2}\right)\sin\left(\frac{\psi'+\phi'}{2}\right), \\ x_1 &= \sqrt{x^2}\cos\left(\frac{\theta'}{2}\right)\cos\left(\frac{\psi'+\phi'}{2}\right). \end{aligned} \tag{3.2}$$

The factor of $\frac{1}{2}$ in the arguments of the trigonometric functions produces a change in sign of the transition function under a 2π rotation that will account for the noninvariance of A. The fact that $n = \frac{1}{2}$ is because, on restricting the base space to a circle S^1, SU(2) becomes the double cover of S^1 and can therefore have a half-integral weight n. Similar remarks apply to $n = \frac{3}{2}, \frac{5}{2}$, etc. This is why we found above that c_2 was even.

A further matter of geometrical interest is the role played by the curvature of the hyperbolic space H^3. We can import a parameter R into the metric ds^2 if we write

$$ds^2 = \frac{r^2}{R^2}\left\{R^2 d\theta^2 + \frac{R^2}{r^2}(dr^2 + dx_1^2 + dx_2^2)\right\}.$$

When we delete the conformal factor this corresponds to working on $S^1(R) \times H^3(R)$, where $S^1(R)$ is a circle of radius R, and $H^3(R)$ is a hyperbolic space of scalar curvature $-6/R^2$. The "ball model" for $H^3(R)$ is now the interior of a ball of radius $2R$. On $S^1(R) \times H^3(R)$ the Bogomolny equation is

$$D_k\phi^a = -(R/2\sqrt{f_R})\epsilon_{ijk}F^a_{jk}, \tag{3.3}$$

where $f_R = (1 - r^2/4R^2)^{-2}$. With a similar ansatz to the previous one we can again exactly solve the field equations. For the Higgs field ϕ we find

$$\phi^a = \frac{1}{R}\frac{dT}{ds}\frac{x^a}{r}, \tag{3.4}$$

with T as before but with $s = 2\tanh^{-1}(r/2R)$. To evaluate $|\phi|_{\rm as}$ we let $r \to 2R$ and we obtain

$$|\phi|_{\rm as} = (B-1)/R = 2n/R. \tag{3.5}$$

We see that by choosing r appropriately we can satisfy any boundary condition for $|\phi|_{\rm as}$. In particular if we choose $R = 2n$ we have $|\phi|_{\rm as} = 1$. If we then let $R \to \infty$, with $|\phi|_{\rm as}$ fixed, H^3 becomes the flat space $\mathbf{R}^3$ and our solution becomes an ordinary $\mathbf{R}^3$-monopole. We shall now prove that this actually happens by carefully evaluating the limit. From above,

$$\begin{aligned}\phi^a &= \frac{1}{R}\frac{dT}{ds}\frac{x^a}{r}\\ &= \frac{1}{R}\left\{\frac{1}{\tanh s} - \frac{B}{\tanh(Bs)}\right\}\frac{x^a}{r}.\end{aligned} \tag{3.6}$$

Now it is *important* to evaluate the limit $R \to \infty$ at *fixed* r. Since $r = 2R\tanh(s/2)$, for large R, $\tanh(s/2) \to 0$. Thus we must replace s by s_R, where

$$s_R/2 = r/2R.$$

The result of this is

$$\lim_{R\to\infty} \phi^a = \lim_{R\to\infty} \frac{1}{R}\left\{\frac{1}{\tanh(r/2R)} - \frac{(R+1)}{\tanh((1+R)r/2r)}\right\}\frac{x^a}{r}. \tag{3.7}$$

The limit is straightforward and gives the result

$$\phi^a_{\rm PS} = \left\{\frac{1}{r} - \frac{1}{\tanh r}\right\}\frac{x^a}{r}. \tag{3.8}$$

This is precisely the Higgs field for the Prasad–Sommerfield monopole.[10] An exactly similar result holds for the connection A_i^a. Thus we have verified that the limit of a hyperbolic monopole in $H^3(R)$ as $R \to \infty$ is indeed an ordinary monopole. This precise limiting procedure above is natural in Atiyah's approach and corresponds to the rescaling mechanism of Chakrabarti.

Now we return to hyperbolic space H^3 and would like to examine the question of the location of the monopole. This is really the same [for SU(2)] as discussing the zeros of the Higgs field ϕ. The existence of a zero of ϕ is forced by the nonvanishing of the magnetic charge k. The magnetic charge is the winding number of the map $\hat{\phi}: \partial H^3 \to S^2_{\rm su(2)}$, where $S^2_{\rm su(2)}$ is a sphere in the Lie algebra. A standard homotopy argument shows that $\hat{\phi}$ only extends to the interior H^3 in a singularityfree manner if the winding number k is zero. Hence for $k = 0$, ϕ has a zero, and in general k counts the zeros of ϕ. We find that there is one zero of ϕ at the origin of the ball H^3. Because of the dual interpretation of a hyperbolic monopole as an instanton, it is of interest to ask for something in the instanton picture to which the zero of ϕ corresponds. First of all, the origin of the ball is transformed, under i, to the point $r = 2$, $x_2 = 0$, $x_3 = 0$ in the upper half-plane model. This in turn corresponds to a circle of radius 2 in $\mathbf{R}^4$ centered at the origin and lying in the x_0-x_1 plane. To describe the axially symmetric instanton, it is appropriate, because of the use of the conformal correspondence (1.3), to use the Jackiw *et al.*[11] conformally invariant construction. This constructs an instanton with index c_2 by choosing $c_2 + 1$ centers y_i and sizes λ_i with

$$A = \sigma_{\mu\nu}\,\partial_\nu \ln\left\{\sum_{i=1}^{c_2+1} \frac{\lambda_i^2}{(x-y_i)^2}\right\}dx^\mu, \tag{3.9}$$

where we use the notation of Ref. 11. Now c_2 is proportional to k via the weight n, and, just as the integral (2.11) for k is actually all concentrated in a small sphere around the zero of ϕ, the integral for c_2 is concentrated around small spheres encircling each center y_i. This leads to the following form for A: to construct an axially symmetric A of weight n one should choose all the sizes λ_i to be equal, and all the centers to lie equally spaced on the circle of radius 2 in the x_0-x_1 plane. For a general R one works on $H^3(R)$ and the circle of radius 2 above has radius $2R$. Further as $R \to \infty$, the centers y_i become more and more numerous, and their angular separation smaller and smaller. Thus an ordinary monopole is like a ring of instantons. In this connection cf. also Chakrabarti.[1]

IV. HYPERBOLIC MONOPOLES AND INSTANTON SINGULARITIES

It is natural to consider the properties of a hyperbolic monopole for a general value of $|\phi|_{\rm as}$. We have seen how a half-integral n corresponds with an instanton but not one with axial symmetry. Now let n be such that $n/2$ is nonintegral, i.e., so that $|\phi|_{\rm as} \notin \mathbf{Z}$. Then the hyperbolic monopole (ϕ, A) exists—but the corresponding instanton A does not. Indeed were A to exist it would have nonintegral c_2. For an example of an instanton with nonintegral c_2, cf. Ref. 12. We shall now show that this state of affairs gives rise to an instanton A with a singularity. Recall that, regarded as an instanton, $\phi = A_\theta$, and that then $|\phi|$ is no longer an invariant. This is because of the freedom to make θ-dependent gauge transformations.

Now all along the "axis" $\mathbf{R}^2$ we have $|\phi| = |\phi|_{\rm as} = B - 1$. Also for an instanton to have finite action there exists a gauge in which $A_\mu \to g^{-1}\partial_\mu g$ at infinity. This suggests that, if we approach infinity along $\mathbf{R}^2$, there exists a gauge for which $A_\theta \to 0$ at infinity. We now construct this latter gauge.

Choose as gauge group element $g(x)$ with

$$g(x) = \exp[-\theta\phi_{\rm as}]. \tag{4.1}$$

Under $g(x)$, ϕ transforms to ϕ^g with

$$\begin{aligned}\phi^g &= g^{-1}\phi g + g^{-1}\partial_\theta g\\ &= \phi - \phi_{\rm as}.\end{aligned} \tag{4.2}$$

So $|\theta^g|_{\rm as}$ is zero as required. The important point is that this gauge transformation has to be implemented on the A_i as well. The gauge transform A_i^g of A_i is given by

$$A_i^g = \exp[\theta\phi_{\rm as}]A_i\exp[-\theta\phi_{\rm as}]. \tag{4.3}$$

With the effort of some linear algebra we eventually find that

$$\begin{aligned}A_i^g = A_i\cos(\theta|\phi|_{\rm as}) &- [A_i, \hat{\phi}_{\rm as}]\sin(\theta|\phi|_{\rm as})\\ &+ 2\,\mathrm{tr}(A_i\hat{\phi}_{\rm as})\cos(\theta|\phi|_{\rm as}) - 2\mathrm{tr}(A_i\hat{\phi}_{\rm as})\hat{\phi}_{\rm as}.\end{aligned} \tag{4.4}$$

Thus A_i^g has acquired θ dependence of a special kind. It is immediate that, for $|\phi|_{\rm as} \notin \mathbf{Z}$, A_i^g is not single valued and therefore possesses a branch singularity that will be located on the "axis" $\mathbf{R}^2$.

So the notion of a hyperbolic monopole is wider than that of an axially symmetric instanton. This makes it more plausible that the conversion of hyperbolic monopoles into ordinary $\mathbf{R}^3$-monopoles has a rigorous basis as this conversion requires a continuous variation of $|\phi|$. Also the question of whether there is a connection between the solutions of the second-order equations in H^3 and those of the second-order equations in $\mathbf{R}^3$ is of considerable interest. These matters are currently being investigated.

ACKNOWLEDGMENT

The author thanks Professor M. F. Atiyah for a private communication and A. Chakrabarti for correspondence.

[1]M. F. Atiyah, Commun. Math. Phys. **93**, 437 (1984); "Magnetic monopoles in hyperbolic space," in *Proceedings of the International Colloquium on Vector Bundles* (Tata Institute, Bombay, 1984); A. Chakrabarti, J. Math. Phys. **27**, 340 (1986).
[2]E. Bogomolny, Sov. Nucl. Phys. **24**, 449 (1976).
[3]E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[4]P. Forgács and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[5]Our notation is $|\phi| = \sqrt{\phi^a\phi^a} = \sqrt{-2\,\mathrm{tr}(\phi^2)}$, $\phi = \phi^a\sigma_a/2i$, σ^a the Pauli matrices.
[6]In a general curved space-time the self-duality equations need not reduce to an equation of Bogomolny type when one imposes invariance under a coordinate transformation, such as time translation or rotation through an angle; it is completely straightforward to check that they do so here, however. For a discussion of this more general situation, cf. A. Comtet, P. Forgács, and P. Horváthy, Phys. Rev. D **30**, 468 (1984).
[7]H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, Phys. Rev. D **20**, 1898 (1979).
[8]M. F. Atiyah and R. Bott, Topology **23**, (1984); M. F. Atiyah, in Ref. 1.
[9]For example, cf. C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic, New York, 1983).
[10]M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[11]R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1976).
[12]P. Forgács, Z. Horvath, and L. Palla, Phys. Rev. Lett. **46**, 392 (1981).

Higher-order interaction energies for a system of *N* arbitrary molecules in the light of spherical tensor theory

Piotr Piecuch
Institute of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland

(Received 12 November 1984; accepted for publication 12 March 1986)

The Rayleigh–Schrödinger perturbation treatment accompanied by a multipole expansion of the interaction potential can be used in the quantum-mechanical studies of long-range intermolecular forces. The total third-order interaction energy in a system of N molecules separates into several induction and dispersion categories. Only some of them are purely pairwise additive. In the present paper the closed expressions for all possible categories of the third-order anisotropic interaction energy in a collection of N arbitrary molecules are derived, in which the orientational dependence is pushed to its limits. The derivation is based on the spherical tensor formalism. The formulas for the induction energies are directly related to the spherical multipole moments and irreducible (hyper)polarizabilities of the interacting molecules in body-fixed frames. The same is achieved for the dispersion categories after employing some simple approximations. The present paper can be treated as an extension of the spherical tensor description of the two-body long-range molecular interactions for the most important quantum-mechanical pairwise nonadditive forces.

I. INTRODUCTION

The Hamiltonian H for a system of N interacting molecules can be written as

$$H = \sum_{i=1}^{N} H_i^{(0)} + V, \tag{1}$$

where $H_i^{(0)}$ is the Hamiltonian for the ith isolated molecule and V is the operator for the interaction energy. The operator V can be expressed as follows:

$$V = \frac{1}{2} \sum_{i,j=1}^{N} V_{ij} \quad (V_{ij} = V_{ji}, \quad V_{ii} = 0), \tag{2}$$

where V_{ij} is the potential energy operator including Coulomb interactions between molecules i and j. In the present paper we consider long-range intermolecular interactions, i.e., each molecule is far enough away from the other molecules so that neither exchange effects nor charge overlap among molecules is important, and close enough to the other molecules so that very long-range retardation effects[1] can be ignored. In this case we can obtain a very good approximation of the real potential V by expanding each V_{ij} in a multipole series. Such an expanded potential V can be used in quantum-mechanical calculations of long-range forces.[2] For example, we can apply simple Rayleigh–Schrödinger perturbation theory for the description of long-range interactions, treating this multipole approximation of V as a perturbation.[3–6]

In such a formulation of the problem, the application of the spherical tensor formalism seems to be very convenient. It was shown that in the case of two molecules, this technique leads to the separation of interaction energy into the terms including irreducible tensors localized on interacting molecules, providing closed expressions for the interaction energy in all orders of the perturbation.[7–14] In these expressions the orientational dependence is simplified to the utmost and thus they are very suitable for studying the anisotropy of the interactions between two molecules.[15] It was also shown that from orientationally dependent interaction energy expressions one can easily obtain clear orientationally averaged (i.e., isotropic) formulas.[7,8,12] These formulas are convenient for the understanding of some properties of the rare gases and liquids, where the molecules rotate more or less freely; they also can be important for the description of the interactions in molecular crystals, where a given molecule often sees surrounding molecules with many different orientations and the anisotropic interactions cancel to a large extent.[16]

However, long-range interactions in a system of N molecules ($N \geqslant 3$) are not pairwise additive,[6,17,18] so this clear description based on the spherical tensor formalism concerning long-range interactions between pairs of molecules is inadequate when the interactions between N molecules are considered. The long-range interaction energy arising in the first order of the perturbation is, of course, pairwise additive, because it represents classical electrostatic interaction energy. The second-order interaction energy separates into a purely pairwise additive dispersion term and an induction term, which contains pairwise additive as well as pairwise nonadditive components.[19] So the nonadditivity can first occur in the second order of the perturbation, and in such systems as molecular crystals it is not negligible and even can be important.[20] Recently we have extended the use of the spherical tensor formalism for these nonadditive induction terms appearing in the second-order interaction energy.[18] Therefore we have generalized the spherical tensor description of long-range molecular interactions in the first two orders of the perturbation to a system of N interacting molecules. This generalization once more shows that the spherical tensor formalism gives us the best method for the theoretical description of long-range intermolecular forces.

To obtain a more complete theoretical picture of the long-range interactions in a system of N molecules we must consider the third-order interactions where, in contrast with the second order, the nonadditivity of the dispersion forces

 0022-2488/86/082165-23$02.50

appears.[17] It should be noted that the third-order nonadditive dispersion interactions are not only interesting from the theoretical point of view; they are worth consideration because of their importance at physically meaningful intermediate intermolecular distances. Although these nonadditive dispersion effects are often small, it has been known for some time that many interesting physical properties or phenomena may be under their control. They are important for the understanding of several properties of imperfect gases (e.g., third virial coefficients of inert gases[21]) because they have influence on their equation of state[22]; they also have influence on the effective pair potentials and radial distribution functions in low density gases.[21] The third-order nonadditive dispersion forces become important when the interaction between two molecules embedded in a nonpolar medium is considered. This fact can be applied for the understanding of the interactions between base pairs in a single DNA double helix as well as for the explaining of the reduction of the effective intermolecular pair potentials in homogeneous nonpolar liquids.[21,23] This fact is also meaningful when the interactions between molecules immersed in a liquid solvent are considered. Such considerations, in which the nonadditivity of the dispersion forces is assumed, show that the solvent plays a crucial role in the modification of two-body intermolecular potentials. Obtained in this way, effective pair potentials are very suitable for the description of the thermodynamical properties of solutions, solute–solute, solute–solvent, and solvent–solvent interactions; they are also useful when the importance of the solvent effects in molecular associations between long polymer chains in various solutions, particularly in biological systems, is examined.[21,23,24] The nonadditive dispersion forces are an important contribution to the lattice energy and other properties of the rare gas solids, i.e., elastic constants.[21,25,26] They also can play any role in the polymorphism in the solid state.[27] Finally these forces contribute very strongly to the physical adsorption and the interactions between adsorbed molecules[21,28]; they are also large in the interactions of colloids.[29]

Stogryn[17] has proved that the total third-order interaction energy in a system of N molecules W_3 separates into five categories

$$W_3 = W_B + W_A + W_D + W_{BA} + W_{CD}, \tag{3}$$

where W_B and W_A represent induction interactions, while W_D, W_{BA}, and W_{CD} represent dispersion interactions. In formula (3) only the last dispersion term W_{CD} is purely pairwise additive; the remaining terms include pairwise nonadditive components.[17] Taking into account the necessity of considering the third-order interactions for a system of N molecules, together with the advantages of the spherical tensor treatment of two-body long-range interactions in all orders of the perturbation theory and similar advantages of the spherical tensor description of long-range forces between N molecules in the first two orders of the perturbation, in the present paper we extend the application of the spherical tensor formalism for the third-order long-range anisotropic interaction energy expressions in a system of N arbitrary molecules. As we will see, the closed equations for W_B, W_A, W_D, W_{BA}, and W_{CD} derived in this paper have the same nice properties as the previous results for two interacting molecules and N interacting molecules, when perturbation treatment up to the second order is used, i.e., they are related to the irreducible tensors localized on interacting molecules in body-fixed frames, and the orientational dependence that appears in them is pushed to its limits. Our formulas describing the third-order electric induction interactions in a collection of N molecules are directly expressed in terms of the irreducible tensors describing electrical properties of each isolated molecule [spherical multipole moments and irreducible spherical (hyper)polarizabilities[30]] in the local coordinate system fixed in it. This fact is important from the practical point of view, because spherical multipole moments and (hyper)polarizabilities are very convenient when higher-order electric interactions are considered.[9–14,18,30,31] So in the last section of the present paper we briefly describe how, by using simple approximations, we can simplify the exact formulas for the third-order dispersion interactions between N molecules and connect them with the above-mentioned molecular properties. At the end of this section let us notice that the set of the expressions obtained in the present paper, together with the earlier results concerning the interactions between N molecules in the first- and the second-order perturbation theory,[7–10,13,18] form a practically complete spherical tensor description of the long-range molecular interactions including quantum-mechanical many-body effects, because in a certain sense all possible kinds of long-range forces appear when perturbation treatment up to the third order is used (viz., classical electrostatic forces between molecules having permanent multipoles, pairwise additive as well as pairwise nonadditive induction forces, and pairwise additive as well as pairwise nonadditive dispersion forces). In higher orders only more complicated categories of the above types of interactions occur[32] and we can expect that they are much smaller than the above-mentioned categories. In the near future we will demonstrate that this practically complete spherical tensor description of long-range interactions between N arbitrary molecules can be useful not only when the anisotropy of the intermolecular forces is considered, but also when the isotropic interactions including quantum-mechanical many-body effects are examined.[33] This fact again shows that the spherical tensor formalism is the best and a very effective way to describe the long-range intermolecular forces.

II. INTERACTION ENERGY OPERATORS AND MOLECULAR PROPERTIES OCCURRING IN THE SPHERICAL TENSOR THEORY OF LONG-RANGE INTERACTIONS BETWEEN N MOLECULES

As we have explained in the previous section, in the case of long-range intermolecular forces, we can replace each interaction energy operator V_{ij} $(i,j = 1,2,...,N)$ by its multipole expansion. Many forms of the multipole expansions are known in the literature.[6,34–41] However, for our purpose the most suitable is the following one[7–13,18,35–37,42,43]:

$$V_{ij} = (4\pi)^{1/2} \sum_{l_i=0}^{\infty} \sum_{l_j=0}^{\infty} (-1)^{l_j} \binom{2l_i + 2l_j}{2l_i}^{1/2} \times [l_i + l_j]^{-1} R_{ij}^{-l_i - l_j - 1} \times \sum_{\eta_{ij} = -l_i - l_j}^{l_i + l_j} (-1)^{\eta_{ij}} Y_{l_i + l_j}^{-\eta_{ij}} (\hat{R}_{ij}) \times [\hat{\mathbf{Q}}_{l_i} \cdot \mathbf{D}^{l_i}(\omega_i^{-1}) \otimes \hat{\mathbf{Q}}_{l_j} \cdot \mathbf{D}^{l_j}(\omega_j^{-1})]_{l_i + l_j}^{\eta_{ij}}. \quad (4)$$

Here $\hat{\mathbf{Q}}_{l_i}$ and $\hat{\mathbf{Q}}_{l_j}$ are the spherical multipole moment operators for molecules i and j, respectively, defined as usual.[7-11,13,18,30] They are related to the local coordinate systems fixed in molecules. The intermolecular vector $\mathbf{R}_{ij}$ points from i to j and $(R_{ij}, \hat{R}_{ij})$ are the spherical components of $\mathbf{R}_{ij}$ in the global coordinate system fixed in the space; ω_i and ω_j are the sets of the Euler angles describing the orientations of the local coordinate systems fixed in molecules i and j, respectively, with respect to the space-fixed local coordinate systems with axes parallel to the axes of the global space-fixed frame; $\mathbf{D}^j(\omega)$ is the matrix, which represents a rotation ω in the $(2j+1)$-dimensional irreducible representation of the SO(3) group; and Y_L^N denotes the usual surface spherical harmonic according to the phase convention given by Condon and Shortley.[44] The irreducible tensor product between two sets of irreducible tensors $\mathbf{T}_{k_1} = \{T_{k_1}^{q_1}: q_1 = -k_1,...,k_1\}$ and $\mathbf{T}_{k_2} = \{T_{k_2}^{q_2}: q_2 = -k_2,...,k_2\}$ is defined as follows[45]:

$$[\mathbf{T}_{k_1} \otimes \mathbf{T}_{k_2}]_k^q = \sum_{q_1 q_2} T_{k_1}^{q_1} T_{k_2}^{q_2} \langle k_1 q_1, k_2 q_2 | kq \rangle,$$

where $\langle k_1 q_1, k_2 q_2 | kq \rangle$ denotes the Clebsch–Gordan coefficient. According to the Wigner abbreviation,

$$[j_1, j_2, ..., j_k] = [(2j_1 + 1)(2j_2 + 1) \cdots (2j_k + 1)]^{1/2}.$$

Equation (4) can be obtained from all forms of the multipole series described earlier.[10,43] It can be also derived directly.[7,13,35-37,42] In the previous considerations concerning the spherical tensor description of long-range intermolecular forces, it has been proved that the multipole expansion (4) is very useful.[7-13,18] Therefore it seems to be natural to use it in the considerations presented in this work, which can be treated as an extension of the spherical tensor theory of long-range molecular interactions for a system of N interacting molecules.

As it was pointed out in the Introduction our theory requires several kinds of molecular properties describing the behavior of a polarizable molecular charge distribution in an external electric field. These are the following: spherical multipole moments $\mathbf{Q}_l$, irreducible polarizability tensors $\boldsymbol{\alpha}_l\{l'l''\}$, and irreducible hyperpolarizability tensors $\boldsymbol{\beta}_L\{(l'l'')l,l'''\}_{[(1+2)+3]}$, $\boldsymbol{\beta}_L\{(l''l''')l,l'\}_{[(2+3)+1]}$, and $\boldsymbol{\beta}_L\{(l'''l')l,l''\}_{[(3+1)+2]}$. We must notice that here and elsewhere in the present paper, the above spherical tensors are defined in molecular (body-fixed) frames. Multipole moments $\mathbf{Q}_l$ and irreducible polarizabilities $\boldsymbol{\alpha}_l\{l'l''\}$ are defined in the usual way in accordance with the classical work of Gray and Lo[30] published in 1976. To define the irreducible hyperpolarizabilities $\boldsymbol{\beta}_L\{(l'l'')l,l'''\}_{[(1+2)+3]}$, $\boldsymbol{\beta}_L\{(l''l''')l,l'\}_{[(2+3)+1]}$, and $\boldsymbol{\beta}_L\{(l'''l')l,l''\}_{[(3+1)+2]}$, we must first introduce the unsymmetrized spherical reducible hyperpolarizability tensors $\boldsymbol{\beta}_{l'l''l'''}$. They are defined in the following way:

$$\beta_{l'l''l'''}^{m'm''m'''} = \sum_{p,q \neq g} \frac{\langle g|\hat{Q}_{l'}^{m'}|p\rangle \langle p|\hat{Q}_{l''}^{m''}|q\rangle \langle q|\hat{Q}_{l'''}^{m'''}|g\rangle}{\epsilon(p)\epsilon(q)} - \sum_{p \neq g} \frac{\langle g|\hat{Q}_{l'}^{m'}|p\rangle Q_{l''}^{m''} \langle p|\hat{Q}_{l'''}^{m'''}|g\rangle}{\epsilon(p)^2}, \quad (5)$$

where $|g\rangle$ denotes the ground eigenstate of a given molecule, while $|p\rangle$ and $|q\rangle$ are the excited molecular eigenstates; $\epsilon(p) = E_p - E_g$ and $\epsilon(q) = E_q - E_g$, where E_g, E_p, E_q are the molecular energies corresponding to the states $|g\rangle$, $|p\rangle$, $|q\rangle$, respectively. We see that this definition differs from the analogous one introduced by Gray and Lo in their classical paper.[30] However, if we symmetrize the tensor $\boldsymbol{\beta}_{l'l''l'''}$ given by Eq. (5) in the way described by Gray and Lo, we get the symmetrized spherical hyperpolarizability tensor $\bar{\boldsymbol{\beta}}_{l'l''l'''}$ introduced by these authors[30]; indeed

$$\bar{\beta}_{l'l''l'''}^{m'm''m'''} = S\{\beta_{l'l''l'''}^{m'm''m'''}\}, \quad (6)$$

where S implies a summation over all possible terms in which the $\binom{m}{l}$ pairs are permuted. Therefore the difference between our definition of the unsymmetrized spherical hyperpolarizability tensor $\boldsymbol{\beta}_{l'l''l'''}$ and the analogous one given by Gray and Lo[30] is unimportant, because in the description of physical phenomena, which are based on the interaction between the molecule and the external electric field, symmetrized (hyper)polarizability tensors occur.[2,17,30,32] Irreducible hyperpolarizabilities $\boldsymbol{\beta}_L\{(l'l'')l,l'''\}_{[(1+2)+3]}$, $\boldsymbol{\beta}_L\{(l''l''')l,l'\}_{[(2+3)+1]}$, and $\boldsymbol{\beta}_L\{(l'''l')l,l''\}_{[(3+1)+2]}$ can be obtained by standard coupling methods from the reducible tensors $\boldsymbol{\beta}_{l'l''l'''}$ defined by Eq. (5) in the following way:

$$\beta_L^M\{(l'l'')l,l'''\}_{[(1+2)+3]} = \sum_{m'm''m'''m} \beta_{l'l''l'''}^{m'm''m'''} \langle l'm',l''m''|lm\rangle \langle lm,l'''m'''|LM\rangle,$$

$$\beta_L^M\{(l''l''')l,l'\}_{[(2+3)+1]} = \sum_{m'm''m'''m} \beta_{l'l''l'''}^{m'm''m'''} \langle l''m'',l'''m'''|lm\rangle \langle lm,l'm'|LM\rangle,$$

$$\beta_L^M\{(l'''l')l,l''\}_{[(3+1)+2]} = \sum_{m'm''m'''m} \beta_{l'l''l'''}^{m'm''m'''} \langle l'''m''',l'm'|lm\rangle \langle lm,l''m''|LM\rangle. \quad (7)$$

Here we have used square brackets in $\boldsymbol{\beta}_L\{\cdots\}_{[\dots]}$ to indicate the order of couplings. For a convenience, at the end of this section we introduce the following abbreviation:

$$\tilde{\boldsymbol{\beta}}_L\{(l'l'')l,l'''\} = \boldsymbol{\beta}_L\{(l'l'')l,l'''\}_{[(1+2+3]} + \boldsymbol{\beta}_L\{l'l'')l,l'''\}_{[(2+3)+1]} + \boldsymbol{\beta}_L\{(l'l'')l,l'''\}_{[(3+1)+2]}. \quad (8)$$

Note that

$$\tilde{\beta}_L^M\{(l'l'')l,l'''\}$$

$$= \sum_{m'm''m'''m} (\beta_{l'l''l'''}^{m'm''m'''} + \beta_{l''l'''l'}^{m''m'''m'} + \beta_{l'''l'l''}^{m'''m'm''}) \times \langle l'm',l''m''|lm\rangle \langle lm,l'''m'''|LM\rangle$$

$$= \frac{1}{2} \sum_{m'm''m'''m} \bar{\beta}_{l'l''l'''}^{m'm''m'''} \times \langle l'm'',l''m''|lm\rangle \langle lm,l'''m'''|LM\rangle, \tag{9}$$

where $\bar{\beta}_{l'l''l'''}^{m'm''m'''}$ are the symmetrized spherical hyperpolarizabilities defined by Gray and Lo[30] [see Eq. (6)]. This means that $\tilde{\boldsymbol{\beta}}_L\{(l'l'')l,l'''\}$ can be called symmetrized irreducible spherical hyperpolarizabilities.

III. INTRODUCTORY REMARKS ON THE THIRD-ORDER INTERACTIONS BETWEEN N ARBITRARY MOLECULES

In this section we prepare ourselves to the derivation of the formulas describing the third-order anisotropic interactions in a collection of N molecules. Application of the Rayleigh–Schrödinger perturbation method to the "ground state" eigenfunction $|G\rangle$ of the unperturbed Hamiltonian $H^{(0)} = \Sigma_{i=1}^N H_i^{(0)}$, namely the product

$$|G\rangle = |g_1 g_2 \cdots g_N\rangle = |g_1\rangle |g_2\rangle \cdots |g_N\rangle$$

of the ground states of the Hamiltonians $H_1^{(0)}, H_2^{(0)}, \ldots, H_N^{(0)}$ leads to the following expressions for the third-order interaction energies[17]:

$$W_B = \sum_{a,b,c,d=1}^{N} \Bigg[\sum_{p_a \neq g_a,\, q_a \neq g_a} \frac{\langle g_a g_b|V_{ab}|p_a g_b\rangle \langle p_a g_c|V_{ac}|q_a g_c\rangle \langle q_a g_d|V_{ad}|g_a g_d\rangle}{\epsilon(p_a)\epsilon(q_a)} - \sum_{q_a \neq g_a} \frac{\langle g_a g_b|V_{ab}|q_a g_b\rangle \langle g_a g_c|V_{ac}|g_a g_c\rangle \langle q_a g_d|V_{ad}|g_a g_d\rangle}{\epsilon(q_a)^2} \Bigg], \tag{10}$$

$$W_A = 2 \sum_{a,b,c,d=1}^{N} \sum_{p_b \neq g_b,\, q_c \neq g_c} \frac{\langle g_a g_b|V_{ab}|g_a p_b\rangle \langle p_b g_c|V_{bc}|g_b q_c\rangle \langle q_c g_d|V_{cd}|g_c g_d\rangle}{\epsilon(p_b)\epsilon(q_c)}, \tag{11}$$

$$W_D = \sum_{a,b,c=1}^{N} \sum_{p_a \neq g_a,\, p_b \neq g_b,\, p_c \neq g_c} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_b g_c|V_{bc}|g_b p_c\rangle \langle p_a p_c|V_{ac}|g_a g_c\rangle}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(p_a) + \epsilon(p_c)]}, \tag{12}$$

$$W_{BA} = \sum_{a,b,c=1}^{N} \Bigg[2 \sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a p_b|V_{ab}|q_a g_b\rangle \langle q_a g_c|V_{ac}|g_a g_c\rangle}{[\epsilon(p_a) + \epsilon(p_b)]\epsilon(q_a)} - 2 \sum_{\substack{p_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle g_a p_b|V_{ab}|g_a g_b\rangle \langle p_a g_c|V_{ac}|g_a g_c\rangle}{[\epsilon(p_a) + \epsilon(p_b)]\epsilon(p_a)}$$
$$+ \sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a g_c|V_{ac}|q_a g_c\rangle \langle q_a p_b|V_{ab}|g_a g_b\rangle}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(q_a) + \epsilon(p_b)]} - \langle g_a g_c|V_{ac}|g_a g_c\rangle \sum_{\substack{p_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a p_b|V_{ab}|g_a g_b\rangle}{[\epsilon(p_a) + \epsilon(p_b)]^2} \Bigg], \tag{13}$$

$$W_{CD} = \frac{1}{2} \sum_{a,b,=1}^{N} \Bigg[\sum_{\substack{p_a, q_a \neq g_a \\ p_b, q_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a p_b|V_{ab}|q_a q_b\rangle \langle q_a q_b|V_{ab}|g_a g_b\rangle}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(q_a) + \epsilon(q_b)]} - 2 \sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a p_b|V_{ab}|q_a g_b\rangle \langle q_a p_b|V_{ab}|g_a g_b\rangle}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(q_a) + \epsilon(p_b)]}$$
$$+ \langle g_a g_b|V_{ab}|g_a g_b\rangle \sum_{\substack{p_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a g_b|V_{ab}|p_a p_b\rangle \langle p_a p_b|V_{ab}|g_a g_b\rangle}{[\epsilon(p_a) + \epsilon(p_b)]^2} \Bigg]. \tag{14}$$

Use of Eq. (4) in formulas (10)–(14) together with some algebraical manipulations involving the summation indices[46] leads to the following expressions for W_B, W_A, W_D, W_{BA}, and W_{CD}:

$$W_B = \frac{1}{3} \sum_{a,b,c,d=1}^{N} \sum_{\substack{l_a' l_a'' l_a''' l_b l_c l_d \\ n_a' n_a'' n_a''' n_b n_c n_d \\ m_a' m_a'' m_a''' m_b m_c m_d}} C_{l_a' l_b}^{n_a' n_b} C_{l_a'' l_c}^{n_a'' n_c} C_{l_a''' l_d}^{n_a''' n_d} (\beta_{l_a' l_a'' l_a'''}^{m_a' m_a'' m_a'''} + \beta_{l_a'' l_a''' l_a'}^{m_a'' m_a''' m_a'} + \beta_{l_a''' l_a' l_a''}^{m_a''' m_a' m_a''})$$
$$\times Q_{l_b}^{m_b} Q_{l_c}^{m_c} Q_{l_d}^{m_d} D_{m_a' n_a'}^{l_a'}(\omega_a^{-1}) D_{m_a'' n_a''}^{l_a''}(\omega_a^{-1}) D_{m_a''' n_a'''}^{l_a'''}(\omega_a^{-1}) D_{m_b n_b}^{l_b}(\omega_b^{-1}) D_{m_c n_c}^{l_c}(\omega_c^{-1}) D_{m_d n_d}^{l_d}(\omega_d^{-1}), \tag{15}$$

where $\beta^{m'_a m''_a m'''_a}_{l'_a l''_a l'''_a}$ is the spherical hyperpolarizability of molecule a defined by Eq. (5);

$$W_A = 2 \sum_{a,b,c,d=1}^{N} \sum_{\substack{l_a l'_b l''_b l'_c l''_c l_d \\ n_a n'_b n''_b n'_c n''_c n_d \\ m_a m'_b m''_b m'_c m''_c m_d}} C^{n_a n'_b}_{l_a l'_b} C^{n''_b n'_c}_{l''_b l'_c} C^{n''_c n_d}_{l''_c l_d} Q^{m_a}_{l_a} \alpha^{m'_b m''_b}_{l'_b l''_b} \alpha^{m'_c m''_c}_{l'_c l''_c} Q^{m_d}_{l_d}$$

$$\times D^{l_a}_{m_a n_a}(\omega_a^{-1}) D^{l'_b}_{m'_b n'_b}(\omega_b^{-1}) D^{l''_b}_{m''_b n''_b}(\omega_b^{-1}) D^{l'_c}_{m'_c n'_c}(\omega_c^{-1}) D^{l''_c}_{m''_c n''_c}(\omega_c^{-1}) D^{l_d}_{m_d n_d}(\omega_d^{-1}), \tag{16}$$

where $\alpha^{m'_b m''_b}_{l'_b l''_b}$ and $\alpha^{m'_c m''_c}_{l'_c l''_c}$ are the unsymmetrized spherical polarizabilities of molecules b and c, respectively, defined by Gray and Lo[30];

$$W_D = \frac{2}{3} \sum_{a,b,c=1}^{N} \sum_{\substack{l'_a l''_a l'_b l''_b l'_c l''_c \\ n'_a n''_a n'_b n''_b n'_c n''_c \\ m'_a m''_a m'_b m''_b m'_c m''_c}} C^{n''_a n'_b}_{l''_a l'_b} C^{n''_b n'_c}_{l''_b l'_c} C^{n''_c n'_a}_{l''_c l'_a} \, {}^{D}t^{m'_a m''_a m'_b m''_b m'_c m''_c}_{l'_a l''_a l'_b l''_b l'_c l''_c}$$

$$\times D^{l'_a}_{m'_a n'_a}(\omega_a^{-1}) D^{l''_a}_{m''_a n''_a}(\omega_a^{-1}) D^{l'_b}_{m'_b n'_b}(\omega_b^{-1}) D^{l''_b}_{m''_b n''_b}(\omega_b^{-1}) D^{l'_c}_{m'_c n'_c}(\omega_c^{-1}) D^{l''_c}_{m''_c n''_c}(\omega_c^{-1}), \tag{17}$$

where

$${}^{D}t^{m'_a m''_a m'_b m''_b m'_c m''_c}_{l'_a l''_a l'_b l''_b l'_c l''_c} = \sum_{p_a \neq g_a, p_b \neq g_b, p_c \neq g_c} D \langle g_a|\hat{Q}^{m'_a}_{l'_a}|p_a\rangle\langle p_a|\hat{Q}^{m''_a}_{l''_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle\langle g_c|\hat{Q}^{m'_c}_{l'_c}|p_c\rangle\langle p_c|\hat{Q}^{m''_c}_{l''_c}|g_c\rangle$$

and

$$D = \frac{\epsilon(p_a) + \epsilon(p_b) + \epsilon(p_c)}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(p_b) + \epsilon(p_c)][\epsilon(p_c) + \epsilon(p_a)]};$$

$$W_{BA} = \sum_{a,b,c=1}^{N} \sum_{\substack{l'_a l''_a l'''_a l'_b l''_b l_c \\ n'_a n''_a n'''_a n'_b n''_b n_c \\ m'_a m''_a m'''_a m'_b m''_b m_c}} C^{n'_a n'_b}_{l'_a l'_b} C^{n''_a n''_b}_{l''_a l''_b} C^{n'''_a n_c}_{l'''_a l_c} \, {}^{BA}t^{m'_a m''_a m'''_a m'_b m''_b m_c}_{l'_a l''_a l'''_a l'_b l''_b l_c}$$

$$\times D^{l'_a}_{m'_a n'_a}(\omega_a^{-1}) D^{l''_a}_{m''_a n''_a}(\omega_a^{-1}) D^{l'''_a}_{m'''_a n'''_a}(\omega_a^{-1}) D^{l'_b}_{m'_b n'_b}(\omega_b^{-1}) D^{l''_b}_{m''_b n''_b}(\omega_b^{-1}) D^{l_c}_{m_c n_c}(\omega_c^{-1}), \tag{18}$$

where

$$\begin{aligned}
{}^{BA}t^{m'_a m''_a m'''_a m'_b m''_b m_c}_{l'_a l''_a l'''_a l'_b l''_b l_c} = &\sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a|\hat{Q}^{m'_a}_{l'_a}|p_a\rangle\langle p_a|\hat{Q}^{m''_a}_{l''_a}|q_a\rangle\langle q_a|\hat{Q}^{m'''_a}_{l'''_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(p_a) + \epsilon(p_b)]\epsilon(q_a)} \\
&+ \sum_{\substack{q_a, p_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a|\hat{Q}^{m'''_a}_{l'''_a}|p_a\rangle\langle p_a|\hat{Q}^{m'_a}_{l'_a}|q_a\rangle\langle q_a|\hat{Q}^{m''_a}_{l''_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(q_a) + \epsilon(p_b)]\epsilon(p_a)} \\
&+ \sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a|\hat{Q}^{m''_a}_{l''_a}|p_a\rangle\langle p_a|\hat{Q}^{m'''_a}_{l'''_a}|q_a\rangle\langle q_a|\hat{Q}^{m'_a}_{l'_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(p_a) + \epsilon(p_b)][\epsilon(q_a) + \epsilon(p_b)]} \\
&- \sum_{p_a \neq g_a, p_b \neq g_b} \frac{\langle g_a|\hat{Q}^{m'_a}_{l'_a}|p_a\rangle Q^{m''_a}_{l''_a}\langle p_a|\hat{Q}^{m'''_a}_{l'''_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(p_a) + \epsilon(p_b)]\epsilon(p_a)} \\
&- \sum_{p_a \neq g_a, p_b \neq g_b} \frac{\langle g_a|\hat{Q}^{m'''_a}_{l'''_a}|p_a\rangle Q^{m'_a}_{l'_a}\langle p_a|\hat{Q}^{m''_a}_{l''_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(p_a) + \epsilon(p_b)]\epsilon(p_a)} \\
&- \sum_{p_a \neq g_a, p_b \neq g_b} \frac{\langle g_a|\hat{Q}^{m''_a}_{l''_a}|p_a\rangle Q^{m'''_a}_{l'''_a}\langle p_a|\hat{Q}^{m'_a}_{l'_a}|g_a\rangle\langle g_b|\hat{Q}^{m'_b}_{l'_b}|p_b\rangle\langle p_b|\hat{Q}^{m''_b}_{l''_b}|g_b\rangle Q^{m_c}_{l_c}}{[\epsilon(p_a) + \epsilon(p_b)]^2},
\end{aligned}$$

$$W_{CD} = \frac{1}{2} \sum_{a,b=1}^{N} \sum_{\substack{l'_a l''_a l'''_a l'_b l''_b l'''_b \\ n'_a n''_a n'''_a n'_b n''_b n'''_b \\ m'_a m''_a m'''_a m'_b m''_b m'''_b}} C^{n'_a n'_b}_{l'_a l'_b} C^{n''_a n''_b}_{l''_a l''_b} C^{n'''_a n'''_b}_{l'''_a l'''_b} \, {}^{CD}t^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}$$

$$\times D^{l'_a}_{m'_a n'_a}(\omega_a^{-1}) D^{l''_a}_{m''_a n''_a}(\omega_a^{-1}) D^{l'''_a}_{m'''_a n'''_a}(\omega_a^{-1}) D^{l'_b}_{m'_b n'_b}(\omega_b^{-1}) D^{l''_b}_{m''_b n''_b}(\omega_b^{-1}) D^{l'''_b}_{m'''_b n'''_b}(\omega_b^{-1}), \tag{19}$$

where

$$
{}^{CD}t^{m_a' m_a'' m_a''' m_b' m_b'' m_b'''}_{l_a' l_a'' l_a''' l_b' l_b'' l_b'''} = \sum_{\substack{p_a, q_a \neq g_a \\ p_b, q_b \neq g_b}} \frac{\langle g_a|\hat{Q}^{m_a'}_{l_a'}|p_a\rangle\langle p_a|\hat{Q}^{m_a''}_{l_a''}|q_a\rangle\langle q_a|\hat{Q}^{m_a'''}_{l_a'''}|g_a\rangle\langle g_b|\hat{Q}^{m_b'}_{l_b'}|p_b\rangle\langle p_b|\hat{Q}^{m_b''}_{l_b''}|q_b\rangle\langle q_b|\hat{Q}^{m_b'''}_{l_b'''}|g_b\rangle}{[\epsilon(p_a)+\epsilon(p_b)][\epsilon(q_a)+\epsilon(q_b)]}
$$

$$
- \sum_{\substack{p_a, q_a \neq g_a \\ p_b \neq g_b}} \frac{\langle g_a|\hat{Q}^{m_a'}_{l_a'}|p_a\rangle\langle p_a|\hat{Q}^{m_a''}_{l_a''}|q_a\rangle\langle q_a|\hat{Q}^{m_a'''}_{l_a'''}|g_a\rangle\langle g_b|\hat{Q}^{m_b'}_{l_b'}|p_b\rangle Q^{m_b''}_{l_b''}\langle p_b|\hat{Q}^{m_b'''}_{l_b'''}|g_b\rangle}{[\epsilon(p_a)+\epsilon(p_b)][\epsilon(q_a)+\epsilon(p_b)]}
$$

$$
- \sum_{\substack{p_b, q_b \neq g_b \\ p_a \neq g_a}} \frac{\langle g_a|\hat{Q}^{m_a'}_{l_a'}|p_a\rangle Q^{m_a''}_{l_a''}\langle p_a|\hat{Q}^{m_a'''}_{l_a'''}|g_a\rangle\langle g_b|\hat{Q}^{m_b'}_{l_b'}|p_b\rangle\langle p_b|\hat{Q}^{m_b''}_{l_b''}|q_b\rangle\langle q_b|\hat{Q}^{m_b'''}_{l_b'''}|g_b\rangle}{[\epsilon(p_a)+\epsilon(p_b)][\epsilon(q_b)+\epsilon(p_a)]}
$$

$$
+ \sum_{p_a \neq g_a, p_b \neq g_b} \frac{\langle g_a|\hat{Q}^{m_a'}_{l_a'}|p_a\rangle Q^{m_a''}_{l_a''}\langle p_a|\hat{Q}^{m_a'''}_{l_a'''}|g_a\rangle\langle g_b|\hat{Q}^{m_b'}_{l_b'}|p_b\rangle Q^{m_b''}_{l_b''}\langle p_b|\hat{Q}^{m_b'''}_{l_b'''}|g_b\rangle}{[\epsilon(p_a)+\epsilon(p_b)]^2}.
$$

For brevity in Eqs. (15)–(19) we have introduced the quantities $C^{n_i n_j}_{l_i l_j}$ $(i,j = 1,2,...,N,\ i \neq j)$, which are defined as[11,12,18,47]

$$
C^{n_i n_j}_{l_i l_j} = (-1)^{l_j}(4\pi)^{1/2}[l_i+l_j]^{-1}\binom{2l_i+2l_j}{2l_i}^{1/2} R_{ij}^{-l_i-l_j-1} \sum_{\eta_{ij}=-l_i-l_j}^{l_i+l_j} (-1)^{\eta_{ij}} Y^{-\eta_{ij}}_{l_i+l_j}(\hat{R}_{ij})\langle l_i n_i, l_j n_j | l_i+l_j \eta_{ij}\rangle. \quad (20)
$$

The general sum over a, b, c, and d in Eq. (15) can be divided into three parts corresponding to the quantum-mechanical two-, three-, and four-body interactions, viz.,

$$
W_B = \sum_{a,b,c,d=1}^{N} (\cdots) = W_{B,2} + W_{B,3} + W_{B,4}, \quad (21)
$$

where

$$
W_{B,2} = \sum_{\substack{a,b,c,d=1 \\ (b=c=d)}}^{N} (\cdots), \quad W_{B,3} = \sum_{\substack{a,b,c,d=1 \\ (b=c\neq d)}}^{N} (\cdots) + \sum_{\substack{a,b,c,d=1 \\ (c=d\neq b)}}^{N} (\cdots)
$$

$$
+ \sum_{\substack{a,b,c,d=1 \\ (b=d\neq c)}}^{N} (\cdots) = 3 \sum_{\substack{a,b,c,d=1 \\ (b=c\neq d)}}^{N} (\cdots), \quad W_{B,4} = \sum_{\substack{a,b,c,d=1 \\ (b\neq c\neq d)}}^{N} (\cdots).
$$

Here we have used the obvious fact that the three summations that occur in the three-body part $W_{B,3}$ are equal to each other [see Eq. (15)]. Analogously, Eq. (16) for the energy W_A can be decomposed into particular contributions corresponding to the two-, three-, and four-body interactions:

$$
W_A = \sum_{a,b,c,d=1}^{N} (\cdots) = W_{A,2} + W_{A,3} + W_{A,4}, \quad (22)
$$

where

$$
W_{A,2} = \sum_{\substack{a,b,c,d=1 \\ (a=c,\, b=d)}}^{N} (\cdots), \quad W_{A,3} = W^{\mathrm{I}}_{A,3} + W^{\mathrm{II}}_{A,3},
$$

$$
W^{\mathrm{I}}_{A,3} = \sum_{\substack{a,b,c,d=1 \\ (a=c,\, b\neq d)}}^{N} (\cdots) + \sum_{\substack{a,b,c,d=1 \\ (a\neq c,\, b=d)}}^{N} (\cdots) = 2 \sum_{\substack{a,b,c,d=1 \\ (a=c,\, b\neq d)}}^{N} (\cdots),
$$

$$
W^{\mathrm{II}}_{A,3} = \sum_{\substack{a,b,c,d=1 \\ (a\neq c,\, a=d)}}^{N} (\cdots), \quad W_{A,4} = \sum_{\substack{a,b,c,d=1 \\ (a\neq c,d;\, b\neq d)}}^{N} (\cdots).
$$

Here we have used the fact that the two summations that appear in $W^{\mathrm{I}}_{A,3}$ are equal [see Eq. (16)]. Note that the category W_A includes two completely different three-body contributions $W^{\mathrm{I}}_{A,3}$ and $W^{\mathrm{II}}_{A,3}$. The energy W_D given by Eq. (17) is purely pairwise nonadditive and contains only three-body terms. However, the summation over a, b, and c in Eq. (18) for W_{BA} can be divided into two contributions corresponding to quantum-mechanical two- and three-body forces:

$$
W_{BA} = \sum_{a,b,c=1}^{N} (\cdots) = W_{BA,2} + W_{BA,3}, \quad (23)
$$

where

$$W_{BA,2} = \sum_{\substack{a,b,c=1\\(b=c)}}^{N} (\cdots), \quad W_{BA,3} = \sum_{\substack{a,b,c=1\\(b\neq c)}}^{N} (\cdots).$$

Finally the category W_{CD} described by Eq. (19) is purely pairwise additive.

Equations (15)–(23) form the set of the initial formulas for the considerations presented in this work. They define all possible types of the third-order interactions in a collection of N arbitrary molecules. In the next two sections we derive the closed expressions for them possessing the advantages described in the Introduction.

IV. DERIVATION OF THE GENERAL EXPRESSIONS FOR THE ENERGIES W_B, W_A, W_D, W_{BA}, AND W_{CD}

Our derivation consists of two steps: At first we modify the general equations (15)–(19) for the energies W_B, W_A, W_D, W_{BA}, and W_{CD}, extensively applying the spherical tensor formalism. This is realized in the present section. Then we make further simplifications for those parts of the general interaction energy expressions that correspond to the particular categories of the third-order long-range forces defined by Eqs. (21)–(23). These additional simplifications are demonstrated in the next section.

If we reduce the products of the matrix elements $D^{l_i}_{m_i n_i}(\omega_i^{-1})$, which occur in Eqs. (15)–(19), by the well-known formula for the Kronecker product of two irreducible representations of the SO(3) group given by[48]

$$D^{j_1}_{\mu_1\mu_1'}(\omega)D^{j_2}_{\mu_2\mu_2'}(\omega) = \sum_{j\mu\mu'} \langle j_1\mu_1, j_2\mu_2 | j\mu\rangle \langle j_1\mu_1', j_2\mu_2' | j\mu'\rangle D^{j}_{\mu\mu'}(\omega), \tag{24}$$

then expand the products of the two spherical harmonics $Y^{-\eta_{ab}'}_{l_a'+l_b'}(\hat{R}_{ab})Y^{-\eta_{ab}''}_{l_a''+l_b''}(\hat{R}_{ab})$, which appear in Eq. (18) in $C^{n_a'n_b'}_{l_a'l_b'}C^{n_a''n_b''}_{l_a''l_b''}$, and the products of the three spherical harmonics $Y^{-\eta_{ab}'}_{l_a'+l_b'}(\hat{R}_{ab})Y^{-\eta_{ab}''}_{l_a''+l_b''}(\hat{R}_{ab})Y^{-\eta_{ab}'''}_{l_a'''+l_b'''}(\hat{R}_{ab})$, which appear in Eq. (19) in $C^{n_a'n_b'}_{l_a'l_b'}C^{n_a''n_b''}_{l_a''l_b''}C^{n_a'''n_b'''}_{l_a'''l_b'''}$, according to the Gaunt formula[48]

$$Y^{\mu_1}_{j_1}(\hat{R})Y^{\mu_2}_{j_2}(\hat{R}) = \sum_{j\mu} (4\pi)^{-1/2}[j_1,j_2][j]^{-1}\langle j_1\mu_1, j_2\mu_2 | j\mu\rangle \langle j_1 0, j_2 0 | j0\rangle Y^{\mu}_{j}(\hat{R}), \tag{25}$$

and finally rewrite the resulting formulas in terms of the irreducible tensors describing the electrical properties of the interacting molecules in body-fixed frames, discussed in Sec. II, we get the following expressions for W_B, W_A, W_D, W_{BA}, and W_{CD}:

$$\begin{aligned}
W_B = {} & \frac{1}{3}(4\pi)^{3/2} \sum_{a,b,c,d=1}^{N} \sum_{\substack{l_a'l_a''l_a'''l_aL_al_bl_cl_d\\ N_an_bn_cn_d,\eta_{ab}\eta_{ac}\eta_{ad}}} (-1)^{l_b+l_c+l_d} R_{ab}^{-l_a'-l_b-1} R_{ac}^{-l_a''-l_c-1} R_{ad}^{-l_a'''-l_d-1} \\
& \times \binom{2l_a'+2l_b}{2l_b}^{1/2} \binom{2l_a''+2l_c}{2l_c}^{1/2} \binom{2l_a'''+2l_d}{2l_d}^{1/2} \\
& \times [l_a'+l_b, l_a''+l_c, l_a'''+l_d]^{-1} Y^{-\eta_{ab}}_{l_a'+l_b}(\hat{R}_{ab}) Y^{-\eta_{ac}}_{l_a''+l_c}(\hat{R}_{ac}) Y^{-\eta_{ad}}_{l_a'''+l_d}(\hat{R}_{ad}) \\
& \times \left[\tilde{\boldsymbol{\beta}}_{L_a}\{(l_a'l_a'')l_a, l_a'''\}\cdot \mathbf{D}^{L_a}(\omega_a^{-1})\right]^{N_a} \left[\mathbf{Q}_{l_b}\cdot \mathbf{D}^{l_b}(\omega_b^{-1})\right]^{n_b} \left[\mathbf{Q}_{l_c}\cdot \mathbf{D}^{l_c}(\omega_c^{-1})\right]^{n_c} \left[\mathbf{Q}_{l_d}\cdot \mathbf{D}^{l_d}(\omega_d^{-1})\right]^{n_d} \\
& \times C_B(l_a', l_a'', l_a''', l_a; L_aN_a, l_bn_b, l_cn_c, l_dn_d; \eta_{ab}, \eta_{ac}, \eta_{ad}),
\end{aligned} \tag{26}$$

where

$$\begin{aligned}
& C_B(l_a', l_a'', l_a''', l_a; L_aN_a, l_bn_b, l_cn_c, l_dn_d; \eta_{ab}, \eta_{ac}, \eta_{ad}) \\
& \quad = (-1)^{\eta_{ab}+\eta_{ac}+\eta_{ad}} \sum_{n_an_a'n_a''n_a'''} \langle l_a'n_a', l_a''n_a'' | l_an_a\rangle \langle l_an_a, l_a'''n_a''' | L_aN_a\rangle \langle l_a'n_a', l_bn_b | l_a'+l_b\,\eta_{ab}\rangle \\
& \qquad \times \langle l_a''n_a'', l_cn_c | l_a''+l_c\,\eta_{ac}\rangle \langle l_a'''n_a''', l_dn_d | l_a'''+l_d\,\eta_{ad}\rangle,
\end{aligned}$$

$$\begin{aligned}
W_A = {} & 2(4\pi)^{3/2} \sum_{a,b,c,d=1}^{N} \sum_{\substack{l_al_b'l_b''l_bl_c'l_c''l_cl_d\\ n_an_bn_cn_d,\eta_{ab}\eta_{bc}\eta_{cd}}} (-1)^{l_b'+l_c'+l_d} R_{ab}^{-l_a-l_b'-1} R_{bc}^{-l_b''-l_c'-1} R_{cd}^{-l_c''-l_d-1} \\
& \times \binom{2l_a+2l_b'}{2l_a}^{1/2} \binom{2l_b''+2l_c'}{2l_b''}^{1/2} \binom{2l_c''+2l_d}{2l_d}^{1/2} \\
& \times [l_a+l_b', l_b''+l_c', l_c''+l_d]^{-1} Y^{-\eta_{ab}}_{l_a+l_b'}(\hat{R}_{ab}) Y^{-\eta_{bc}}_{l_b''+l_c'}(\hat{R}_{bc}) Y^{-\eta_{cd}}_{l_c''+l_d}(\hat{R}_{cd}) \\
& \times \left[\mathbf{Q}_{l_a}\cdot \mathbf{D}^{l_a}(\omega_a^{-1})\right]^{n_a} \left[\boldsymbol{\alpha}_{l_b}\{l_b'l_b''\}\cdot \mathbf{D}^{l_b}(\omega_b^{-1})\right]^{n_b} \left[\boldsymbol{\alpha}_{l_c}\{l_c'l_c''\}\cdot \mathbf{D}^{l_c}(\omega_c^{-1})\right]^{n_c} \left[\mathbf{Q}_{l_d}\cdot \mathbf{D}^{l_d}(\omega_d^{-1})\right]^{n_d} \\
& \times C_A(l_b', l_b'', l_c', l_c''; l_an_a, l_bn_b, l_cn_c, l_dn_d; \eta_{ab}, \eta_{bc}, \eta_{cd}),
\end{aligned} \tag{27}$$

where

$$C_A(l'_b,l''_b,l'_c,l''_c;l_an_a,l_bn_b,l_cn_c,l_dn_d;\eta_{ab},\eta_{bc},\eta_{cd})$$

$$=(-1)^{\eta_{ab}+\eta_{bc}+\eta_{cd}}\sum_{n'_bn''_bn'_cn''_c}\langle l'_bn'_b,l''_bn''_b|l_bn_b\rangle\langle l'_cn'_c,l''_cn''_c|l_cn_c\rangle\langle l_an_a,l'_bn'_b|l_a+l'_b\eta_{ab}\rangle$$

$$\times\langle l''_bn''_b,l'_cn'_c|l''_b+l'_c\eta_{bc}\rangle\langle l''_cn''_c,l_dn_d|l''_c+l_d\eta_{cd}\rangle,$$

$$W_D=\frac{2}{3}(4\pi)^{3/2}\sum_{a,b,c=1}^{N}\sum_{\substack{l'_al''_al_al'_bl''_bl_bl'_cl''_cl_c\\ n_an_bn_c,\eta_{ab}\eta_{bc}\eta_{ca}}}(-1)^{l'_a+l'_b+l'_c}R_{ab}^{-l''_a-l'_b-1}R_{bc}^{-l''_b-l'_c-1}R_{ca}^{-l''_c-l'_a-1}$$

$$\times\binom{2l''_a+2l'_b}{2l''_a}^{1/2}\binom{2l''_b+2l'_c}{2l''_b}^{1/2}\binom{2l''_c+2l'_a}{2l''_c}^{1/2}$$

$$\times[l''_a+l'_b,l''_b+l'_c,l''_c+l'_a]^{-1}Y_{l''_a+l'_b}^{-\eta_{ab}}(\widehat{R}_{ab})Y_{l''_b+l'_c}^{-\eta_{bc}}(\widehat{R}_{bc})Y_{l''_c+l'_a}^{-\eta_{ca}}(\widehat{R}_{ca})$$

$$\times C_D(l'_a,l''_a,l'_b,l''_b,l'_c,l''_c;l_an_a,l_bn_b,l_cn_c;\eta_{ab},\eta_{bc},\eta_{ca})$$

$$\times\sum_{m'_am''_am_am'_bm''_bm_bm'_cm''_cm_c}{}^{D}t^{\,m'_am''_am'_bm''_bm'_cm''_c}_{\,l'_al''_al'_bl''_bl'_cl''_c}\langle l'_am'_a,l''_am''_a|l_am_a\rangle D^{l_a}_{m_an_a}(\omega_a^{-1})$$

$$\times\langle l'_bm'_b,l''_bm''_b|l_bm_b\rangle D^{l_b}_{m_bn_b}(\omega_b^{-1})\langle l'_cm'_c,l''_cm''_c|l_cm_c\rangle D^{l_c}_{m_cn_c}(\omega_c^{-1}), \qquad (28)$$

where

$$C_D(l'_a,l''_a,l'_b,l''_b,l'_c,l''_c;l_an_a,l_bn_b,l_cn_c;\eta_{ab},\eta_{bc},\eta_{ca})$$

$$=(-1)^{\eta_{ab}+\eta_{bc}+\eta_{ca}}\sum_{n'_an''_an'_bn''_bn'_cn''_c}\langle l'_an'_a,l''_an''_a|l_an_a\rangle\langle l'_bn'_b,l''_bn''_b|l_bn_b\rangle\langle l'_cn'_c,l''_cn''_c|l_cn_c\rangle$$

$$\times\langle l''_an''_a,l'_bn'_b|l''_a+l'_b\eta_{ab}\rangle\langle l''_bn''_b,l'_cn'_c|l''_b+l'_c\eta_{bc}\rangle\langle l''_cn''_c,l'_an'_a|l''_c+l'_a\eta_{ca}\rangle,$$

$$W_{BA}=4\pi\sum_{a,b,c=1}^{N}\sum_{\substack{l'_al''_al'''_al_aL_al'_bl''_bl_bl_c\\ l_{ab}n_{ab}\eta_{ac},N_an_bn_c}}(-1)^{l'_b+l''_b+l_c}R_{ab}^{-l'_a-l'_b-l''_a-l''_b-2}R_{ac}^{-l'''_a-l_c-1}$$

$$\times\binom{2l'_a+2l'_b}{2l'_a}^{1/2}\binom{2l''_a+2l''_b}{2l''_a}^{1/2}\binom{2l'''_a+2l_c}{2l'''_a}^{1/2}$$

$$\times[l_{ab},l'''_a+l_c]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle Y_{l_{ab}}^{-n_{ab}}(\widehat{R}_{ab})Y_{l'''_a+l_c}^{-\eta_{ac}}(\widehat{R}_{ac})$$

$$\times C_{BA}(l'_a,l''_a,l'''_a,l'_b,l''_b,l_a;L_aN_a,l_bn_b,l_cn_c;l_{ab}-n_{ab},\eta_{ac})$$

$$\times\sum_{m'_am''_am'''_am_aM_am'_bm''_bm_bm_c}{}^{BA}t^{\,m'_am''_am'''_am'_bm''_bm_c}_{\,l'_al''_al'''_al'_bl''_bl_c}\langle l'_am'_a,l''_am''_a|l_am_a\rangle\langle l_am_a,l'''_am'''_a|L_aM_a\rangle$$

$$\times D^{L_a}_{M_aN_a}(\omega_a^{-1})\langle l'_bm'_b,l''_bm''_b|l_bm_b\rangle D^{l_b}_{m_bn_b}(\omega_b^{-1})D^{l_c}_{m_cn_c}(\omega_c^{-1}), \qquad (29)$$

where

$$C_{BA}(l'_a,l''_a,l'''_a,l'_bl''_b,l_a;L_aN_a,l_bn_b,l_cn_c;l_{ab}-n_{ab},\eta_{ac})$$

$$=\sum_{n'_an''_an'''_an'_bn''_b,\eta'_{ab}\eta''_{ab}}(-1)^{\eta'_{ab}+\eta''_{ab}+\eta_{ac}}\langle l'_an'_a,l''_an''_a|l_an_a\rangle\langle l_an_a,l'''_an'''_a|L_aN_a\rangle\langle l'_bn'_b,l''_bn''_b|l_bn_b\rangle\langle l'_an'_a,l'_bn'_b|l'_a+l'_b\eta'_{ab}\rangle$$

$$\times\langle l''_an''_a,l''_bn''_b|l''_a+l''_b\eta''_{ab}\rangle\langle l'''_an'''_a,l_cn_c|l'''_a+l_c\eta_{ac}\rangle\langle l'_a+l'_b-\eta'_{ab},l''_a+l''_b-\eta''_{ab}|l_{ab}-n_{ab}\rangle,$$

$$W_{CD}=\frac{1}{2}(4\pi)^{1/2}\sum_{a,b=1}^{N}\sum_{l'_al''_al'''_al_aL_al'_bl''_bl'''_bl_bL_bl_{ab}L_{ab}N_aN_bN_{ab}}(-1)^{l'_b+l''_b+l'''_b}R_{ab}^{-l'_a-l'_b-l''_a-l''_b-l'''_a-l'''_b-3}$$

$$\times\binom{2l'_a+2l'_b}{2l'_a}^{1/2}\binom{2l''_a+2l''_b}{2l''_a}^{1/2}\binom{2l'''_a+2l'''_b}{2l'''_a}^{1/2}[L_{ab}]^{-1}$$

$$\times\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle Y_{L_{ab}}^{-N_{ab}}(\widehat{R}_{ab})$$

$$\times C_{CD}(l'_a,l''_a,l'''_a,l'_b,l''_b,l'''_b,l_a,l_b,l_{ab};L_aN_a,L_bN_b;L_{ab}-N_{ab})$$

$$\times\sum_{m'_am''_am'''_am_aM_am'_bm''_bm'''_bm_bM_b}{}^{CD}t^{\,m'_am''_am'''_am'_bm''_bm'''_b}_{\,l'_al''_al'''_al'_bl''_bl'''_b}\langle l'_am'_a,l''_am''_a|l_am_a\rangle\langle l_am_a,l'''_am'''_a|L_aM_a\rangle$$

$$\times D^{L_a}_{M_aN_a}(\omega_a^{-1})\langle l'_bm'_b,l''_bm''_b|l_bm_b\rangle\langle l_bm_b,l'''_bm'''_b|L_bM_b\rangle D^{L_b}_{M_bN_b}(\omega_b^{-1}), \qquad (30)$$

where

$$C_{CD}(l'_a,l''_a,l'''_a,l'_b,l''_b,l'''_b,l_a,l_b,l_{ab};L_aN_a,L_bN_b;L_{ab}-N_{ab})$$
$$=\sum_{n'_an''_an'''_an'_bn''_bn'''_bn_an_b\eta'_{ab}\eta''_{ab}\eta'''_{ab}n_{ab}}(-1)^{\eta'_{ab}+\eta''_{ab}+\eta'''_{ab}}\langle l'_an'_a,l''_an''_a|l_an_a\rangle$$
$$\times\langle l_an_a,l'''_an'''_a|L_aN_a\rangle\langle l'_bn'_b,l''_bn''_b|l_bn_b\rangle\langle l_bn_b,l'''_bn'''_b|L_bN_b\rangle$$
$$\times\langle l'_an'_a,l'_bn'_b|l'_a+l'_b\eta'_{ab}\rangle\langle l''_an''_a,l''_bn''_b|l''_a+l''_b\eta''_{ab}\rangle\langle l'''_an'''_a,l'''_bn'''_b|l'''_a+l'''_b\eta'''_{ab}\rangle$$
$$\times\langle l'_a+l'_b-\eta'_{ab},l''_a+l''_b-\eta''_{ab}|l_{ab}-n_{ab}\rangle\langle l_{ab}-n_{ab},l'''_a+l'''_b-\eta'''_{ab}|L_{ab}-N_{ab}\rangle.$$

It is seen that all the above operations lead to the simplification of the orientational dependence of the expressions for W_B, W_A, W_D, W_{BA}, and W_{CD}. However, at the same time they introduce any number of Clebsch–Gordan coefficients, which are collected in the quantities $C_B(\cdots)$, $C_A(\cdots)$, $C_D(\cdots)$, $C_{BA}(\cdots)$, and $C_{CD}(\cdots)$, and which define the respective coupling schemes. Now we replace these coupling schemes by the other ones, which are physically appealing and allow us to express each of the equations for W_B, W_A, W_D, W_{BA}, and W_{CD} as an interaction between irreducible tensors localized on interacting molecules. We realize this modification of the quantities $C_X(\cdots)$ $(X=B,A,D,BA,CD)$ by applying very elegant and powerful graphical methods of the quantum theory of angular momentum described by Yutsis *et al.*[49] (see also Ref. 50).

A. $C_B(\cdots)$

Replacing the Clebsch-Gordan coefficients occurring in $C_B(\cdots)$ by the Wigner 3-j symbols[51] and making use of the fact that in $C_B(\cdots)$ only the terms with $n'_a+n''_a=n_a$, $n_a+n'''_a=N_a$ are nonzero, we easily find that[49,50]

$$C_B(\cdots)=(-1)^{l'_a+l''_a+l'''_a+l_b+l_c+l_d}[l_a,L_a,l'_a+l_b,l''_a+l_c,l'''_a+l_d]\sum_{n_an'_an''_an'''_a}(-1)^{l_a-n_a+l'_a-n'_a+l''_a-n''_a+l'''_a-n'''_a}$$
$$\times\begin{pmatrix}l'_a & l''_a & l_a\\ n'_a & n''_a & -n_a\end{pmatrix}\begin{pmatrix}l_a & l'''_a & L_a\\ n_a & n'''_a & -N_a\end{pmatrix}\begin{pmatrix}l'_a & l_b & l'_a+l_b\\ -n'_a & -n_b & \eta_{ab}\end{pmatrix}$$
$$\times\begin{pmatrix}l''_a & l_c & l''_a+l_c\\ -n''_a & -n_c & \eta_{ac}\end{pmatrix}\begin{pmatrix}l'''_a & l_d & l'''_a+l_d\\ -n'''_a & -n_d & \eta_{ad}\end{pmatrix}$$
$$=(-1)^{l'_a+l''_a+l'''_a+l_b+l_c+l_d}[l_a,L_a,l'_a+l_b,l''_a+l_c,l'''_a+l_d]C'_B,\tag{31}$$

where C'_B is represented by Fig. 1(a). Here

$$\begin{pmatrix}j_1 & j_2 & j_3\\ \mu_1 & \mu_2 & \mu_3\end{pmatrix}$$

denotes the Wigner 3-j symbol. The jm-coefficient C'_B can be decomposed and rewritten in terms of generalized Wigner coefficients[49,50,52] in the following way[49,50]:

$$C'_B=\sum_{l_{bc}l_{bcd}l_{abc}L_{abcd}}R_BV_B[l_{bc},l_{bcd},l_{abc},L_{abcd}]^2,\tag{32}$$

where the generalized Wigner coefficient V_B is represented

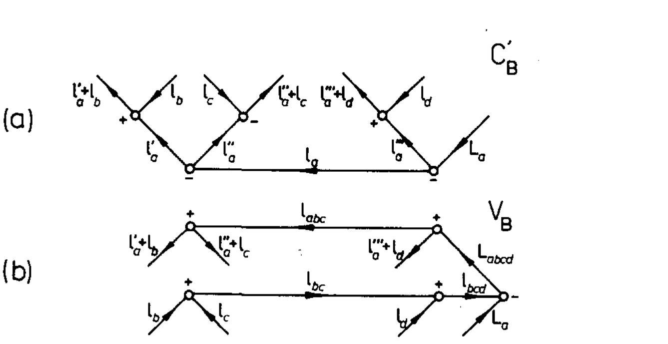


FIG. 1. Graphic representation of (a) C'_B, (b) V_B.

by the diagram shown in Fig. 1(b), while the j-coefficient R_B has a graphic representation illustrated in Fig. 2(a). Let us recall that according to general rules of diagrammatic approach[49,50] one obtains the diagram R_B joining the lines L_a, $l_b,l_c,l_d,l'_a+l_b,l''_a+l_c,l'''_a+l_d$ of the diagrams C'_B and V_B and leaving their directions as they were in diagram C'_B. We see that if we change the direction of the line l_{abc} in the diagram R_B, this diagram can be divided into two parts by the lines l_{abc},l_{bc},l_a. The change of direction of a given line j introduces the phase factor $(-1)^{2j}$ (see Refs. 49 and 50); in our case all parameters j are natural, so the change of direction of the line l_{abc} in R_B does not introduce any phase factor. The fact that the set of the three lines l_{abc},l_{bc},l_a divides R_B into two parts with a smaller number of parameters j, i.e., X_1 and X_2 shown in Fig. 2(b), means that[49,50]

$$R_B=X_1\cdot X_2.\tag{33}$$

The diagrams X_1 and X_2 can be redrawn in the forms presented in Fig. 2(c). Now let us compare these new shapes of X_1 and X_2 with the graphic representation of the Wigner 9-j symbol illustrated in Fig. 2(d). We see that we must change the directions of some lines in X_1 and X_2 to express them in terms of the 9-j symbols. But as we have explained earlier such changes do not introduce any additional phase factors in our case. So

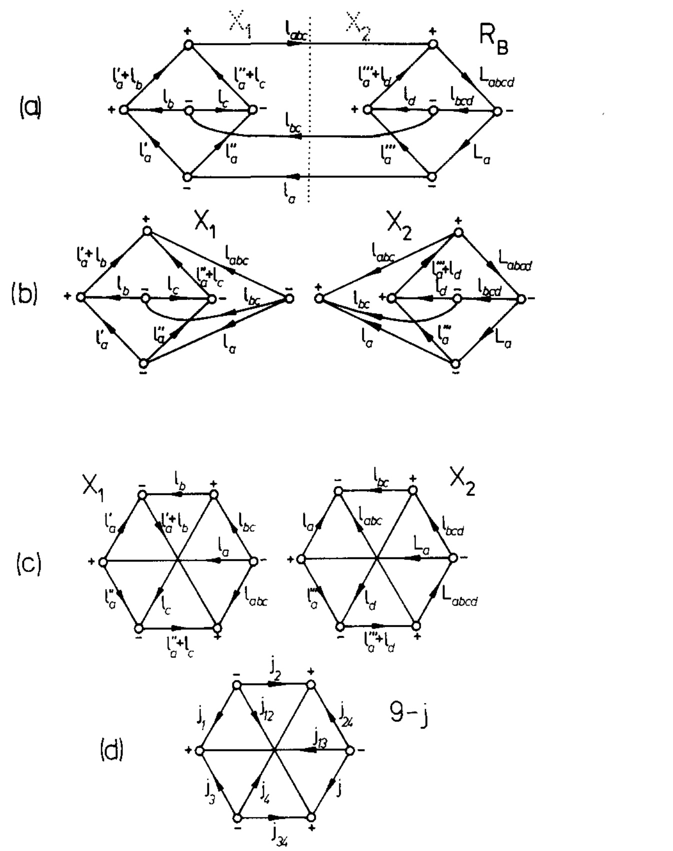


FIG. 2. (a) Graphic representation of R_B. The dotted line denotes how this diagram divides into parts X_1 and X_2. (b) Graphic representations of X_1 and X_2. The forms are obtained directly from the diagram illustrated in Fig. 2(a). (c) Redrawn forms of the diagrams X_1 and X_2. (d) Graphic representation of the Wigner 9-j symbol

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & j_4 & j_{34} \\ j_{13} & j_{24} & j \end{Bmatrix}.$$

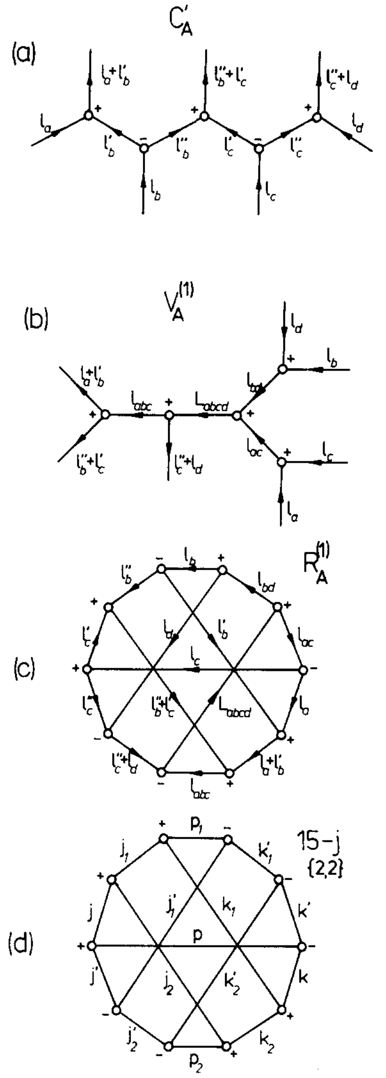


FIG. 3. (a) Graphic representation of C'_A. (b) Graphic representation of $V_A^{(1)}$. (c) Graphic representation of $R_A^{(1)}$. (d) Graphic representation of the 15-j symbol of type {2,2},

$$\left\langle \begin{matrix} j_1 & & j'_1 & j & & j' & j_2 & & j'_2 \\ & p_1 & & & p & & & p_2 & \\ k_1 & & k'_1 & k & & k' & k_2 & & k'_2 \end{matrix} \right\rangle.$$

$$X_1 = \begin{Bmatrix} l'_a & l_b & l'_a + l_b \\ l''_a & l_c & l''_a + l_c \\ l_a & l_{bc} & l_{abc} \end{Bmatrix}, \quad X_2 = \begin{Bmatrix} l_a & l_{bc} & l_{abc} \\ l'''_a & l_d & l'''_a + l_d \\ L_a & l_{bcd} & L_{abcd} \end{Bmatrix}, \tag{34}$$

where the expressions between curly brackets are the Wigner 9-j symbols.[51] Of course, V_B can be written as[49,50]

$$\begin{aligned} V_B = & \sum_{n_{bc} n_{bcd} n_{abc} N_{abcd}} (-1)^{l_{abc} - n_{abc} + l_{bc} - n_{bc} + l_{bcd} - n_{bcd} + L_{abcd} - N_{abcd}} \begin{pmatrix} l'_a + l_b & l''_a + l_c & l_{abc} \\ \eta_{ab} & \eta_{ac} & -n_{abc} \end{pmatrix} \begin{pmatrix} l_{abc} & l'''_a + l_d & L_{abcd} \\ n_{abc} & \eta_{ad} & -N_{abcd} \end{pmatrix} \\ & \times \begin{pmatrix} l_b & l_c & l_{bc} \\ -n_b & -n_c & -n_{bc} \end{pmatrix} \begin{pmatrix} l_{bc} & l_d & l_{bcd} \\ -n_{bc} & -n_d & n_{bcd} \end{pmatrix} \begin{pmatrix} L_a & l_{bcd} & l_{bcd} \\ -N_a & -n_{bcd} & N_{abcd} \end{pmatrix} \\ = & \sum_{n_{abc} n_{bc} n_{bcd} N_{abcd}} (-1)^{L_{abcd} - N_{abcd}} [l_{abc}, L_{abcd}, l_{bc}, l_{bcd}, L_{abcd}]^{-1} \langle l'_a + l_b - \eta_{ab}, l''_a + l_c - \eta_{ac} | l_{abc} - n_{abc} \rangle \\ & \times \langle l_{abc} - n_{abc}, l'''_a + l_d - \eta_{ad} | L_{abcd} - N_{abcd} \rangle \langle l_b n_b, l_c n_c | l_{bc} n_{bc} \rangle \\ & \times \langle l_{bc} n_{bc}, l_d n_d | l_{bcd} n_{bcd} \rangle \langle L_a N_a, l_{bcd} n_{bcd} | L_{abcd} N_{abcd} \rangle, \end{aligned} \tag{35}$$

because[48]

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ -\mu_1 & -\mu_2 & -\mu_3 \end{pmatrix} = (-1)^{j_1+j_2+j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix}.$$

From Eqs. (31)–(35) we get the following modified formula for $C_B(\cdots)$:

$$C_B(\cdots) = (-1)^{l'_a + l''_a + l'''_a + l_b + l_c + l_d} \sum_{\substack{l_{abc} l_{bc} l_{bcd} L_{abcd} \\ n_{abc} n_{bc} n_{bcd} N_{abcd}}} (-1)^{L_{abcd} - N_{abcd}} [l'_a + l_b, l''_a + l_c, l'''_a + l_d, l_a, L_a, l_{bc}, l_{bcd}, l_{abc}]$$
$$\times \begin{Bmatrix} l'_a & l_b & l'_a + l_b \\ l''_a & l_c & l''_a + l_c \\ l_a & l_{bc} & l_{abc} \end{Bmatrix} \begin{Bmatrix} l_a & l_{bc} & l_{abc} \\ l'''_a & l_d & l'''_a + l_d \\ L_a & l_{bcd} & L_{abcd} \end{Bmatrix} \langle l'_a + l_b - \eta_{ab}, l''_a + l_c - \eta_{ac} | l_{abc} - n_{abc} \rangle$$
$$\times \langle l_{abc} - n_{abc}, l'''_a + l_d - \eta_{ad} | L_{abcd} - N_{abcd} \rangle \langle l_b n_b, l_c n_c | l_{bc} n_{bc} \rangle$$
$$\times \langle l_{bc} n_{bc}, l_d n_d | l_{bcd} n_{bcd} \rangle \langle L_a N_a, l_{bcd} n_{bcd} | L_{abcd} N_{abcd} \rangle. \tag{36}$$

B. $C_A(\cdots)$

Similarly as in the case of $C_B(\cdots)$, after replacing the Clebsch–Gordan coefficients occurring in Eq. (27) for $C_A(\cdots)$ by the 3-j symbols and making use of the fact that in $C_A(\cdots)$ only the terms with $n'_b + n''_b = n_b$, $n'_c + n''_c = n_c$ are nonvanishing, one obtains[49,50]

$$C_A(\cdots) = (-1)^{l_a + l'_b + l''_b + l'_c + l''_c + l_d} [l_b, l_c, l_a + l'_b, l''_b + l'_c, l''_c + l_d] C'_A, \tag{37}$$

where the jm-coefficient C'_A is represented by Fig. 3(a). Again C'_A can be decomposed and rewritten in terms of generalized Wigner coefficients.[49,50,52] In particular,[49,50]

$$C'_A = \sum_{l_{ac} l_{bd} l_{abc} L_{abcd}} R_A^{(1)} V_A^{(1)} [l_{ac}, l_{bd}, l_{abc}, L_{abcd}]^2, \tag{38}$$

where the generalized Wigner coefficient $V_A^{(1)}$ is represented by the diagram shown in Fig. 3(b), while the j-coefficient $R_A^{(1)}$ has a graphic representation illustrated in Fig. 3(c). In Fig. 3(d) we have given a graphic representation of the 15-j symbol of type {2,2} described by Yutsis and Bandzaitis,[53] which is proportional to the 15-j symbol of the third kind considered by Yutsis *et al.*[54] [in Fig. 3(d) we have left arrows that denote the directions of lines, because, as we have explained in Sec. IV A, they are immaterial in the case of a j-coefficient with natural j's]. If we now compare the diagrams presented in Figs. 3(c) and 3(d) we see that $R_A^{(1)}$ can be expressed in terms of the 15-j symbol. To do this we must only change the orientations of the loops (l'_b, l''_b, l_b), (l_b, l_d, l_{bd}) and $(l_{ac}, l_{bd}, L_{abcd})$ in the diagram $R_A^{(1)}$. This means that[49,50]

$$R_A^{(1)} = \left\langle \begin{matrix} l''_b & & l_d & l'_c & & l''_c & l''_b + l'_c & & l''_c + l_d \\ & l_b & & & l_c & & & l_{abc} & \\ l'_b & & l_{bd} & l_a & & l_{ac} & l_a + l'_b & & L_{abcd} \end{matrix} \right\rangle (-1)^{l'_b + l''_b + l_d + l_{ac} + L_{abcd}}, \tag{39}$$

where the expression between brackets $\langle\quad\rangle$ is the 15-j symbol of type {2,2}. One can easily check that $V_A^{(1)}$ can be put into the following form[48–50]:

$$V_A^{(1)} = (-1)^{l_a + l_c + l_{ac} + L_{abcd}} [l_{abc}, L_{abcd}, l_{ac}, l_{bd}, L_{abcd}]^{-1} \sum_{n_{ac} n_{bd} n_{abc} N_{abcd}} (-1)^{N_{abcd}} \langle l_a + l'_b - \eta_{ab}, l''_b + l'_c - \eta_{bc} | l_{abc} - n_{abc} \rangle$$
$$\times \langle l_{abc} - n_{abc}, l''_c + l_d - \eta_{cd} | L_{abcd} - N_{abcd} \rangle \langle l_c n_c, l_a n_a | l_{ac} n_{ac} \rangle \langle l_b n_b, l_d n_d | l_{bd} n_{bd} \rangle \langle l_{ac} n_{ac}, l_{bd} n_{bd} | L_{abcd} N_{abcd} \rangle. \tag{40}$$

Therefore from Eqs. (37)–(40) we get the following modified formula for $C_A(\cdots)$:

$$C_A(\cdots) = (-1)^{l'_c + l''_c + l_c} [l_a + l'_b, l''_b + l'_c, l''_c + l_d, l_b, l_c] \sum_{\substack{l_{ac} l_{bd} l_{abc} L_{abcd} \\ n_{ac} n_{bd} n_{abc} N_{abcd}}} (-1)^{N_{abcd}} [l_{ac}, l_{bd}, l_{abc}]$$
$$\times \left\langle \begin{matrix} l''_b & & l_d & l'_c & & l''_c & l''_b + l'_c & & l''_c + l_d \\ & l_b & & & l_c & & & l_{abc} & \\ l'_b & & l_{bd} & l_a & & l_{ac} & l_a + l'_b & & L_{abcd} \end{matrix} \right\rangle$$
$$\times \langle l_a + l'_b - \eta_{ab}, l''_b + l'_c - \eta_{ab} | l_{abc} - n_{abc} \rangle \langle l_{abc} - n_{abc}, l''_c + l_d - \eta_{cd} | L_{abcd} - N_{abcd} \rangle$$
$$\times \langle l_c n_c, l_a n_a | l_{ac} n_{ac} \rangle \langle l_b n_b, l_d n_d | l_{bd} n_{bd} \rangle \langle l_{ac} n_{ac}, l_{bd} n_{bd} | L_{abcd} N_{abcd} \rangle. \tag{41}$$

However, we can decompose and rewrite C'_A in terms of generalized Wigner coefficients in another way, which also will be

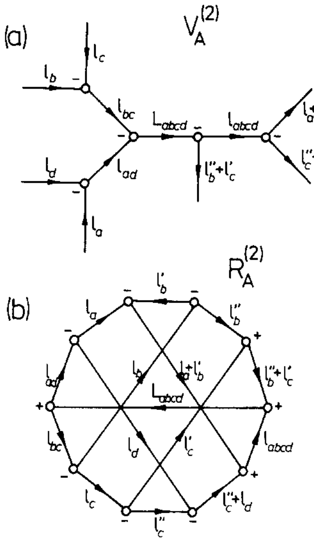


FIG. 4. Graphic representation of (a) $V_A^{(2)}$, (b) $R_A^{(2)}$.

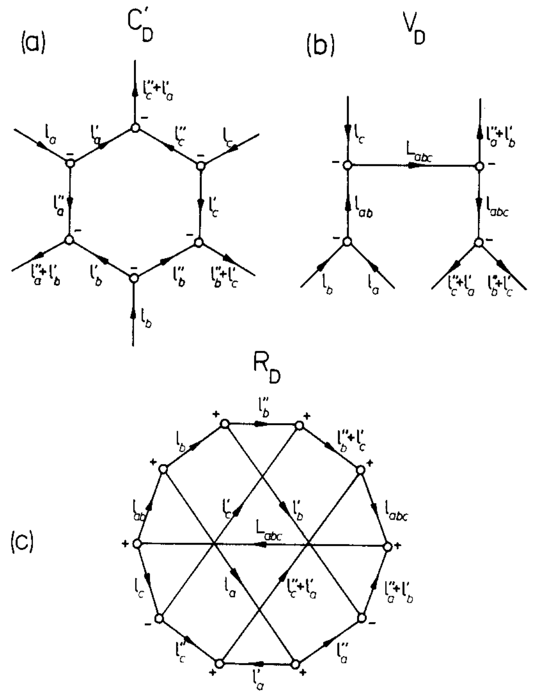


FIG. 5. Graphic representation of (a) C'_D, (b) V_D, (c) R_D.

interesting for us, viz.,[49,50]

$$C'_A = \sum_{l_{ad} l_{bc} l_{abcd} L_{abcd}} R_A^{(2)} V_A^{(2)} [l_{ad}, l_{bc}, l_{abcd}, L_{abcd}]^2, \tag{42}$$

where the generalized Wigner coefficient $V_A^{(2)}$ is represented by the diagram shown in Fig. 4(a) and the j-coefficient $R_A^{(2)}$ has a graphic representation illustrated in Fig. 4(b). If we again compare Fig. 4(b) with the graphic representation of the 15-j symbol of type {2,2} given in Fig. 3(d), we find that[49,50]

$$R_A^{(2)} = \left\langle \begin{array}{cccccccccc} l_a & & l_b & l_{ad} & & l_{bc} & l_d & & l_c \\ & l'_b & & & L_{abcd} & & & l''_c & \\ l_a + l'_b & & l''_b & l_{abcd} & & l''_b + l'_c & l''_c + l_d & & l'_c \end{array} \right\rangle (-1)^{l''_b + l'_c + l_a + l_d + l_{ad} + l_{abcd} + L_{abcd}}. \tag{43}$$

It is evident that[48–50]

$$\begin{aligned} V_A^{(2)} = {} & (-1)^{l'_b + l''_c + l_{ad} + l_{abcd} + L_{abcd}} [l_{ad}, l_{bc}, L_{abcd}, l_{abcd}, L_{abcd}]^{-1} \\ & \times \sum_{n_{ad} n_{bc} n_{abcd} N_{abcd}} (-1)^{N_{abcd}} \langle l''_c + l_d - \eta_{cd}, l_a + l'_b - \eta_{ab} | l_{abcd} - n_{abcd} \rangle \\ & \times \langle l_{abcd} - n_{abcd}, l''_b + l'_c - \eta_{bc} | L_{abcd} - N_{abcd} \rangle \langle l_d n_d, l_a n_a | l_{ad} n_{ad} \rangle \langle l_b n_b, l_c n_c | l_{bc} n_{bc} \rangle \langle l_{ad} n_{ad}, l_{bc} n_{bc} | L_{abcd} N_{abcd} \rangle . \end{aligned} \tag{44}$$

From Eqs. (37) and (42)–(44) we obtain the alternative modified expression for $C_A(\cdots)$, i.e.,

$$\begin{aligned} C_A(\cdots) = {} & [l_a + l'_b, l''_b + l'_c, l''_c + l_d, l_b, l_c] \sum_{\substack{l_{ad} l_{bc} l_{abcd} L_{abcd} \\ n_{ad} n_{bc} n_{abcd} N_{abcd}}} (-1)^{N_{abcd}} [l_{ad}, l_{bc}, l_{abcd}] \\ & \times \left\langle \begin{array}{cccccccccc} l_a & & l_b & l_{ad} & & l_{bc} & l_d & & l_c \\ & l'_b & & & L_{abcd} & & & l''_c & \\ l_a + l'_b & & l''_b & l_{abcd} & & l''_b + l'_c & l''_c + l_d & & l'_c \end{array} \right\rangle \\ & \times \langle l''_c + l_d - \eta_{cd}, l_a + l'_b - \eta_{ab} | l_{abcd} - n_{abcd} \rangle \langle l_{abcd} - n_{abcd}, l''_b + l'_c - \eta_{bc} | L_{abcd} - N_{abcd} \rangle \\ & \times \langle l_d n_d, l_a n_a | l_{ad} n_{ad} \rangle \langle l_b n_b, l_c n_c | l_{bc} n_{bc} \rangle \langle l_{ad} n_{ad}, l_{bc} n_{bc} | L_{abcd} N_{abcd} \rangle . \end{aligned} \tag{45}$$

C. $C_D(\cdots)$

From the defining relation for $C_D(\cdots)$ [see Eq. (28)] we immediately have[49,50]

$$C_D(\cdots) = (-1)^{l'_a + l''_a + l'_b + l''_b + l'_c + l''_c}[l_a, l_b, l_c, l''_a + l'_b, l''_b + l'_c, l''_c + l'_a]C'_D, \tag{46}$$

where the jm-coefficient C'_D is represented by Fig. 5(a). As in the case of C'_B and C'_A we can write[49,50]

$$C'_D = \sum_{l_{ab} l_{abc} L_{abc}} R_D V_D [l_{ab}, l_{abc}, L_{abc}]^2, \tag{47}$$

where the generalized Wigner coefficient V_D is represented by the diagram shown in Fig. 5(b), while the j-coefficient R_D has a graphic representation given in Fig. 5(c). If we compare Figs. 3(d) and 5(c) we get[49,50]

$$R_D = \left\langle \begin{matrix} l_b & & l'_c & & l_{ab} & & l_c & l_a & & l''_c \\ & l''_b & & & & L_{abc} & & & l'_a & \\ l'_b & & l''_b + l'_c & & l''_a + l'_b & & l_{abc} & l''_a & & l''_c + l'_a \end{matrix} \right\rangle (-1)^{l'_a + l''_a + l'_b + l''_b + l'_c + l''_c + L_{abc}}. \tag{48}$$

It is clear that the algebraic expression corresponding to the diagram V_D given in Fig. 5(b) is[48–50]

$$V_D = [l_{ab}, L_{abc}, l_{abc}, L_{abc}]^{-1} \sum_{n_{ab} n_{abc} N_{abc}} (-1)^{L_{abc} + N_{abc}} \langle l_a n_a, l_b n_b | l_{ab} n_{ab} \rangle$$

$$\times \langle l_{ab} n_{ab}, l_c n_c | L_{abc} N_{abc} \rangle \langle l''_b + l'_c - \eta_{bc}, l''_c + l'_a - \eta_{ca} | l_{abc} - \eta_{abc} \rangle \langle l''_a + l'_b - \eta_{ab}, l_{abc} - n_{abc} | L_{abc} - N_{abc} \rangle. \tag{49}$$

From Eqs. (46)–(49) we obtain the following modified formula for $C_D(\cdots)$:

$$C_D(\cdots) = [l''_a + l'_b, l''_b + l'_c, l''_c + l'_a, l_a, l_b, l_c] \sum_{\substack{l_{ab} l_{abc} L_{abc} \\ n_{ab} n_{abc} N_{abc}}} (-1)^{N_{abc}} [l_{ab}, l_{abc}]$$

$$\times \left\langle \begin{matrix} l_b & & l'_c & & l_{ab} & & l_c & l_a & & l''_c \\ & l''_b & & & & L_{abc} & & & l'_a & \\ l'_b & & l''_b + l'_c & & l''_a + l'_b & & l_{abc} & l''_a & & l''_c + l'_a \end{matrix} \right\rangle$$

$$\times \langle l_a n_a, l_b n_b | l_{ab} n_{ab} \rangle \langle l_{ab} n_{ab}, l_c n_c | L_{abc} N_{abc} \rangle \langle l''_b + l'_c - \eta_{bc}, l''_c + l'_a - \eta_{ca} | l_{abc} - n_{abc} \rangle$$

$$\times \langle l''_a + l'_b - \eta_{ab}, l_{abc} - n_{abc} | L_{abc} - N_{abc} \rangle. \tag{50}$$

D. $C_{BA}(\cdots)$

It is easy to see that[49,50]

$$C_{BA}(\cdots) = (-1)^{l'_a + l''_a + l'''_a + l'_b + l''_b + l_c}[l_a, L_a, l_b, l''_a + l'_b, l''_a + l''_b, l'''_a + l_c, l_{ab}]C'_{BA}, \tag{51}$$

where the jm-coefficient C'_{BA} is represented by the diagram illustrated in Fig. 6(a). In language of diagrams we can write that[49,50]

$$C'_{BA} = \sum_{l_{bc} L_{abc}} R_{BA} V_{BA} [l_{bc}, L_{abc}]^2, \tag{52}$$

where the generalized Wigner coefficient V_{BA} has a graphical representation given in Fig. 6(b), while the j-coefficient R_{BA} is represented by the diagram shown in Fig. 7(a). As in the case of R_B (see Sec. IV A) after changing the direction of the line l_{ab} in the diagram R_{BA} we find that the three lines l_a, l_b, l_{ab} divide R_{BA} into two parts X_1 and X_2 with less numbers of parameters j, which are illustrated in Fig. 7(b). This means that[49,50]

$$R_{BA} = X_1 \cdot X_2. \tag{53}$$

The diagrams X_1 and X_2 can be redrawn in the new forms presented in Fig. 7(c). If we compare these new shapes of X_1 and X_2 with the diagram representing the Wigner 9-j symbol [see Fig. 2(d)], we get

$$X_1 = \begin{Bmatrix} l'_a & l''_a & l_a \\ l'_b & l''_b & l_b \\ l'_a + l'_b & l''_a + l''_b & l_{ab} \end{Bmatrix}, \quad X_2 = \begin{Bmatrix} l_a & l'''_a & L_a \\ l_b & l_c & l_{bc} \\ l_{ab} & l'''_a + l_c & L_{abc} \end{Bmatrix}. \tag{54}$$

According to the usual interpretation of the diagram V_{BA} given in Fig. 6(b) we obtain[48–50]

$$V_{BA} = (-1)^{L_{abc}}[l_{bc}, L_{abc}, L_{abc}]^{-1} \sum_{n_{bc} N_{abc}} (-1)^{N_{abc}} \langle l_b n_b, l_c n_c | l_{bc} n_{bc} \rangle$$

$$\times \langle L_a N_a, l_{bc} n_{bc} | L_{abc} N_{abc} \rangle \langle l_{ab} - n_{ab}, l'''_a + l_c - \eta_{ac} | L_{abc} - N_{abc} \rangle. \tag{55}$$

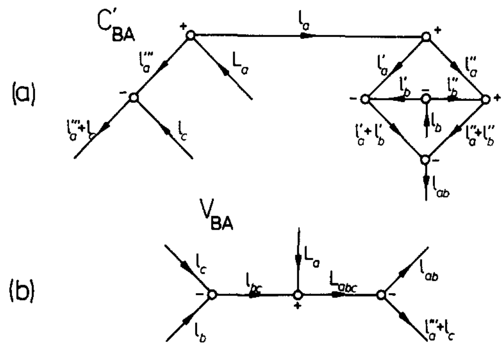


FIG. 6. Graphic representation of (a) C'_{BA}, (b) V_{BA}.

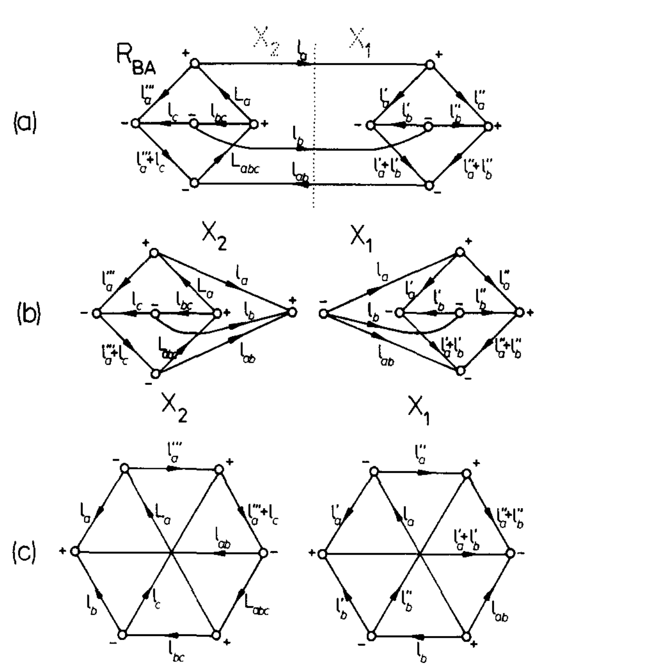


FIG. 7. (a) Graphic representation of R_{BA}. Dotted line denotes how this diagram divides into parts X_1 and X_2. (b) Graphic representations of X_1 and X_2. Forms obtained directly from the diagram illustrated in Fig. 7 (a). (c) Redrawn forms of the diagrams X_1 and X_2.

From Eqs. (51)–(55) we get the following modified expression for $C_{BA}(\cdots)$:

$$C_{BA}(\cdots) = (-1)^{l'_a+l''_a+l'''_a+l'_b+l''_b+l_c}[l'_a+l'_b, l''_a+l''_b, l'''_a+l_c, l_{ab}, l_a, L_a, l_b]$$
$$\times \sum_{\substack{l_{bc}L_{abc}\\ n_{bc}N_{abc}}} (-1)^{L_{abc}+N_{abc}}[l_{bc}] \begin{Bmatrix} l'_a & l''_a & l_a \\ l'_b & l''_b & l_b \\ l'_a+l'_b & l''_a+l''_b & l_{ab} \end{Bmatrix} \begin{Bmatrix} l_a & l'''_a & L_a \\ l_b & l_c & l_{bc} \\ l_{ab} & l'''_a+l_c & L_{abc} \end{Bmatrix}$$
$$\times \langle l_b n_b, l_c n_c | l_{bc} n_{bc}\rangle \langle L_a N_a, l_{bc} n_{bc} | L_{abc} N_{abc}\rangle \langle l_{ab} - n_{ab}, l'''_a + l_c - \eta_{ac} | L_{abc} - N_{abc}\rangle. \tag{56}$$

E. $C_{CD}(\cdots)$

Let us notice that the quantity $C_{CD}(\cdots)$ given by Eq. (30) is defined in the same way as the quantity $C(\cdots)$, which occurs in the spherical tensor theory of the third-order interactions between two molecules.[11,12] This is no wonder, because the energy W_{CD} represents purely pairwise additive interactions. Therefore we can directly apply the result of the considerations concerning the quantity $C(\cdots)$ (see Refs. 11 and 12) to $C_{CD}(\cdots)$. In this way we find the following modified formula for $C_{CD}(\cdots)$[55]:

$$C_{CD}(\cdots) = (-1)^{l'_a+l''_a+l'''_a+l'_b+l''_b+l'''_b+L_{ab}+N_{ab}}[l'_a+l'_b, l''_a+l''_b, l'''_a+l'''_b, l_a, l_b, l_{ab}, L_a, L_b]$$
$$\times \begin{Bmatrix} l'_a & l''_a & l_a \\ l'_b & l''_b & l_b \\ l'_a+l'_b & l''_a+l''_b & l_{ab} \end{Bmatrix} \begin{Bmatrix} l_a & l'''_a & L_a \\ l_b & l'''_b & L_b \\ l_{ab} & l'''_a+l'''_b & L_{ab} \end{Bmatrix} \langle L_a N_a, L_b N_b | L_{ab} N_{ab}\rangle. \tag{57}$$

Inserting the modified formulas for the quantities $C_X(\cdots)$ ($X = B,A,D,BA,CD$) obtained in Secs. IV A–IV E [see Eqs. (36), (41), (45), (50), (56), and (57)] into the respective expressions for the energies W_X ($X = B,A,D,BA,CD$) given by Eqs. (26)–(30) we obtain clear general formulas describing the third-order interactions in a collection of N arbitrary molecules. According to our program, these formulas, together with some supplementary simplifications, are discussed in the next section of the present work.

V. RESULTS

The mathematical considerations described in the previous sections yield the following final expressions for the energies W_B, W_A, W_D, W_{BA}, and W_{CD}:

$$
\begin{aligned}
W_B = \frac{1}{3}(4\pi)^{3/2} \sum_{a,b,c,d=1}^{N} \sum_{l'_a l''_a l'''_a l_a L_a l_b l_c l_d l_{bc} l_{bcd} l_{abc} L_{abcd}} & (-1)^{l'_a + l''_a + l'''_a} R_{ab}^{-l'_a - l_b - 1} R_{ac}^{-l''_a - l_c - 1} R_{ad}^{-l'''_a - l_d - 1} \\
& \times \binom{2l'_a + 2l_b}{2l_b}^{1/2} \binom{2l''_a + 2l_c}{2l_c}^{1/2} \binom{2l'''_a + 2l_d}{2l_d}^{1/2} [l_a, L_a, l_{bc}, l_{bcd}, l_{abc}] \\
& \times \begin{Bmatrix} l'_a & l_b & l'_a + l_b \\ l''_a & l_c & l''_a + l_c \\ l_a & l_{bc} & l_{abc} \end{Bmatrix} \begin{Bmatrix} l_a & l_{bc} & l_{abc} \\ l'''_a & l_d & l'''_a + l_d \\ L_a & l_{bcd} & L_{abcd} \end{Bmatrix} \\
& \times \sum_{N_{abcd} = -L_{abcd}}^{L_{abcd}} (-1)^{L_{abcd} - N_{abcd}} \left[\left[\mathbf{Y}_{l'_a + l_b}(\hat{R}_{ab}) \otimes \mathbf{Y}_{l''_a + l_c}(\hat{R}_{ac}) \right]_{l_{abc}} \otimes \mathbf{Y}_{l'''_a + l_d}(\hat{R}_{ad}) \right]_{L_{abcd}}^{-N_{abcd}} \\
& \times \left[\tilde{\boldsymbol{\beta}}_{L_a}\{(l'_a l''_a) l_a, l'''_a\} \cdot \mathbf{D}^{L_a}(\boldsymbol{\omega}_a^{-1}) \otimes \left[\left[\mathbf{Q}_{l_b} \cdot \mathbf{D}^{l_b}(\boldsymbol{\omega}_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\boldsymbol{\omega}_c^{-1}) \right]_{l_{bc}} \otimes \mathbf{Q}_{l_d} \cdot \mathbf{D}^{l_d}(\boldsymbol{\omega}_d^{-1}) \right]_{l_{bcd}} \right]_{L_{abcd}}^{N_{abcd}},
\end{aligned}
\tag{58}
$$

$$
\begin{aligned}
W_A = 2(4\pi)^{3/2} \sum_{a,b,c,d=1}^{N} \sum_{l_a l'_b l''_b l_b l'_c l''_c l_c l_d l_{ac} l_{bd} l_{abc} L_{abcd}} & (-1)^{l'_b + l''_c + l_d + l_c} R_{ab}^{-l_a - l'_b - 1} R_{bc}^{-l''_b - l'_c - 1} R_{cd}^{-l''_c - l_d - 1} \\
& \times \binom{2l_a + 2l'_b}{2l_a}^{1/2} \binom{2l''_b + 2l'_c}{2l''_b}^{1/2} \binom{2l''_c + 2l_d}{2l_d}^{1/2} [l_b, l_c, l_{ac}, l_{bd}, l_{abc}] \\
& \times \left\langle \begin{matrix} l''_b & & l_d & l'_c & & l''_c & l''_b + l'_c & & l''_c + l_d \\ & l_b & & & l_c & & & l_{abc} & \\ l'_b & & l_{bd} & l_a & & l_{ac} & l_a + l'_b & & L_{abcd} \end{matrix} \right\rangle \\
& \times \sum_{N_{abcd} = -L_{abcd}}^{L_{abcd}} (-1)^{N_{abcd}} \left[\left[\mathbf{Y}_{l_a + l'_b}(\hat{R}_{ab}) \otimes \mathbf{Y}_{l''_b + l'_c}(\hat{R}_{bc}) \right]_{l_{abc}} \otimes \mathbf{Y}_{l''_c + l_d}(\hat{R}_{cd}) \right]_{L_{abcd}}^{-N_{abcd}} \\
& \times \left[\left[\boldsymbol{\alpha}_{l_c}\{l'_c l''_c\} \cdot \mathbf{D}^{l_c}(\boldsymbol{\omega}_c^{-1}) \otimes \mathbf{Q}_{l_a} \cdot \mathbf{D}^{l_a}(\boldsymbol{\omega}_a^{-1}) \right]_{l_{ac}} \otimes \left[\boldsymbol{\alpha}_{l_b}\{l'_b l''_b\} \cdot \mathbf{D}^{l_b}(\boldsymbol{\omega}_b^{-1}) \otimes \mathbf{Q}_{l_d} \cdot \mathbf{D}^{l_d}(\boldsymbol{\omega}_d^{-1}) \right]_{l_{bd}} \right]_{L_{abcd}}^{N_{abcd}}
\end{aligned}
\tag{59a}
$$

or

$$
\begin{aligned}
W_A = 2(4\pi)^{3/2} \sum_{a,b,c,d=1}^{N} \sum_{l_a l'_b l''_b l_b l'_c l''_c l_c l_d l_{ad} l_{bc} l_{abcd} L_{abcd}} & (-1)^{l'_b + l'_c + l_d} R_{ab}^{-l_a - l'_b - 1} R_{bc}^{-l''_b - l'_c - 1} R_{cd}^{-l''_c - l_d - 1} \\
& \times \binom{2l_a + 2l'_b}{2l_a}^{1/2} \binom{2l''_b + 2l'_c}{2l''_b}^{1/2} \binom{2l''_c + 2l_d}{2l_d}^{1/2} [l_b, l_c l_{ad}, l_{bc}, l_{abcd}] \\
& \times \left\langle \begin{matrix} l_a & & l_b & l_{ad} & & l_{bc} & l_d & & l_c \\ & l'_b & & & L_{abcd} & & & l''_c & \\ l_a + l'_b & & l''_b & l_{abcd} & & l''_b + l'_c & l''_c + l_d & & l'_c \end{matrix} \right\rangle \\
& \times \sum_{N_{abcd} = -L_{abcd}}^{L_{abcd}} (-1)^{N_{abcd}} \left[\left[\mathbf{Y}_{l''_c + l_d}(\hat{R}_{cd}) \otimes \mathbf{Y}_{l_a + l'_b}(\hat{R}_{ab}) \right]_{l_{abcd}} \otimes \mathbf{Y}_{l''_b + l'_c}(\hat{R}_{bc}) \right]_{L_{abcd}}^{-N_{abcd}} \\
& \times \left[\left[\mathbf{Q}_{l_d} \cdot \mathbf{D}^{l_d}(\boldsymbol{\omega}_d^{-1}) \otimes \mathbf{Q}_{l_a} \cdot \mathbf{D}^{l_a}(\boldsymbol{\omega}_a^{-1}) \right]_{l_{ad}} \otimes \left[\boldsymbol{\alpha}_{l_b}\{l'_b l''_b\} \cdot \mathbf{D}^{l_b}(\boldsymbol{\omega}_b^{-1}) \otimes \boldsymbol{\alpha}_{l_c}\{l'_c l''_c\} \cdot \mathbf{D}^{l_c}(\boldsymbol{\omega}_c^{-1}) \right]_{l_{bc}} \right]_{L_{abcd}}^{N_{abcd}},
\end{aligned}
\tag{59b}
$$

$$
\begin{aligned}
W_D = \frac{2}{3}(4\pi)^{3/2} \sum_{a,b,c=1}^{N} \sum_{l'_a l''_a l_a l'_b l''_b l_b l'_c l''_c l_c l_{ab} l_{abc} L_{abc}} & (-1)^{l'_a + l'_b + l'_c} R_{ab}^{-l''_a - l'_b - 1} R_{bc}^{-l''_b - l'_c - 1} R_{ca}^{-l''_c - l'_a - 1} \\
& \times \binom{2l''_a + 2l'_b}{2l''_a}^{1/2} \binom{2l''_b + 2l'_c}{2l''_b}^{1/2} \binom{2l''_c + 2l'_a}{2l''_c}^{1/2} [l_a, l_b, l_c, l_{ab}, l_{abc}] \\
& \times \left\langle \begin{matrix} l_b & & l'_c & l_{ab} & & l_c & l_a & & l''_c \\ & l''_b & & & L_{abc} & & & l'_a & \\ l'_b & & l''_b + l'_c & l''_a + l'_b & & l_{abc} & l''_a & & l''_c + l'_a \end{matrix} \right\rangle \\
& \times \sum_{N_{abc} = -L_{abc}}^{L_{abc}} (-1)^{N_{abc}} \left[\mathbf{Y}_{l''_a + l'_b}(\hat{R}_{ab}) \otimes \left[\mathbf{Y}_{l''_b + l'_c}(\hat{R}_{bc}) \otimes \mathbf{Y}_{l''_c + l'_a}(\hat{R}_{ca}) \right]_{l_{abc}} \right]_{L_{abc}}^{-N_{abc}} \\
& \times {}^{D}T_{L_{abc}}^{N_{abc}}\{((l'_a l''_a) l_a, (l'_b l''_b) l_b) l_{ab}, (l'_c l''_c) l_c\},
\end{aligned}
\tag{60}
$$

where

$$
{}^{D}T^{N_{abc}}_{L_{abc}}\{((l'_a l''_a)l_a,(l'_b l''_b)l_b)l_{ab},(l'_c l''_c)l_c\}
$$

$$
= \sum_{m'_a m''_a m_a m'_b m''_b m_b m'_c m''_c m_c, n_a n_b n_{ab} n_c} {}^{D}t^{m'_a m''_a m'_b m''_b m'_c m''_c}_{l'_a l''_a l'_b l''_b l'_c l''_c} \langle l'_a m'_a, l''_a m''_a | l_a m_a\rangle D^{l_a}_{m_a n_a}(\omega_a^{-1}) \langle l'_b m'_b, l''_b m''_b | l_b m_b\rangle
$$

$$
\times D^{l_b}_{m_b n_b}(\omega_b^{-1}) \langle l'_c m'_c, l''_c m''_c | l_c m_c\rangle D^{l_c}_{m_c n_c}(\omega_c^{-1}) \langle l_a n_a, l_b n_b | l_{ab} n_{ab}\rangle \langle l_{ab} n_{ab}, l_c n_c | L_{abc} N_{abc}\rangle
$$

$$
= \sum_{p_a\neq g_a,\, p_b\neq g_b,\, p_c\neq g_c} D\Big[\Big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle \otimes \langle p_a|\hat{\mathbf{Q}}_{l''_a}|g_a\rangle\big]_{l_a} \cdot \mathbf{D}^{l_a}(\omega_a^{-1})
$$

$$
\otimes \big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1})\Big]_{l_{ab}}
$$

$$
\otimes \big[\langle g_c|\hat{\mathbf{Q}}_{l'_c}|p_c\rangle \otimes \langle p_c|\hat{\mathbf{Q}}_{l''_c}|g_c\rangle\big]_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]^{N_{abc}}_{L_{abc}},
$$

$$
W_{BA} = 4\pi \sum_{a,b,c=1}^{N} \sum_{l'_a l''_a l'''_a l_a L_a l'_b l''_b l_b l_c l_{ab} l_{bc} L_{abc}} (-1)^{l'_a+l''_a+l'''_a} R_{ab}^{-l'_a-l'_b-l''_a-l''_b-2} R_{ac}^{-l'''_a-l_c-1}
$$

$$
\times \binom{2l'_a+2l'_b}{2l'_a}^{1/2} \binom{2l''_a+2l''_b}{2l''_a}^{1/2} \binom{2l'''_a+2l_c}{2l'''_a}^{1/2}
$$

$$
\times [l_a, L_a, l_b, l_{bc}, l'_a+l'_b, l''_a+l''_b] \langle l'_a+l'_b 0, l''_a+l''_b 0 | l_{ab} 0\rangle \begin{Bmatrix} l'_a & l''_a & l_a \\ l'_b & l''_b & l_b \\ l'_a+l'_b & l''_a+l''_b & l_{ab} \end{Bmatrix} \begin{Bmatrix} l_a & l'''_a & L_a \\ l_b & l_c & l_{bc} \\ l_{ab} & l'''_a+l_c & L_{abc} \end{Bmatrix}
$$

$$
\times \sum_{N_{abc}=-L_{abc}}^{L_{abc}} (-1)^{L_{abc}-N_{abc}} \big[\mathbf{Y}_{l_{ab}}(\hat{R}_{ab}) \otimes \mathbf{Y}_{l'''_a+l_c}(\hat{R}_{ac})\big]^{-N_{abc}}_{L_{abc}} {}^{BA}T^{N_{abc}}_{L_{abc}}\{((l'_a l''_a)l_a, l'''_a)L_a, ((l'_b l''_b)l_b, l_c)l_{bc}\}, \qquad (61)
$$

where

$$
{}^{BA}T^{N_{abc}}_{L_{abc}}\{((l'_a l''_a)l_a, l'''_a)L_a, ((l'_b l''_b)l_b, l_c)l_{bc}\}
$$

$$
= \sum_{m'_a m''_a m'''_a m_a M_a m'_b m''_b m_b m_c, n_b n_c n_{bc} N_a} {}^{BA}t^{m'_a m''_a m'''_a m'_b m''_b m_c}_{l'_a l''_a l'''_a l'_b l''_b l_c} \langle l'_a m'_a, l''_a m''_a | l_a m_a\rangle \langle l_a m_a, l'''_a m'''_a | L_a M_a\rangle
$$

$$
\times D^{L_a}_{M_a N_a}(\omega_a^{-1}) \langle l'_b m'_b, l''_b m''_b | l_b m_b\rangle D^{l_b}_{m_b n_b}(\omega_b^{-1}) D^{l_c}_{m_c n_c}(\omega_c^{-1}) \langle L_a N_a, l_{bc} n_{bc} | L_{abc} N_{abc}\rangle \langle l_b n_b, l_c n_c | l_{bc} n_{bc}\rangle
$$

$$
= \sum_{\substack{p_a, q_a\neq g_a \\ p_b\neq g_b}} [\epsilon(p_a)+\epsilon(p_b)]^{-1} \epsilon(q_a)^{-1} \Big[\Big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle \otimes \langle p_a|\hat{\mathbf{Q}}_{l''_a}|q_a\rangle\big]_{l_a} \otimes \langle q_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}}
$$

$$
- \sum_{p_a\neq g_a,\, p_b\neq g_b} [\epsilon(p_a)+\epsilon(p_b)]^{-1} \epsilon(p_a)^{-1} \Big[\Big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle \otimes \mathbf{Q}_{l''_a}\big]_{l_a} \otimes \langle p_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}}
$$

$$
+ \sum_{\substack{p_a, q_a\neq g_a \\ p_b\neq g_b}} [\epsilon(q_a)+\epsilon(p_b)]^{-1} \epsilon(p_a)^{-1} \Big[\Big[\big[\langle p_a|\hat{\mathbf{Q}}_{l'_a}|q_a\rangle \otimes \langle q_a|\hat{\mathbf{Q}}_{l''_a}|g_a\rangle\big]_{l_a} \otimes \langle g_a|\hat{\mathbf{Q}}_{l'''_a}|p_a\rangle\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}}
$$

$$
- \sum_{p_a\neq g_a,\, p_b\neq g_b} [\epsilon(p_a)+\epsilon(p_b)]^{-1} \epsilon(p_a)^{-1} \Big[\Big[\big[\mathbf{Q}_{l'_a} \otimes \langle p_a|\hat{\mathbf{Q}}_{l''_a}|g_a\rangle\big]_{l_a} \otimes \langle g_a|\hat{\mathbf{Q}}_{l'''_a}|p_a\rangle\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}}
$$

$$
+ \sum_{\substack{p_a, q_a\neq g_a \\ p_b\neq g_b}} [\epsilon(p_a)+\epsilon(p_b)]^{-1} [\epsilon(q_a)+\epsilon(p_b)]^{-1}
$$

$$
\times \Big[\Big[\big[\langle q_a|\hat{\mathbf{Q}}_{l'_a}|g_a\rangle \otimes \langle g_a|\hat{\mathbf{Q}}_{l''_a}|p_a\rangle\big]_{l_a} \otimes \langle p_a|\hat{\mathbf{Q}}_{l'''_a}|q_a\rangle\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}}
$$

$$
- \sum_{p_a\neq g_a,\, p_b\neq g_b} [\epsilon(p_a)+\epsilon(p_b)]^{-2} \Big[\Big[\big[\langle p_a|\hat{\mathbf{Q}}_{l'_a}|g_a\rangle \otimes \langle g_a|\hat{\mathbf{Q}}_{l''_a}|p_a\rangle\big]_{l_a} \otimes \mathbf{Q}_{l'''_a}\Big]_{L_a} \cdot \mathbf{D}^{L_a}(\omega_a^{-1})
$$

$$
\otimes \Big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle \otimes \langle p_b|\hat{\mathbf{Q}}_{l''_b}|g_b\rangle\big]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_c} \cdot \mathbf{D}^{l_c}(\omega_c^{-1})\Big]_{l_{bc}}\Big]^{N_{abc}}_{L_{abc}},
$$

$$W_{CD}=\frac{1}{2}(4\pi)^{1/2}\sum_{a,b=1}^{N}\sum_{l'_a l''_a l'''_a l_a L_a l'_b l''_b l'''_b l_b L_b l_{ab} L_{ab}}(-1)^{l'_b+l''_b+l'''_b}R_{ab}^{-l'_a-l'_b-l''_a-l''_b-l'''_a-l'''_b-3}$$
$$\times\binom{2l'_a+2l'_b}{2l'_a}^{1/2}\binom{2l''_a+2l''_b}{2l''_a}^{1/2}\binom{2l'''_a+2l'''_b}{2l'''_a}^{1/2}$$
$$\times[l_a,L_a,l_b,L_b,l_{ab},l'_a+l'_b,l''_a+l''_b,l'''_a+l'''_b][L_{ab}]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle$$
$$\times\begin{Bmatrix} l'_a & l''_a & l_a\\ l'_b & l''_b & l_b\\ l'_a+l'_b & l''_a+l''_b & l_{ab}\end{Bmatrix}\begin{Bmatrix} l_a & l'''_a & L_a\\ l_b & l'''_b & L_b\\ l_{ab} & l'''_a+l'''_b & L_{ab}\end{Bmatrix}$$
$$\times\sum_{N_{ab}=-L_{ab}}^{L_{ab}}(-1)^{N_{ab}}Y_{L_{ab}}^{-N_{ab}}(\hat{R}_{ab})\,{}^{CD}T_{L_{ab}}^{N_{ab}}\{((l'_a l''_a)l_a,l'''_a)L_a,((l'_b l''_b)l_b,l'''_b)L_b\},\tag{62}$$

where

$${}^{CD}T_{L_{ab}}^{N_{ab}}\{((l'_a l''_a)l_a,l'''_a)L_a,((l'_b l''_b)l_b,l'''_b)L_b\}$$
$$=\sum_{m'_a m''_a m'''_a m_a M_a m'_b m''_b m'''_b m_b M_b,N_a N_b}{}^{CD}t_{l'_a l''_a l'''_a l'_b l''_b l'''_b}^{m'_a m''_a m'''_a m'_b m''_b m'''_b}\langle l'_a m'_a,l''_a m''_a|l_a m_a\rangle\langle l_a m_a,l'''_a m'''_a|L_a M_a\rangle D_{M_a N_a}^{L_a}(\omega_a^{-1})$$
$$\times\langle l'_b m'_b,l''_b m''_b|l_b m_b\rangle\langle l_b m_b,l'''_b m'''_b|L_b M_b\rangle D_{M_b N_b}^{L_b}(\omega_b^{-1})\langle L_a N_a,L_b N_b|L_{ab}N_{ab}\rangle$$
$$=\sum_{\substack{p_a,q_a\neq g_a\\ p_b,q_b\neq g_b}}[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_a)+\epsilon(q_b)]^{-1}\Big[\big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle\otimes\langle p_a|\hat{\mathbf{Q}}_{l''_a}|q_a\rangle\big]_{l_a}\otimes\langle q_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\big]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})$$
$$\otimes\big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle\otimes\langle p_b|\hat{\mathbf{Q}}_{l''_b}|q_b\rangle\big]_{l_b}\otimes\langle q_b|\hat{\mathbf{Q}}_{l'''_b}|g_b\rangle\big]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\Big]_{L_{ab}}^{N_{ab}}$$
$$-\sum_{\substack{p_a,q_a\neq g_a\\ p_b\neq g_b}}[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_a)+\epsilon(p_b)]^{-1}$$
$$\times\Big[\big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle\otimes\langle p_a|\hat{\mathbf{Q}}_{l''_a}|q_a\rangle\big]_{l_a}\otimes\langle q_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\big]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})$$
$$\otimes\big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle\otimes\mathbf{Q}_{l''_b}\big]_{l_b}\otimes\langle p_b|\hat{\mathbf{Q}}_{l'''_b}|g_b\rangle\big]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\Big]_{L_{ab}}^{N_{ab}}$$
$$-\sum_{\substack{p_a\neq g_a\\ p_b,q_b\neq g_b}}[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_b)+\epsilon(p_a)]^{-1}\Big[\big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle\otimes\mathbf{Q}_{l''_a}\big]_{l_a}$$
$$\otimes\langle p_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\big]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})\otimes\big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle\otimes\langle p_b|\hat{\mathbf{Q}}_{l''_b}|q_b\rangle\big]_{l_b}\otimes\langle q_b|\hat{\mathbf{Q}}_{l'''_b}|g_b\rangle\big]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\Big]_{L_{ab}}^{N_{ab}}$$
$$+\sum_{p_a\neq g_a,\,p_b\neq g_b}[\epsilon(p_a)+\epsilon(p_b)]^{-2}\Big[\big[\big[\langle g_a|\hat{\mathbf{Q}}_{l'_a}|p_a\rangle\otimes\mathbf{Q}_{l''_a}\big]_{l_a}\otimes\langle p_a|\hat{\mathbf{Q}}_{l'''_a}|g_a\rangle\big]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})$$
$$\otimes\big[\big[\langle g_b|\hat{\mathbf{Q}}_{l'_b}|p_b\rangle\otimes\mathbf{Q}_{l''_b}\big]_{l_b}\otimes\langle p_b|\hat{\mathbf{Q}}_{l'''_b}|g_b\rangle\big]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\Big]_{L_{ab}}^{N_{ab}}.$$

In the case of W_{CD} we have applied the following identity:

$$\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle(-1)^{l'_a+l''_a+l'''_a+l'_b+l''_b+l'''_b+L_{ab}}$$
$$=\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle,\tag{63}$$

which results from the well-known property of Clebsch–Gordan coefficients of type $\langle j_1 0,j_2 0|j0\rangle$, i.e.,[48]

$$\langle j_1 0,j_2 0|j0\rangle=(-1)^{j_1+j_2-j}\langle j_1 0,j_2 0|j0\rangle.\tag{64}$$

We have used curly brackets in ${}^{D}T_{L_{abc}}^{N_{abc}}\{\cdots\}$, ${}^{BA}T_{L_{abc}}^{N_{abc}}\{\cdots\}$, and ${}^{CD}T_{L_{ab}}^{N_{ab}}\{\cdots\}$, which appear in the formulas (60)–(62) for the dispersion energies W_D, W_{BA}, and W_{CD}, respectively, to indicate the coupling schemes.

Now we briefly describe simplifications in the general interaction energy expressions (58)–(62), which are possible for some specific parts of them corresponding to the particular categories of the third-order long-range forces defined by Eqs. (21)–(23). These simplifications can be made for $W_{B,2}$, $W_{B,3}$, $W_{A,2}$, $W_{A,3}^{\mathrm{I}}$, $W_{A,3}^{\mathrm{II}}$, and $W_{BA,2}$.

In the case of the pairwise additive category $W_{B,2}$ we must first relabel the indices that occur in the terms of Eq. (58) where $b=c=d$ as follows: $l_b\to l'_b$, $l_c\to l''_b$, $l_d\to l'''_b$, $l_{bc}\to l_b$, $l_{bcd}\to L_b$, $l_{abc}\to l_{ab}$, $L_{abcd}\to L_{ab}$, $N_{abcd}\to N_{ab}$. Using Eqs. (24) and (25) we can easily derive the following formulas:

$$[[\mathbf{Y}_{l'_a+l'_b}(\hat{R}_{ab})\otimes\mathbf{Y}_{l''_a+l''_b}(\hat{R}_{ab})]_{l_{ab}}\otimes\mathbf{Y}_{l'''_a+l'''_b}(\hat{R}_{ab})]_{L_{ab}}$$
$$=(4\pi)^{-1}[l'_a+l'_b,l''_a+l''_b,l'''_a+l'''_b][L_{ab}]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle\mathbf{Y}_{L_{ab}}\hat{R}_{ab}), \tag{65}$$

$$[[\mathbf{Q}_{l'_b}\cdot\mathbf{D}^{l'_b}(\omega_b^{-1})\otimes\mathbf{Q}_{l''_b}\cdot\mathbf{D}^{l''_b}(\omega_b^{-1})]_{l_b}\otimes\mathbf{Q}_{l'''_b}\cdot\mathbf{D}^{l'''_b}(\omega_b^{-1})]_{L_b}=[[\mathbf{Q}_{l'_b}\otimes\mathbf{Q}_{l''_b}]_{l_b}\otimes\mathbf{Q}_{l'''_b}]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1}). \tag{66}$$

If we now substitute Eqs. (65) and (66) into this part of Eq. (58), which corresponds to $W_{B,2}$, i.e., contains the terms where $b=c=d$, and apply the identity (63) we get the final maximally simplified expression for the two-body category $W_{B,2}$.

From Eq. (21) it follows that the three-body category $W_{B,3}$ is completely determined by this part of the sum over a, b, c, and d in Eq. (58) that contains the terms where $b=c\neq d$. Therefore to obtain the maximally simplified expression for $W_{B,3}$ we first relabel the indices that appear in the terms of Eq. (58) where $b=c\neq d$ as follows: $l_b\to l'_b$, $l_c\to l''_b$, $l_{bc}\to l_b$, $l_{bcd}\to l_{bd}$, $l_{abc}\to l_{ab}$, $L_{abcd}\to L_{abd}$, $N_{abcd}\to N_{abd}$. Then using the same arguments as in the case of Eqs. (65) and (66) we can write that

$$[\mathbf{Y}_{l'_a+l'_b}(\hat{R}_{ab})\otimes\mathbf{Y}_{l''_a+l''_b}(\hat{R}_{ab})]_{l_{ab}}=(4\pi)^{-1/2}[l'_a+l'_b,l''_a+l''_b][l_{ab}]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\mathbf{Y}_{l_{ab}}(\hat{R}_{ab}), \tag{67}$$

$$[\mathbf{Q}_{l'_b}\cdot\mathbf{D}^{l'_b}(\omega_b^{-1})\otimes\mathbf{Q}_{l''_b}\cdot\mathbf{D}^{l''_b}(\omega_b^{-1})]_{l_b}=[\mathbf{Q}_{l'_b}\otimes\mathbf{Q}_{l''_b}]_{l_b}\cdot\mathbf{D}^{l_b}(\omega_b^{-1}). \tag{68}$$

According to Eq. (21), if we insert Eqs. (67) and (68) into this part of Eq. (58), which includes the terms where $b=c\neq d$, and multiply the resulting formula by 3, we obtain the final maximally simplified expression for the three-body category $W_{B,3}$.

In the case of the pairwise additive category $W_{A,2}$ we first relabel the indices that occur in the terms of Eq. (59a) where $c=a$, $d=b$ as follows: $l'_c\to l''_a$, $l''_c\to l'''_a$, $l_c\to l_a$, $l_a\to l'_a$, $l_{ac}\to L_a$, $l_d\to l'''_b$, $l_{bd}\to L_b$, $l_{abc}\to l_{ab}$, $L_{abcd}\to L_{ab}$, $N_{abcd}\to N_{ab}$. Then applying the same arguments as in the case of Eqs. (65) and (66), we can write the following expressions:

$$[[\mathbf{Y}_{l'_a+l'_b}(\hat{R}_{ab})\otimes\mathbf{Y}_{l''_a+l''_b}(\hat{R}_{ba})]_{l_{ab}}\otimes\mathbf{Y}_{l'''_a+l'''_b}(\hat{R}_{ab})]_{L_{ab}}$$
$$=(-1)^{l''_a+l''_b}(4\pi)^{-1}[l'_a+l'_b,l''_a+l''_b,l'''_a+l'''_b][L_{ab}]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle$$
$$\times\langle l_{ab}0,l'''_a+l'''_b0|L_{ab}0\rangle\mathbf{Y}_{L_{ab}}(\hat{R}_{ab}), \tag{69}$$

$$[\alpha_{l_a}\{l''_al'''_a\}\cdot\mathbf{D}^{l_a}(\omega_a^{-1})\otimes\mathbf{Q}_{l'_a}\cdot\mathbf{D}^{l'_a}(\omega_a^{-1})]_{L_a}=[\alpha_{l_a}\{l''_al'''_a\}\otimes\mathbf{Q}_{l'_a}]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1}), \tag{70}$$

$$[\alpha_{l_b}\{l'_bl''_b\}\cdot\mathbf{D}^{l_b}(\omega_b^{-1})\otimes\mathbf{Q}_{l'''_b}\cdot\mathbf{D}^{l'''_b}(\omega_b^{-1})]_{L_b}=[\alpha_{l_b}\{l'_bl''_b\}\otimes\mathbf{Q}_{l'''_b}]_{L_b}\cdot\mathbf{D}^{L_b}(\omega_b^{-1}). \tag{71}$$

Inserting Eqs. (69)–(71) into this part of Eq. (59a), which corresponds to $W_{A,2}$, i.e., contains the terms where $c=a$, $d=b$, we get the final maximally simplified formula for the two-body category $W_{A,2}$.

According to Eq. (22), to obtain the maximally simplified formula for the three-body energy $W^{\mathrm{I}}_{A,3}$ we have to examine this part of the sum over a, b, c, and d in Eq. (59a), which includes the terms where $c=a$, $b\neq d$. Therefore in this case we first relabel the indices that occur in the terms of Eq. (59a) where $c=a$, $b\neq d$ in the following way: $l_a\to l'_a$, $l'_c\to l''_a$, $l''_c\to l'''_a$, $l_c\to l_a$, $l_{ac}\to L_a$, $l_{abc}\to l_{ab}$, $L_{abcd}\to L_{abd}$, $N_{abcd}\to N_{abd}$. Using Eqs. (24) and (25) we immediately find that

$$[\mathbf{Y}_{l'_a+l'_b}(\hat{R}_{ab})\otimes\mathbf{Y}_{l''_a+l''_b}(\hat{R}_{ba})]_{l_{ab}}=(-1)^{l''_a+l''_b}(4\pi)^{-1/2}[l'_a+l'_b,l''_a+l''_b]$$
$$\times[l_{ab}]^{-1}\langle l'_a+l'_b0,l''_a+l''_b0|l_{ab}0\rangle\mathbf{Y}_{l_{ab}}(\hat{R}_{ab}), \tag{72}$$

$$[\alpha_{l_a}\{l''_al'''_a\}\cdot\mathbf{D}^{l_a}(\omega_a^{-1})\otimes\mathbf{Q}_{l'_a}\cdot\mathbf{D}^{l'_a}(\omega_a^{-1})]_{L_a}=[\alpha_{l_a}\{l''_al'''_a\}\otimes\mathbf{Q}_{l'_a}]_{L_a}\cdot\mathbf{D}^{L_a}(\omega_a^{-1}). \tag{73}$$

From Eq. (22) it follows that if we substitute Eqs. (72) and (73) into this part of Eq. (59a), which contains the terms where $c=a$, $b\neq d$, and multiply the resulting expression by 2, we get the final maximally simplified formula for the three-body category $W^{\mathrm{I}}_{A,3}$.

Contrary to the energy $W^{\mathrm{I}}_{A,3}$, in the case of $W^{\mathrm{II}}_{A,3}$, which also represents three-body interactions, we must apply Eq. (59b), which differs from Eq. (59a) by the order of couplings. According to Eq. (22), we first relabel the indices that appear in the terms of Eq. (59b) where $c\neq a$, $d=a$ as follows: $l_d\to l'_a$, $l_a\to l''_a$, $l_{ad}\to l_a$, $l_{abcd}\to l_{abc}$, $L_{abcd}\to L_{abc}$, $N_{abcd}\to N_{abc}$. Then with the help of Eq. (24) we can write

$$[\mathbf{Q}_{l'_a}\cdot\mathbf{D}^{l'_a}(\omega_a^{-1})\otimes\mathbf{Q}_{l''_a}\cdot\mathbf{D}^{l''_a}(\omega_a^{-1})]_{l_a}=[\mathbf{Q}_{l'_a}\otimes\mathbf{Q}_{l''_a}]_{l_a}\cdot\mathbf{D}^{l_a}(\omega_a^{-1}). \tag{74}$$

Inserting Eq. (74) into this part of Eq. (59b), which contains the terms where $c\neq a$, $d=a$, we find the final maximally simplified expression for the category $W^{\mathrm{II}}_{A,3}$.

The last category of the third-order interactions between N molecules, for which further simplifications can be made, is the pairwise additive part of the energy W_{BA}, i.e., $W_{BA,2}$. According to Eq. (23), in this case we must first relabel the indices that occur in the terms of Eq. (61) where $b=c$ in the following way: $l_c\to l'''_b$, $l_{bc}\to L_b$, $L_{abc}\to L_{ab}$, $N_{abc}\to N_{ab}$. Then making use of Eqs. (24) and (25) we simply obtain the following equations:

$$\left[\left[\langle g_b | \widehat{\mathbf{Q}}_{l_b'} | p_b \rangle \otimes \langle p_b | \widehat{\mathbf{Q}}_{l_b''} | g_b \rangle \right]_{l_b} \cdot \mathbf{D}^{l_b}(\omega_b^{-1}) \otimes \mathbf{Q}_{l_b'''} \cdot \mathbf{D}^{l_b'''}(\omega_b^{-1}) \right]_{L_b}$$

$$= \left[\left[\langle g_b | \widehat{\mathbf{Q}}_{l_b'} | p_b \rangle \otimes \langle p_b | \widehat{\mathbf{Q}}_{l_b''} | g_b \rangle \right]_{l_b} \otimes \mathbf{Q}_{l_b'''} \right]_{L_b} \cdot \mathbf{D}^{L_b}(\omega_b^{-1}), \tag{75}$$

$$\left[\mathbf{Y}_{l_{ab}}(\widehat{R}_{ab}) \otimes \mathbf{Y}_{l_a''' + l_b'''}(\widehat{R}_{ab}) \right]_{L_{ab}} = (4\pi)^{-1/2} [l_{ab}, l_a''' + l_b'''] [L_{ab}]^{-1} \langle l_{ab} 0, l_a''' + l_b''' 0 | L_{ab} 0 \rangle \mathbf{Y}_{L_{ab}}(\widehat{R}_{ab}). \tag{76}$$

If we now insert Eqs. (75) and (76) into the part of Eq. (61) that contains the terms where $b = c$ and apply the relation (63), we get the final maximally simplified formula for the two-body dispersion category $W_{BA,2}$.

From Eqs. (21)–(23) and (58)–(62) it follows that any further simplifications are not possible for the remaining third-order energies $W_{B,4}$, $W_{A,4}$, W_D, $W_{BA,3}$, and W_{CD}. Therefore the maximally simplified formulas for them are directly given by the adequate parts of the general equations (58)–(62).

Our expressions for all possible categories of the anisotropic long-range molecular interactions arising from the third order of the perturbation theory described in the present section have very readable, formally and physically appealing forms, viz., the respective irreducible tensors localized on interacting molecules are first coupled together. At the same time the geometric factors describing the orientations of the intermolecular vectors $\mathbf{R}_{ij}$ $(i,j = 1,2,...,N)$ are also coupled together to a tensor that contains all the information about the geometry of a system. Then both these resultant tensors, which have the same order, are coupled to a scalar by the phase factor of type $(-1)^{L-N}$ or $(-1)^N$, because[48,52] $\langle L - N, LN | 00 \rangle = (-1)^{L-N} [L]^{-1}$.

It should be pointed out that the present formulas for the pairwise additive parts of the third-order interaction energies in a collection of N molecules, i.e., $W_{B,2}$, $W_{A,2}$, $W_{BA,2}$, and W_{CD}, are in complete agreement with the previously obtained expressions describing the adequate categories of the third-order interactions between two molecules.[11,12,14] One can easily verify this statement in the case of $W_{B,2}$, $W_{BA,2}$, and W_{CD}. To prove that it is also true for $W_{A,2}$, we must apply the identity (63) and the relation

$$\left\langle \begin{array}{ccccccccc} l_b'' & & l_b''' & l_a'' & & l_a''' & l_a''+l_b'' & & l_a'''+l_b''' \\ & l_b & & & l_a & & & l_{ab} & \\ l_b' & & L_b & l_a' & & L_a & l_a'+l_b' & & L_{ab} \end{array} \right\rangle$$

$$= \sum_{l_a^{(1)}} [l_a^{(1)}]^2 \begin{Bmatrix} l_a' & l_b' & l_a'+l_b' \\ l_a'' & l_b'' & l_a''+l_b'' \\ l_a^{(1)} & l_b & l_{ab} \end{Bmatrix} \begin{Bmatrix} l_a^{(1)} & l_b & l_{ab} \\ l_a''' & l_b''' & l_a'''+l_b''' \\ L_a & L_b & L_{ab} \end{Bmatrix}$$

$$\times \begin{Bmatrix} l_a'' & l_a''' & l_a \\ L_a & l_a' & l_a^{(1)} \end{Bmatrix} (-1)^{l_a'+l_b'+l_a''+l_b''+l_a'''+l_b'''+L_{ab}}, \tag{77}$$

which simply follows from the definition of the 15-j symbol of type {2,2} given by Yutsis and Bandzaitis.[53] (Here

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}$$

denotes the Wigner 6-j symbol.[51]) For more details see the Appendix.

In the end of this section we would like to mention that the present formulas for

$$W_B = W_{B,2} + W_{B,3} + W_{B,4},$$

$$W_A = W_{A,2} + W^{\mathrm{I}}_{A,3} + W^{\mathrm{II}}_{A,3} + W_{A,4},$$

$W_D, W_{BA} = W_{BA,2} + W_{BA,3}$, and W_{CD}, given by Eqs. (21)–(23), (58)–(63), (65)–(76), are the closed spherical tensor analogs of Eqs. (2.13), (2.16), (2.21), (2.32), and (2.38) in Stogryn's 1971 paper,[17] where the Cartesian notation is applied. Let us recall that the electric induction categories W_B and W_A, which arise from the third order of the quantum-mechanical perturbation theory,[17] are classical in nature and can also be found using only classical electrostatic arguments, as described by Kielich[41] and Stogryn.[32] Stogryn's energy W_D (see Ref. 17) is the maximal generalization of all earlier results on purely pairwise nonadditive three-body dispersion forces.[21,22,25,56–64] $W_{BA} = W_{BA,2} + W_{BA,3}$ describes a category of the third-order dispersion interactions, which had not been considered in the literature before the publication of Stogryn's work.[17] Finally, the purely pairwise additive category of the third-order dispersion interactions W_{CD} had been considered in the literature before the publication of Stogryn's paper,[17] but only for spherically symmetric molecules.[65]

VI. CONCLUSIONS

Let us summarize the preceding results. It is seen that the application of the spherical tensor formalism leads to the closed, physically appealing expressions for all possible types of the anisotropic interactions between N arbitrary molecules that occur in the third order of the perturbation theory, i.e., $W_{B,2}$, $W_{B,3}$, $W_{B,4}$, $W_{A,2}$, $W^{\mathrm{I}}_{A,3}$, $W^{\mathrm{II}}_{A,3}$, $W_{A,4}$, W_D, $W_{BA,2}$, $W_{BA,3}$, and W_{CD}. In contrast with the analogous formulas for them, when the Cartesian notation is used,[17] they have the following advantages: (i) the dependence on the orientations of the molecules in the space and the dependence on the orientations of the intermolecular vectors that occur in them are completely separated and pushed to their limits, and (ii) the physical parts of them are also separated and related to the irreducible tensors localized on interacting molecules in molecular (body-fixed) frames [in the case of the induction categories these tensors are directly connected with the molecular spherical multipole moments and irreducible (hyper)polarizabilities]. From the above nice properties of our energy expressions it follows that they can be additionally simplified if at least one molecule in a system has finite symmetry group; this remark is the same as in the case of the use of spherical tensors to the description of two-body long-range interactions.[9,10,12–14,66,67] So demonstrated in this paper formulas for $W_{B,2}$, $W_{B,3}$, $W_{B,4}$, $W_{A,2}$, $W^{\rm I}_{A,3}$, $W^{\rm II}_{A,3}$, $W_{A,4}$, W_D, $W_{BA,2}$, $W_{BA,3}$, and W_{CD} should be very convenient in practice; in the case of the dispersion categories W_D, $W_{BA,2}$, $W_{BA,3}$, and W_{CD}, this will be seen more clearly in the next section of the present work, where we apply simple approximations to express them through molecular spherical multipole moments and (hyper)polarizabilities. Let us recall that similar advantages are also valid for the spherical tensor theory of the long-range interactions between two molecules in all orders of the perturbation theory[7–14] and for the spherical tensor description of the long-range forces between N molecules, when perturbation treatment up to second order is used.[7–10,13,18] Therefore the spherical tensor formalism really provides the best and a very effective method for the examination of long-range pairwise additive as well as pairwise nonadditive molecular interactions.

VII. CONNECTION OF W_D, $W_{BA,2}$ $W_{BA,3}$, AND W_{CD} WITH ELECTRICAL PROPERTIES OF INTERACTING MOLECULES

In this section we briefly describe some useful approximations for the exact formulas for the dispersion energies W_D, $W_{BA,2}$, $W_{BA,3}$, and W_{CD} obtained in the present paper. They are similar to those described by Stogryn,[17] so we do not report the details here.

Let us apply the Unsöld[68] or, more precisely, the Buckingham approximation[2] to formula (60). This means that we must replace the D that occurs in ${}^{D}T^{N_{abc}}_{L_{abc}}\{\cdots\}$ by

$$\frac{(U_a+U_b+U_c)U_aU_bU_c}{(U_a+U_b)(U_b+U_c)(U_c+U_a)\epsilon(p_a)\epsilon(p_b)\epsilon(p_c)}, \tag{78}$$

where U_a, U_b, and U_c are the characteristic energies associated with molecules a, b, and c, respectively (in the zeroth approximation, their first ionization potentials). In this way we obtain the following approximated form of ${}^{D}T^{N_{abc}}_{L_{abc}}\{\cdots\}$:

$$\begin{aligned}&{}^{D}T^{N_{abc}}_{L_{abc}}\{((l'_al''_a)l_a,(l'_bl''_b)l_b)l_{ab},(l'_cl''_c)l_c\}\\&\approx\frac{(U_a+U_b+U_c)U_aU_bU_c}{(U_a+U_b)(U_b+U_c)(U_c+U_a)}\\&\quad\times\big[\big[\boldsymbol{\alpha}_{l_a}\{l'_al''_a\}\cdot\mathbf{D}^{l_a}(\omega_a^{-1})\\&\quad\otimes\boldsymbol{\alpha}_{l_b}\{l'_bl''_b\}\cdot\mathbf{D}^{l_b}(\omega_b^{-1})\big]_{l_{ab}}\\&\quad\otimes\boldsymbol{\alpha}_{l_c}\{l'_cl''_c\}\cdot\mathbf{D}^{l_c}(\omega_c^{-1})\big]^{N_{abc}}_{L_{abc}}.\end{aligned} \tag{79}$$

Equation (60) together with Eq. (79) represent a useful approximate expression for W_D, related to the irreducible polarizabilities $\boldsymbol{\alpha}_{l_i}\{l'_il''_i\}$ ($i=1,2,\ldots,N$) of interacting molecules in molecule-fixed frames.

A similar approximation for ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$ is not so simple but also possible.[17] Here, $[\epsilon(p_a)+\epsilon(p_b)]^{-1}\epsilon(q_a)^{-1}$ and $[\epsilon(q_a)+\epsilon(p_b)]^{-1}\epsilon(p_a)^{-1}$, which occur in the first and third terms of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, respectively, can be replaced by

$$\frac{U_aU_b}{(U_a+U_b)\epsilon(p_a)\epsilon(q_a)\epsilon(p_b)}, \tag{80}$$

while $[\epsilon(p_a)+\epsilon(p_b)]^{-1}\epsilon(p_a)^{-1}$, which occurs in the second and fourth terms of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, can be replaced by

$$\frac{U_aU_b}{(U_a+U_b)\epsilon(p_a)^2\epsilon(p_b)}. \tag{81}$$

However, $[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_a)+\epsilon(p_b)]^{-1}$, which occurs in the fifth term of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, can be replaced by $U'_a/(U'_a+U_b)$ times (80), where U'_a is a certain characteristic energy associated with molecule a, which can be other than U_a [U'_a is connected with the presence of $\epsilon(q_a)$ in the fifth term of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$]. Finally $[\epsilon(p_a)+\epsilon(p_b)]^{-2}$, which occurs in the sixth term of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, can be replaced by $U_a/(U_a+U_b)$ times (81). In order to obtain an approximate form of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, we first ignore the additional factors $U'_a/(U'_a+U_b)$ and $U_a/(U_a+U_b)$, which occur in the fifth and sixth terms of ${}^{BA}T^{N_{abc}}_{L_{abc}}\{\cdots\}$, use the definitions of the molecular irreducible tensors described in Sec. II, and then correct the omission of the above factors by the insertion of a factor ξ. In this way we find

$$\begin{aligned}&{}^{BA}T^{N_{abc}}_{L_{abc}}\{((l'_al''_a)l_a,l'''_a)L_a,((l'_bl''_b)l_b,l_c)l_{bc}\}\\&\approx\xi\frac{U_aU_b}{(U_a+U_b)}\big[\tilde{\boldsymbol{\beta}}_{L_a}\{(l'_al''_a)l_a,l'''_a\}\\&\quad\cdot\mathbf{D}^{L_a}(\omega_a^{-1})\otimes\big[\boldsymbol{\alpha}_{l_b}\{l'_bl''_b\}\cdot\mathbf{D}^{l_b}(\omega_b^{-1})\\&\quad\otimes\mathbf{Q}_{l_c}\cdot\mathbf{D}^{l_c}(\omega_c^{-1})\big]_{l_{bc}}\big]^{N_{abc}}_{L_{abc}}.\end{aligned} \tag{82}$$

Employing Stogryn's estimate for ξ,[17] we can write that

$$\xi\approx\tfrac{5}{6}\quad\text{and}\quad\tfrac{2}{3}<\xi<1. \tag{83}$$

The application of Eqs. (23), (61), (63), (75), and (76) together with Eq. (82) leads to the useful approximation for $W_{BA}=W_{BA,2}+W_{BA,3}$, related to the spherical multipole moments and irreducible (hyper)polarizabilities of interacting molecules in body-fixed axis systems.

W_{CD} can be simplified by using analogous approximations as in the case of $W_{BA,2}$ and $W_{BA,3}$, i.e., $[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_a)+\epsilon(q_b)]^{-1}$ is replaced by

$$\frac{U_aU_b}{(U_a+U_b)\epsilon(p_a)\epsilon(p_b)}\times\frac{U'_aU'_b}{(U'_a+U'_b)\epsilon(q_a)\epsilon(q_b)}, \tag{84}$$

$[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_a)+\epsilon(p_b)]^{-1}$ is replaced by

$$\frac{U_aU_b}{(U_a+U_b)\epsilon(p_a)\epsilon(p_b)}\times\frac{U'_aU_b}{(U'_a+U_b)\epsilon(q_a)\epsilon(p_b)}, \tag{85}$$

$[\epsilon(p_a)+\epsilon(p_b)]^{-1}[\epsilon(q_b)+\epsilon(p_a)]^{-1}$ is replaced by

$$\frac{U_a U_b}{(U_a+U_b)\epsilon(p_a)\epsilon(p_b)}\times\frac{U_a U'_b}{(U_a+U'_b)\epsilon(q_b)\epsilon(p_a)}, \tag{86}$$

and $[\epsilon(p_a)+\epsilon(p_b)]^{-2}$ is replaced by

$$\left(\frac{U_a U_b}{U_a+U_b}\right)^2\times\frac{1}{\epsilon(p_a)^2\epsilon(p_b)^2}. \tag{87}$$

If we use Eqs. (5), (7), (8), (62), and (84)–(87), and ignore the differences between U_a and U'_a, U_b and U'_b, we obtain the following simplified expression for ${}^{CD}T^{N_{ab}}_{L_{ab}}\{\cdots\}$:

$$\begin{aligned}
&{}^{CD}T^{N_{ab}}_{L_{ab}}\{((l'_a l''_a)l_a,l'''_a)L_a,((l'_b l''_b)l_b,l'''_b)L_b\}\\
&\quad\approx\left(\frac{U_a U_b}{U_a+U_b}\right)^2\left[\boldsymbol{\beta}_{L_a}\{(l'_a l''_a)l_a,l'''_a\}_{[(1+2)+3]}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})\otimes\boldsymbol{\beta}_{L_b}\{(l'_b l''_b)l_b,l'''_b\}_{[(1+2)+3]}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\right]^{N_{ab}}_{L_{ab}}\\
&\quad\approx\frac{1}{9}\left(\frac{U_a U_b}{U_a+U_b}\right)^2\left[\tilde{\boldsymbol{\beta}}_{L_a}\{(l'_a l''_a)l_a,l'''_a\}\cdot\mathbf{D}^{L_a}(\omega_a^{-1})\otimes\tilde{\boldsymbol{\beta}}_{L_b}\{(l'_b l''_b)l_b,l'''_b\}\cdot\mathbf{D}^{L_b}(\omega_b^{-1})\right]^{N_{ab}}_{L_{ab}}.
\end{aligned} \tag{88}$$

If we insert Eq. (88) into Eq. (62) we obtain a useful approximate expression for W_{CD}, related to the irreducible hyperpolarizabilities $\boldsymbol{\beta}_{L_i}\{(l'_i l''_i)l_i,l'''_i\}_{[(1+2)+3]}$ or $\tilde{\boldsymbol{\beta}}_{L_i}\{(l'_i l''_i)l_i,l'''_i\}$ ($i=1,2,\ldots,N$) of interacting molecules in local coordinate systems fixed in them.

Approximate formulas for the dispersion energies W_D, $W_{BA,2}$, $W_{BA,3}$, and W_{CD} obtained in this section should be convenient in practice. Similarly as in the case of the induction categories $W_{B,2}$, $W_{B,3}$, $W_{B,4}$, $W_{A,2}$, $W^{\mathrm{I}}_{A,3}$, $W^{\mathrm{II}}_{A,3}$, $W_{A,4}$, they are directly related to the irreducible tensors describing electrical properties of interacting molecules in molecular (body-fixed) frames. Therefore full advantage can be taken from the point symmetry of the molecules under consideration, because spherical multipole moments and irreducible (hyper)polarizabilities possess very nice transformational properties.[10,30,31]

ACKNOWLEDGMENTS

The author would like to thank Professor W. Kołos and Assistant Professor L. Piela for their kindly invitating him to the Quantum Chemistry Laboratory of Warsaw University in April 1984. He is very much indebted to Assistant Professor B. Jeziorski for a very stimulating and enlightening discussion during this visit.

APPENDIX: RELATION OF THE PRESENT WORK TO THE PREVIOUS RESULTS ON THE INTERACTIONS BETWEEN TWO MOLECULES

In this Appendix we briefly show that expressions for the pairwise additive part of the third-order interactions between N molecules obtained in the present paper are in agreement with the recent results on the third-order interactions between two molecules.[11,12,14] Adapting general formulas for all possible categories of the third-order interactions in the uncoupled form given by Eqs. (15)–(19) to the case of two interacting molecules a and b, and denoting the third-order interaction energy between molecules a and b by $W_{3,ab}$, we easily find

$$\begin{aligned}
W_{3,ab}=&\sum_{\substack{l'_a l''_a l'''_a l'_b l''_b l'''_b\\ n'_a n''_a n'''_a n'_b n''_b n'''_b\\ m'_a m''_a m'''_a m'_b m''_b m'''_b}} C^{n'_a n'_b}_{l'_a l'_b} C^{n''_a n''_b}_{l''_a l''_b} C^{n'''_a n'''_b}_{l'''_a l'''_b} f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} D^{l'_a}_{m'_a n'_a}(\omega_a^{-1})\\
&\times D^{l''_a}_{m''_a n''_a}(\omega_a^{-1}) D^{l'''_a}_{m'''_a n'''_a}(\omega_a^{-1}) D^{l'_b}_{m'_b n'_b}(\omega_b^{-1}) D^{l''_b}_{m''_b n''_b}(\omega_b^{-1}) D^{l'''_b}_{m'''_b n'''_b}(\omega_b^{-1}),
\end{aligned} \tag{A1}$$

where

$$f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} = {}_{(B)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} + {}_{(A)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} + {}_{(BA)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} + {}_{(CD)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}.$$

Here ${}_{(X)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}$ ($X=B,A,BA,CD$) correspond to the energies W_X ($X=B,A,BA,CD$), respectively, and they are defined as

$$\begin{aligned}
{}_{(B)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} =& \tfrac{1}{3}\left(\beta^{m'_a m''_a m'''_a}_{l'_a l''_a l'''_a} + \beta^{m''_a m'''_a m'_a}_{l''_a l'''_a l'_a} + \beta^{m'''_a m'_a m''_a}_{l'''_a l'_a l''_a}\right) Q^{m'_b}_{l'_b} Q^{m''_b}_{l''_b} Q^{m'''_b}_{l'''_b}\\
&+ Q^{m'_a}_{l'_a} Q^{m''_a}_{l''_a} Q^{m'''_a}_{l'''_a}\,\tfrac{1}{3}\left(\beta^{m'_b m''_b m'''_b}_{l'_b l''_b l'''_b} + \beta^{m''_b m'''_b m'_b}_{l''_b l'''_b l'_b} + \beta^{m'''_b m'_b m''_b}_{l'''_b l'_b l''_b}\right),
\end{aligned} \tag{A2}$$

$${}_{(A)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} = 2Q^{m'_a}_{l'_a}\alpha^{m''_a m'''_a}_{l''_a l'''_a}\alpha^{m'_b m''_b}_{l'_b l''_b} Q^{m'''_b}_{l'''_b} + 2\alpha^{m'_a m''_a}_{l'_a l''_a} Q^{m'''_a}_{l'''_a} Q^{m'_b}_{l'_b}\alpha^{m''_b m'''_b}_{l''_b l'''_b}, \tag{A3}$$

$${}_{(BA)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} = {}^{BA}t^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} + {}^{BA}t^{m'_b m''_b m'''_b m'_a m''_a m'''_a}_{l'_b l''_b l'''_b l'_a l''_a l'''_a}, \tag{A4}$$

$$ {}_{(CD)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} = \tfrac{1}{2}\,{}^{CD}t^{\,m'_a m''_a m'''_a m'_b m''_b m'''_b}_{\,l'_a l''_a l'''_a l'_b l''_b l'''_b} + \tfrac{1}{2}\,{}^{CD}t^{\,m'_b m''_b m'''_b m'_a m''_a m'''_a}_{\,l'_b l''_b l'''_b l'_a l''_a l'''_a} = {}^{CD}t^{\,m'_a m''_a m'''_a m'_b m''_b m'''_b}_{\,l'_a l''_a l'''_a l'_b l''_b l'''_b}. \tag{A5} $$

Note that in the case of two molecules $W_D = 0$, so this type of interaction energy is not represented in the expression (A1).

Recently we have proved that a formula of type (A1) can be put into the following closed form[11,12,14]:

$$ \begin{aligned} W_{3,ab} = (4\pi)^{1/2} & \sum_{l'_a l''_a l'''_a l'_b l''_b l'''_b l_a L_a l_b L_b l_{ab} L_{ab}} (-1)^{l'_b + l''_b + l'''_b} R_{ab}^{-l'_a - l'_b - l''_a - l''_b - l'''_a - l'''_b - 3} [l'_a + l'_b, l''_a + l''_b, l'''_a + l'''_b] \\ & \times \binom{2l'_a + 2l'_b}{2l'_a}^{1/2} \binom{2l''_a + 2l''_b}{2l''_a}^{1/2} \binom{2l'''_a + 2l'''_b}{2l'''_a}^{1/2} \\ & \times \begin{Bmatrix} l'_a & l'_b & l'_a + l'_b \\ l''_a & l''_b & l''_a + l''_b \\ l_a & l_b & l_{ab} \end{Bmatrix} \begin{Bmatrix} l_a & l_b & l_{ab} \\ l'''_a & l'''_b & l'''_a + l'''_b \\ L_a & L_b & L_{ab} \end{Bmatrix} \\ & \times \langle l'_a + l'_b 0, l''_a + l''_b 0 | l_{ab} 0\rangle \langle l_{ab} 0, l'''_a + l'''_b 0 | L_{ab} 0\rangle [l_a, l_b, l_{ab}, L_a, L_b] [L_{ab}]^{-1} \sum_{N_{ab} = -L_{ab}}^{L_{ab}} (-1)^{N_{ab}} \\ & \times Y_{L_{ab}}^{-N_{ab}}(\hat{R}_{ab}) \sum_{m'_a m''_a m'''_a m'_b m''_b m'''_b m_a M_a m_b M_b N_a N_b} f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b} \langle l'_a m'_a, l''_a m''_a | l_a m_a\rangle \langle l_a m_a, l'''_a m'''_a | L_a M_a\rangle \\ & \times \langle l'_b m'_b, l''_b m''_b | l_b m_b\rangle \langle l_b m_b, l'''_b m'''_b | L_b M_b\rangle D^{L_a}_{M_a N_a}(\omega_a^{-1}) D^{L_b}_{M_b N_b}(\omega_b^{-1}) \langle L_a N_a, L_b N_b | L_{ab} N_{ab}\rangle. \end{aligned} \tag{A6} $$

Inserting Eqs. (A2), (A4), and (A5) into Eq. (A6) we immediately obtain the closed expressions for $W_{B,2}$, $W_{BA,2}$, and W_{CD}, respectively, given in Sec. V adapted to the case of two interacting molecules [see Eqs. (21), (23), (58), (61), (62), (65), (66), (75), and (76) accompanied by Eq. (63)]. To obtain the closed formula for $W_{A,2}$ corresponding to the case of two interacting molecules a and b we must first change the coupling schemes that occur in Eq. (A6). It is seen that ${}_{(A)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}$ contains two parts, for which two different coupling schemes are required. In the case of $Q^{m'_a}_{l'_a} \alpha^{m''_a m'''_a}_{l''_a l'''_a} \alpha^{m'_b m''_b}_{l'_b l''_b} Q^{m'''_b}_{l'''_b}$ the needed order of couplings is [(2 + 3) + 1] for l_a's and [(1 + 2) + 3] for l_b's. It can be obtained in Eq. (A6) owing to the following property of the Wigner 6-j symbols[48]:

$$ \begin{aligned} & \sum_{m_a} \langle l'_a m'_a, l''_a m''_a | l_a m_a\rangle \langle l_a m_a, l'''_a m'''_a | L_a M_a\rangle \\ & = \sum_{l_a^{(1)} m_a^{(1)}} \langle l''_a m''_a, l'''_a m'''_a | l_a^{(1)} m_a^{(1)}\rangle \langle l_a^{(1)} m_a^{(1)}, l'_a m'_a | L_a M_a\rangle \begin{Bmatrix} l''_a & l'''_a & l_a^{(1)} \\ L_a & l'_a & l_a \end{Bmatrix} (-1)^{l'_a + l''_a + l_a^{(1)}} [l_a, l_a^{(1)}]. \end{aligned} \tag{A7} $$

Similarly, in the case of $\alpha^{m'_a m''_a}_{l'_a l''_a} Q^{m'''_a}_{l'''_a} Q^{m'_b}_{l'_b} \alpha^{m''_b m'''_b}_{l''_b l'''_b}$ the needed order of couplings is [(1 + 2) + 3] for l_a's and [(2 + 3) + 1] for l_b's. Again we can introduce it into Eq. (A6) applying the following property of the 6-j symbols[48]:

$$ \begin{aligned} & \sum_{m_b} \langle l'_b m'_b, l''_b m''_b | l_b m_b\rangle \langle l_b m_b, l'''_b m'''_b | L_b M_b\rangle \\ & = \sum_{l_b^{(1)} m_b^{(1)}} \langle l''_b m''_b, l'''_b m'''_b | l_b^{(1)} m_b^{(1)}\rangle \langle l_b^{(1)} m_b^{(1)}, l'_b m'_b | L_b M_b\rangle \begin{Bmatrix} l''_b & l'''_b & l_b^{(1)} \\ L_b & l'_b & l_b \end{Bmatrix} [l_b, l_b^{(1)}] (-1)^{l'_b + l''_b + l_b^{(1)}}. \end{aligned} \tag{A8} $$

If we now replace $f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}$ by ${}_{(A)}f^{m'_a m''_a m'''_a m'_b m''_b m'''_b}_{l'_a l''_a l'''_a l'_b l''_b l'''_b}$ in Eq. (A6), then use Eq. (A7) to the part of the resulting formula where $Q^{m'_a}_{l'_a} \alpha^{m''_a m'''_a}_{l''_a l'''_a} \alpha^{m'_b m''_b}_{l'_b l''_b} Q^{m'''_b}_{l'''_b}$ occurs and Eq. (A8) to the part of it where $\alpha^{m'_a m''_a}_{l'_a l''_a} Q^{m'''_a}_{l'''_a} Q^{m'_b}_{l'_b} \alpha^{m''_b m'''_b}_{l''_b l'''_b}$ appears, we get a clear expression for the energy $W_{A,2}$ corresponding to the case of two interacting molecules a and b. From Eqs. (22), (59a), (63), (69)–(71), and (77) it immediately follows that this expression is equal to the formula for $W_{A,2}$ given in Sec. V, when the interaction between two molecules a and b is considered.

Therefore the general expressions for the pairwise additive categories $W_{B,2}$, $W_{A,2}$, $W_{BA,2}$, and W_{CD} derived in the present paper are in complete agreement with the previous considerations concerning the third-order interactions between two molecules.[11,12,14]

[1]E. A. Power, Adv. Chem. Phys. **12**, 167 (1967).
[2]A. D. Buckingham, Adv. Chem. Phys. **12**, 107 (1967).
[3]H. Margenau, Rev. Mod. Phys. **11**, 1 (1939).
[4]H. C. Longuet-Higgins, Proc. R. Soc. London Ser. A **235**, 537 (1956).
[5]J. O. Hirschfelder and W. J. Meath, Adv. Chem. Phys. **12**, 3 (1967).
[6]H. Margenau and N. R. Kestner, *Theory of Intermolecular Forces* (Pergamon, Oxford, 1971).
[7]P. E. S. Wormer, Ph.D. thesis, University of Nijmegen, Nijmegen, 1975.
[8]P. E. S. Wormer, F. Mulder, and A. van der Avoird, Int. J. Quantum

Chem. **11**, 959 (1977).
[9] A. van der Avoird, P. E. S. Wormer, F. Mulder, and R. M. Berns, Topics Curr. Chem. **93**, 1 (1980).
[10] R. P. Leavitt, J. Chem. Phys. **72**, 3472 (1980); **73**, 2017E (1980).
[11] P. Piecuch, Int. J. Quantum Chem. **25**, 449 (1984).
[12] P. Piecuch, Chem. Phys. Lett. **106**, 364 (1984).
[13] A. J. Stone and R. J. A. Tough, Chem. Phys. Lett. **110**, 123 (1984).
[14] P. Piecuch, Int. J. Quantum Chem. **28**, 375 (1985).
[15] See the review in Ref. 9; see also Refs. 12 and 14.
[16] See Ref. 8 and references therein.
[17] D. E. Stogryn, Mol. Phys. **22**, 81 (1971).
[18] P. Piecuch, Chem. Phys. Lett. **110**, 496 (1984).
[19] See, e.g., the introduction to Ref. 17, see also Ref. 18.
[20] R. Giua and V. Schettino, Chem. Phys. Lett. **105**, 86 (1984).
[21] See Chap. 5 in Ref. 6 and references therein.
[22] T. Kihara, Adv. Chem. Phys. **1**, 267 (1958), and references therein.
[23] O. Sinanoğlu, Adv. Chem. Phys. **12**, 283 (1967), and references therein.
[24] O. Sinanoğlu, in *Molecular Interactions*, edited by H. Ratajczak and W. J. Orville-Thomas (Wiley, Chichester, 1982), Vol. 3, p. 283, and references therein.
[25] W. Goetze and H. Schmidt, Z. Phys. **192**, 409 (1966), and references therein.
[26] I. J. Zucker and G. G. Chell, J. Phys. C **1**, 1505 (1968), and references therein.
[27] F. O. Rice and E. Teller, *The Structure of Matter* (Wiley, New York, 1949), Chap. 8.
[28] See Chap. 9 in Ref. 6, and references therein.
[29] M. J. Sparnaay, Physica **25**, 217 (1959).
[30] C. G. Gray and B. W. N. Lo, Chem. Phys. **14**, 73 (1976).
[31] P. Piecuch, Acta Phys. Pol. A **66**, 69 (1984).
[32] See, for example, D. E. Stogryn, Mol. Phys. **23**, 897 (1972).
[33] P. Piecuch, "Spherical tensor theory of long-range interactions in a system of N arbitrary molecules including quantum-mechanical many-body effects. III. Isotropic interactions in the first three orders of perturbation theory," Mol. Phys. (submitted).
[34] B. C. Carlson and G. S. Rushbrooke, Proc. Cambridge Philos. Soc. **46**, 626 (1950).
[35] M. E. Rose, J. Math. Phys. (Cambridge, Mass.) **37**, 215 (1958).
[36] O. Steinborn, Chem. Phys. Lett. **3**, 671 (1969).
[37] E. O. Steinborn and K. Ruedenberg, Adv. Quantum Chem. **7**, 1 (1973).
[38] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
[39] A. T. Amos and R. J. Crispin, in *Theoretical Chemistry, Advances and Perspectives*, edited by H. Eyring and D. J. Henderson (Academic, New York, 1976), Vol. 2, p. 1.
[40] L. Jansen, Physica **23**, 599 (1957).
[41] S. Kielich, Acta Phys. Pol. **28**, 459 (1965).
[42] C. G. Gray, Can. J. Phys. **54**, 505 (1976).
[43] P. Piecuch, J. Phys. A **18**, L739 (1985).
[44] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Macmillan, New York, 1935).
[45] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[46] These manipulations are similar to the manipulations applied to the expressions for W_B, W_A, W_D, W_{BA}, and W_{CD} by Stogryn in Ref. 17, when the Cartesian notation is exploited, so we do not report them here.
[47] Note that the multipole expansion operator V_{ij} given by Eq. (4) can be rewritten in terms of $C^{n_in_j}_{l_il_j}$ as follows:

$$V_{ij} = \sum_{l_i,l_j=0}^{\infty} \left[\hat{\mathbf{Q}}_{l_i}\cdot \mathbf{D}^{l_i}(\omega_i^{-1})\right]\cdot \mathbf{C}_{l_il_j}\cdot \left[\hat{\mathbf{Q}}_{l_j}\cdot \mathbf{D}^{l_j}(\omega_j^{-1})\right].$$

[48] A. Messiah, *Mécanique Quantique* (Dunod, Paris, 1960), Vol. 2, pp. 906–926.
[49] A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Matematičeskij Apparat Teorii Momenta Količestva Dviženija* (Mintis, Vil'nius, 1960) [*Mathematical Apparatus of the Theory of Angular Momentum* (Israel Programme for Scientific Translation, Jerusalem, 1962)].
[50] I. Lindgren and J. Morrison, *Atomic Many-Body Theory* (Springer, Berlin, 1982).
[51] E. P. Wigner, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), p. 87.
[52] A. P. Yutsis and A. A. Bandzaitis, *Teorija Momenta Količestva Dviženija v Kvantovoj Mechanike* (Mintis, Vil'nius, 1965).
[53] See p. 194 in Ref. 52.
[54] See Eqs. (20.1)–(20.3) in Ref. 49.
[55] See Eq. (21) or Eq. (33) in Ref. 11.
[56] B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943).
[57] Y. Muto, Nippon Suugaku-Buturigakkai Kizi **17**, 629 (1943).
[58] B. M. Axilrod, J. Chem. Phys. **17**, 1349 (1949); **19**, 719 (1951).
[59] Y. Midzuno and T. Kihara, J. Phys. Soc. Jpn. **11**, 1045 (1956).
[60] R. U. Ayres and R. H. Tredgold, Proc. Phys. Soc. B **69**, 840 (1956).
[61] W. L. Bade, J. Chem. Phys. **27**, 1280 (1957).
[62] N. R. Kestner and O. Sinanoğlu, J. Chem. Phys. **38**, 1730 (1963).
[63] D. E. Stogryn, Phys. Rev. Lett. **24**, 971 (1970).
[64] R. J. Bell, J. Phys. B **3**, 751 (1970).
[65] See, e.g., Y. M. Chan and A. Dalgarno, Mol. Phys. **14**, 101 (1968).
[66] F. Mulder, A. van der Avoird, and P. E. S. Wormer, Mol. Phys. **37**, 159 (1979).
[67] F. Mulder, G. van Dijk, and A. van der Avoird, Mol. Phys. **39**, 407 (1980).
[68] A. Unsöld, Z. Phys. **43**, 563 (1927).

The eigenvalues of the simplified ideal MHD ballooning equation

R. B. Paris[a] and N. Auby
Association Euratom–CEA, Centre d'Etudes Nucléaires, 92260 Fontenay-aux-Roses, France
R. Y. Dagazian
Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 9 July 1985; accepted for publication 2 April 1986)

The investigation of the spectrum of the simplified differential equation describing the variation of the amplitude of the ideal MHD ballooning instability along magnetic field lines constitutes a multiparameter Schrödinger eigenvalue problem. An exact eigenvalue relation for the discrete part of the spectrum is obtained in terms of the oblate spheroidal functions. The dependence of the eigenvalues λ on the two free parameters γ^2 and μ^2 of the equation is discussed, together with certain analytical approximations in the limits of small and large γ^2. A brief review of the principal properties of the spheroidal functions is given in an appendix.

I. INTRODUCTION

An important mode of instability for many thermonuclear plasma confinement systems is the ideal magnetohydrodynamic (MHD) pressure-driven instability known as the interchange mode. A particularly dangerous form of this instability in toroidal magnetic confinement configurations is the so-called ballooning instability. This is a nonlocal version of an interchange mode that adapts itself to fit the varying curvature of magnetic field lines in a torus. Ballooning instabilities can impose a serious limitation on the amount of plasma that can be stably confined in a toroidal device. The driving mechanism of these instabilities results from the interaction of the plasma pressure gradient with local regions of unfavorable magnetic curvature. This causes the plasma to bulge out, or to "balloon," in these regions in an analogous manner to the aneurysms that develop at weak spots in a pressurized elastic container.

The ideal MHD ballooning instability has been investigated by various authors.[1–5] The structure of these modes is characterized by a rapid variation perpendicular to the magnetic field in a magnetic flux surface (high toroidal mode number) with a slow variation along field lines and across flux surfaces. In the limit of high toroidal mode number, the system of ideal MHD equations for incompressible plasma displacements and negligible mode kinetic energy associated with the component of the displacement parallel to the field leads to a second-order ordinary differential equation describing the variation of the mode amplitude along the magnetic field lines on each flux surface.[1–4] The coefficients of this equation in the general toroidal case are extremely complicated functions of the independent variable. Recent theoretical investigations[6–9] have shown that in certain situations these coefficients may be simplified considerably to yield the differential equation describing the variation of the mode amplitude $y(x)$ along the field lines in the form

$$\frac{d}{dx}\left[(1+x^2)\frac{dy}{dx}\right]-\left[\lambda+\gamma^2(1+x^2)-\frac{\mu^2}{1+x^2}\right]y=0,\qquad -\infty<x<\infty\,. \tag{1.1}$$

The parameter γ is essentially the growth rate of the mode and λ and μ^2 are parameters that depend on the pressure gradient and magnetic shear of the equilibrium configuration. The boundary condition that $y(x)\to 0$ as $x\to\pm\infty$ then constitutes a multiparameter eigenvalue problem for any one of the three parameters λ, γ^2, or μ^2 in terms of the other two. Although in the physical problem it is the growth rate parameter γ^2 that is to be determined, it is customary—and entirely equivalent—in the mathematical treatment of Eq. (1.1) to consider γ as the eigenvalue with λ^2 and μ^2 as arbitrary parameters.

The solution of Eq. (1.1) is expressed in terms of radial oblate spheroidal functions, for which the principal properties have been reviewed in Ref. 10. The analytic solution is complicated by the existence of a parameter known as the characteristic exponent ν, which depends in a transcendental manner on λ, γ^2, and μ^2. This exponent arises in the theory in exactly the same way as the characteristic exponent in the solution of Mathieu's differential equation [to which Eq. (1.1) is related in the particular case $\mu^2=\frac{1}{4}$]. Unlike the case of the spheroidal wave functions (which arise when the wave equation is separated in spheroidal coordinates), where ν is necessarily an integer,[11] the characteristic exponent for the ballooning equation (1.1) is found to be complex over most of the range of the parameters and its determination proves to be a central issue in the theory.

The aim of this paper is to discuss the nature and dependence of the spectrum of λ-eigenvalues on the parameters γ^2 and μ^2. An exact eigenvalue relation for the even and odd solutions satisfying $y(x)\to 0$ as $x\to\infty$ is derived in terms of the characteristic exponent and the joining factors for the spheroidal functions. Numerical results illustrating the dependence of the discrete spectrum of even and odd eigenvalues on the parameters γ^2 and μ^2, together with analytical approximations in the limits of small and large positive values of γ^2, are presented and briefly discussed. Representative examples of the corresponding eigenfunctions are also given.

II. QUALITATIVE DISCUSSION OF THE SPECTRUM

We consider the simplified ideal MHD ballooning equation (1.1) expressed in the form

$$L_\gamma y=\lambda y,\qquad -\infty<x<\infty\,, \tag{2.1}$$

[a] Present address: Association Euratom—CEA, Centre d'Etudes Nucléaires de Cadarache, Boîte Postale No. 1, 13108 St. Paul-lez-Durance, France.

 0022-2488/86/082188-15$02.50

where L_γ denotes the differential operator

$$L_\gamma \equiv \frac{d}{dx}(1+x^2)\frac{d}{dx} - \gamma^2(1+x^2) + \frac{\mu^2}{1+x^2}.$$

The solutions $y(x)$ are subject to the boundary condition $y(x) \to 0$ as $x \to \pm\infty$. The parameters γ^2 and μ^2 will be supposed throughout to be real quantities so that, since L_γ is a self-adjoint operator, the λ-eigenvalues must all be real.

The nature of the spectrum of the operator L_γ depends on the sign of γ^2 but not on its absolute value. To see this, we transform the above equation to the standard Schrödinger form by making the change of variables $x = \sinh\zeta$, $\psi(\zeta) = \cosh^{1/2}\zeta\, y(x)$ to find

$$\frac{d^2\psi}{d\zeta^2} + [\Lambda - q(\zeta)]\psi = 0, \quad -\infty < \zeta < \infty, \tag{2.2}$$

where

$$\Lambda = -\lambda - \tfrac{1}{4}, \quad q(\zeta) = \gamma^2\cosh^2\zeta - (\mu^2 - \tfrac{1}{4})\operatorname{sech}^2\zeta. \tag{2.3}$$

When $\gamma^2 > 0$, $q(\zeta) \to +\infty$ as $\zeta \to \pm\infty$, and, by a well-known result,[12,13] the spectrum of Eq. (2.2), and hence that of Eq. (2.1), is an infinite pure point spectrum that is bounded below in Λ. When $\gamma^2 = 0$, $q(\zeta) \to 0$ as $\zeta \to \pm\infty$ and there is a finite point spectrum[12] (which may be null) in $-\infty < \Lambda < 0$ with a continuous spectrum in $0 < \Lambda < \infty$. We remark that, since $y(x) = O(x^{-1/2 \pm i\sqrt{\Lambda}})$ as $x \to \pm\infty$ when $\gamma^2 = 0$, the solutions belonging to the continuous spectrum do not belong to the space $L^2(-\infty, \infty)$ of Lebesgue square integrable functions on the real x axis. For $\gamma^2 < 0$, the solutions $y(x)$ behave like $\exp(\pm i|\gamma|x)/x$ as $x \to \pm\infty$ and any linear combination of the two linearly independent solutions of Eqs. (2.1) is in $L^2(-\infty, \infty)$. The spectrum in this case consequently extends continuously from $-\infty$ to ∞.

In this paper, we shall be concerned with the dependence of the point spectrum of L_γ on the parameters γ^2 and μ^2 when $\gamma^2 > 0$ (corresponding to unstable modes in the physical problem) and the behavior of this spectrum as $\gamma^2 \to 0$ through positive values. In the limiting case $\gamma^2 = 0$, Eq. (2.1) reduces to Legendre's equation of imaginary argument and has been discussed in Refs. 12 (p. 103) and 14. The discrete eigenvalues of the operator L_0 are given by

$$\lambda_n = (\mu - n)(\mu - n - 1),$$
$$n = 0,1,\dots,N, \quad N = [\mu - \tfrac{1}{2}] \quad (\gamma^2 = 0), \tag{2.4}$$

with square brackets denoting the integral part. Thus, when $\gamma^2 = 0$, there is a finite sequence of eigenvalues satisfying $-\frac{1}{4} < \lambda \leqslant \mu(\mu - 1)$, with the number of eigenvalues depending on the value of μ. When $\mu > 1/2$, the sequence always contains at least one eigenvalue given by $\lambda_0 = \mu(\mu - 1)$ and an additional eigenvalue appears each time μ increases by unity; for $\mu^2 \leqslant \frac{1}{4}$, the discrete part of the spectrum is null. This behavior may be understood qualitatively by noticing from Eq. (2.3) that, when $\gamma^2 = 0$, $q(\zeta)$ has a "potential well" of finite depth only when $\mu^2 > \frac{1}{4}$.

The eigenfunctions associated with Eq. (2.4) may be conveniently expressed in terms of terminating hypergeometric functions. Since L_γ is unchanged if we replace x by $-x$, it is possible to consider even and odd solutions separately, which we shall denote by the superscripts e and o, respectively. The even and odd solutions when $\gamma^2 = 0$ are

$$\begin{aligned} y^{(e)}(x) &= (1+x^2)^{-(1/2)\mu} \\ &\quad \times {}_2F_1(-\tfrac{1}{2}\nu - \tfrac{1}{2}\mu, \tfrac{1}{2} + \tfrac{1}{2}\nu - \tfrac{1}{2}\mu; \tfrac{1}{2}; -x^2) \\ &= (1+x^2)^{-1/2-(1/2)\nu} \\ &\quad \times {}_2F_1(\tfrac{1}{2} + \tfrac{1}{2}\nu + \tfrac{1}{2}\mu, \tfrac{1}{2} + \tfrac{1}{2}\nu - \tfrac{1}{2}\mu; \tfrac{1}{2}; x^2/(1+x^2)), \end{aligned} \tag{2.5}$$

$$\begin{aligned} y^{(o)}(x) &= x(1+x^2)^{-(1/2)\mu} \\ &\quad \times {}_2F_1(\tfrac{1}{2} - \tfrac{1}{2}\nu - \tfrac{1}{2}\mu, 1 + \tfrac{1}{2}\nu - \tfrac{1}{2}\mu; \tfrac{3}{2}; -x^2) \\ &= x(1+x^2)^{-1-(1/2)\nu} \\ &\quad \times {}_2F_1(1 + \tfrac{1}{2}\nu + \tfrac{1}{2}\mu, 1 + \tfrac{1}{2}\nu - \tfrac{1}{2}\mu; \tfrac{3}{2}; x^2/(1+x^2)), \end{aligned} \tag{2.6}$$

where the parameter ν is defined by

$$\nu = -\tfrac{1}{2} + (\lambda + \tfrac{1}{4})^{1/2} \quad (\gamma^2 = 0). \tag{2.7}$$

Since $\nu > -\frac{1}{2}$ for the discrete spectrum, these solutions are seen to diverge at $x = \pm\infty$ unless, in Eq. (2.5), $\nu = \mu - 1 - 2n$, $n = 0,1,\dots,N^{(e)}$, $N^{(e)} = [\frac{1}{2}(\mu - \frac{1}{2})]$ and in Eq. (2.6) $\nu = \mu - 2 - 2n$, $n = 0,1,\dots,N^{(o)}$, $N^{(o)} = [\frac{1}{2}(\mu - \frac{3}{2})]$. The even and odd eigenfunctions and eigenvalues of L_0 associated with the discrete spectrum are therefore given by

$$y_n^{(e)}(x) = (1+x^2)^{n-(1/2)\mu}\,{}_2F_1(-n, \mu - n; \tfrac{1}{2}; x^2/(1+x^2)),$$
$$n = 0,1,\dots,[\tfrac{1}{2}(\mu - \tfrac{1}{2})], \tag{2.8}$$
$$\lambda_n^{(e)} = (\mu - 2n)(\mu - 2n - 1) \quad (\mu > \tfrac{1}{2}),$$

$$y_n^{(o)}(x) = x(1+x^2)^{n-(1/2)\mu}\,{}_2F_1(-n, \mu - n; \tfrac{3}{2}; x^2/(1+x^2)),$$
$$n = 0,1,\dots[\tfrac{1}{2}(\mu - \tfrac{3}{2})], \tag{2.9}$$
$$\lambda_n^{(o)} = (\mu - 2n - 1)(\mu - 2n - 2) \quad (\mu > \tfrac{3}{2}).$$

The lowest even and odd eigenfunctions and eigenvalues when $\gamma^2 = 0$ are

$$\begin{aligned} y_0^{(e)}(x) &= (1+x^2)^{-(1/2)\mu}, \\ \lambda_0^{(e)} &= \mu(\mu - 1) \quad (\mu > \tfrac{1}{2}), \\ y_0^{(o)}(x) &= x(1+x^2)^{-(1/2)\mu}, \\ \lambda_0^{(o)} &= (\mu - 1)(\mu - 2) \quad (\mu > \tfrac{3}{2}). \end{aligned} \tag{2.10}$$

Since the hypergeometric functions appearing in Eqs. (2.8) and (2.9) are polynomials of degree n in $x^2/(1+x^2)$, the eigenfunctions, when $\gamma^2 = 0$ and $\mu - \frac{1}{2}$ is not equal to a positive integer, are always in $L^2(-\infty, \infty)$. When $\mu - \frac{1}{2}$ is either an even or an odd integer, however, the eigenfunction corresponding to $n = N^{(e)}$ or $N^{(o)}$, respectively, contains the multiplicative factor $(1+x^2)^{-1/4}$ or $x(1+x^2)^{-3/4}$. This particular eigenfunction therefore is no longer in $L^2(-\infty, \infty)$ and is associated with the eigenvalue $\lambda = -\frac{1}{4}$, which lies at the end point of the continuous (non-L^2) spectrum in $-\infty < \lambda < -\frac{1}{4}$.

Simple bounds on the eigenvalues for $\gamma^2 > 0$ may be derived from consideration of the quadratic forms associated with either Eq. (2.1) or Eq. (2.3). When $\gamma^2 > 0$, the eigenfunctions behave like $\exp(-|\gamma x|)/|x|$ as $x \to \pm\infty$ and we consequently have

$$\int_{-\infty}^{\infty}\left[(1+x^2)\left(\frac{dy}{dx}\right)^2+\left\{\lambda+\gamma^2(1+x^2)-\frac{\mu^2}{1+x^2}\right\}y^2\right]dx$$
$$=\int_{-\infty}^{\infty}\left[\left(\frac{d\psi}{d\zeta}\right)^2+\left(\lambda+\frac{1}{4}\right)\psi^2+q(\zeta)\psi^2\right]d\zeta=0\,.$$

For $\mu^2 \leqslant \frac{1}{4}$, or $\mu^2 > \frac{1}{4}$ and $\gamma^2 \geqslant \mu^2 - \frac{1}{4}$, $q(\zeta)$ is non-negative on $(-\infty,\infty)$. Hence, the eigenvalues must satisfy the condition

$$-\lambda > \tfrac{1}{4} \quad (\mu^2 \leqslant \tfrac{1}{4};\ \ \mu^2 > \tfrac{1}{4},\ \ \gamma^2 \geqslant \mu^2 - \tfrac{1}{4})\,. \qquad (2.11)$$

By suitable rearrangement of the "potential" function $q(\zeta)$ as

$$q(\zeta)=\gamma^2-\mu^2+\tfrac{1}{4}+[\gamma^2\sinh^2\zeta+(\mu^2-\tfrac{1}{4})\tanh^2\zeta\,]$$
$$=2\gamma(\tfrac{1}{4}-\mu^2)^{1/2}+[\gamma\cosh\zeta-(\tfrac{1}{4}-\mu^2)^{1/2}\operatorname{sech}\zeta\,]^2$$
$$(\mu^2 \leqslant \tfrac{1}{4})\,,$$

we then immediately deduce the bounds

$$-\lambda > \gamma^2-\mu^2+\tfrac{1}{2} \quad (\mu^2 \geqslant \tfrac{1}{4};\ \ \mu^2<\tfrac{1}{4},\ \ \gamma^2 \geqslant \tfrac{1}{4}-\mu^2)\,,$$
$$-\lambda > \tfrac{1}{4}+2\gamma(\tfrac{1}{4}-\mu^2)^{1/2} \quad (\mu^2<\tfrac{1}{4},\ \ 0<\gamma^2<\tfrac{1}{4}-\mu^2)\,. \qquad (2.12)$$

Note that the above rearrangements of $q(\zeta)$ correspond to separating off the minimum value of $q(\zeta)$ in the two different cases which arise. For, when $\mu^2 \geqslant \frac{1}{4}$, or when $\mu^2 < \frac{1}{4}$ and $\gamma^2 \geqslant \frac{1}{4}-\mu^2$, $q(\zeta)$ possesses a single minimum at $\zeta = 0$ with $q(0)=\gamma^2-\mu^2+\frac{1}{4}$. When $\mu^2<\frac{1}{4}$ and $0<\gamma^2<\frac{1}{4}-\mu^2$, $q(\zeta)$ possesses in addition two symmetric minima at $\zeta = \pm\zeta_0$, where $\gamma\cosh^2\zeta_0 = (\frac{1}{4}-\mu^2)^{1/2}$, with $q(\pm\zeta_0)=2\gamma(1/4-\mu^2)^{1/2}$ and the point $\zeta=0$ now being a local maximum (see Fig. 1).

The above bounds on the point spectrum are illustrated in Fig. 2 which shows $-\lambda$ as a function of γ^2 with μ^2 as parameter. The inequality Eq. (2.11) indicates that the eigenvalues $-\lambda_n$ $(n=0,1,2,\ldots)$ for $\gamma^2>\max(0,\mu^2-\frac{1}{4})$ must be located in the upper right-hand quadrant with vertex at $(\frac{1}{4},\max(0,\mu^2-\frac{1}{4}))$. The conditions in Eq. (2.12) then define lower bounds on $-\lambda_n$ in this quadrant, where for $\gamma^2>0$ when $\mu^2 \geqslant \frac{1}{4}$ or $\gamma^2 \geqslant \frac{1}{4}-\mu^2$ when $\mu^2<\frac{1}{4}$, they form, for different μ^2, a family of parallel lines of unit positive slope. In the region $0<\gamma^2<\frac{1}{4}-\mu^2$ when $\mu^2<\frac{1}{4}$, the bounds are portions of a family of parabolas with common origin at the point $(\frac{1}{4},0)$, which join "smoothly" (i.e., in value and first derivative) with those in $\gamma^2 \geqslant \frac{1}{4}-\mu^2$. The points of intersection between the two families of curves ($\mu^2<\frac{1}{4}$) lie on the straight line $-\lambda=\frac{1}{4}+2\gamma^2$, which is shown dashed in Fig. 2. This figure clearly suggests that the eigenvalue curves $\lambda(\gamma^2)$ assume a different character according as to whether the value of μ^2 is less than or greater than the Mathieu function case, $\mu^2=\frac{1}{4}$.

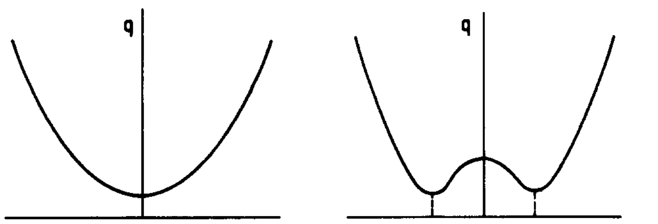


FIG. 1. The "potential" function $q(\zeta)$ for (a) $\mu^2 \geqslant \frac{1}{4}$, or $\mu^2<\frac{1}{4}$ and $\gamma^2 \geqslant \frac{1}{4}-\mu^2$, and (b) $\mu^2<\frac{1}{4}$ and $0<\gamma^2<\frac{1}{4}-\mu^2$.

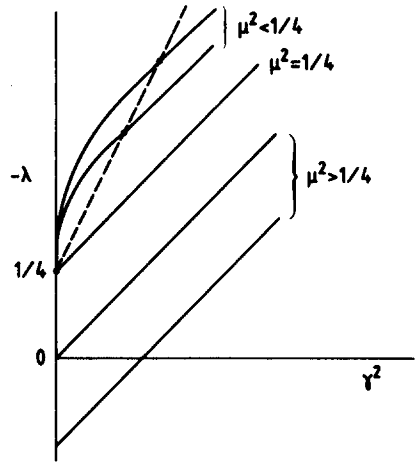


FIG. 2. The lower bounds in Eq. (2.12) on the eigenvalues $-\lambda$ as a function of γ^2 for different values of μ^2. When $\mu^2<\frac{1}{4}$ the bounds pass through the point $-\lambda=\frac{1}{4}$, $\gamma^2=0$.

To conclude this preliminary discussion of the nature of the spectrum, we note that the operator L_γ can be considered as a perturbation of the operator L_0 (with $\gamma^2=0$) associated with Legendre's differential equation of imaginary argument. The spectrum of L_γ is thus seen to be nonuniform in the limit $\gamma^2 \to 0+$: when $\gamma^2>0$, we have a purely point spectrum with an infinite number of eigenvalues for any μ^2 while, when $\gamma^2=0$, the spectrum is finite in $-\infty<-\lambda<\frac{1}{4}$, given by Eq. (2.4), and continuous in $\frac{1}{4}<-\lambda<\infty$. We shall see that the point $-\lambda=\frac{1}{4}$, $\gamma^2=0$, is, in fact, an accumulation point of the spectrum. This concentration of the eigenvalues in the neighborhood of $\gamma^2=0$ can be considered from the viewpoint of perturbation theory for linear operators, as described by Kato[15] and discussed in a specific example by Titchmarsh.[16] We shall find that, although L_γ is a relatively unbounded perturbation of L_0, the change of the finite part of the spectrum of L_0 in $-\infty<-\lambda<\frac{1}{4}$ is nevertheless continuous as γ^2 increases continuously from zero.

III. THE EIGENVALUE RELATION FOR $\gamma^2>0$

We now proceed to derive an exact eigenvalue relation for the point spectrum of the operator L_γ in Eq. (2.1) when $\gamma^2>0$. We shall adopt throughout the convention that when γ^2 (or μ^2) is positive, we take γ (or μ) to be positive. Equation (2.1) is the oblate spheroidal differential equation of order μ, expressed in the canonical form

$$\frac{d}{dz}\left[(1-z^2)\frac{dy}{dz}\right]+\left[\lambda+\gamma^2(1-z^2)-\frac{\mu^2}{1-z^2}\right]y=0\,, \qquad (3.1)$$

when $z=ix$. The standard solutions of this equation for general λ, γ^2, and μ^2 are known as the spheroidal functions.[17,18] A detailed review of the theory of these functions required for the ballooning eigenvalue problem has been given in Ref.

10 and their principal properties are briefly summarized in the Appendix.

The spheroidal functions of the first group involve expansions in series of Legendre functions and are defined by[17]

$$Ps_\nu^\mu(z;\gamma) = \sum_{r=-\infty}^{\infty} (-)^r a_{\nu,2r}^\mu(\gamma^2) P_{\nu+2r}^\mu(z),$$
$$Qs_\nu^\mu(z;\gamma) = \sum_{r=-\infty}^{\infty} (-)^r a_{\nu,2r}^\mu(\gamma^2) Q_{\nu+2r}^\mu(z) \quad (3.2)$$
$$(\mu+\nu \neq 0, \pm 1, \pm 2,\ldots),$$

where $a_{\nu,2r}^\mu(\gamma^2)$ are coefficients independent of z. These functions converge uniformly and absolutely in any domain not including the points $z = \pm 1$ and $z = \infty$. It is assumed that $\mu + \nu$ is not an integer in the definition of $Qs_\nu^\mu(z;\gamma)$, since the Legendre function $Q_{\nu+2r}^\mu$ is then no longer defined when $\mu + \nu + 2r$ is a negative integer. The spheroidal functions of the second group involve the expansions in series of Bessel functions

$$S_\nu^{\mu(j)}(z;\gamma) = (1-z^{-2})^{-(1/2)\mu}[A_\nu^\mu(\gamma^2)]^{-1}\left(\frac{\pi}{2\gamma z}\right)^{1/2} \times \sum_{r=-\infty}^{\infty} a_{\nu,2r}^\mu(\gamma^2)\mathscr{C}_{\nu+1/2+2r}^{(j)}(\gamma z)$$
$$(j=1,2,3,4), \quad (3.3)$$

where

$$\mathscr{C}_\nu^{(1)} = J_\nu(z), \quad \mathscr{C}_\nu^{(2)} = Y_\nu(z), \quad \mathscr{C}_\nu^{(3,4)}(z) = H_\nu^{(1,2)}(z).$$

These solutions are valid in the neighborhood of the point at infinity and are uniformly and absolutely convergent for $|z|>1$. The coefficients $a_{\nu,2r}^\mu(\gamma^2)$ satisfy the three-term recurrence relation given in Eq. (A1) and are normalized such that $a_{\nu,0}^\mu(0)=1$. The quantity $A_\nu^\mu(\gamma^2)$ is a normalizing factor defined in Eq. (A10), so chosen that the solutions $S_\nu^{\mu(j)}(z;\gamma)$ possess the leading asymptotic behavior $(\pi/2\gamma z)^{1/2}\mathscr{C}_{\nu+1/2}(\gamma z)$ as $|z|\to\infty$ in $|\arg(\gamma z)|<\pi$.

The parameter ν appearing in the above expansions is known as the characteristic exponent of Eq. (3.1) and is analogous to that encountered in the solutions of Mathieu's equation.[19] The value of ν depends on the parameters λ,γ^2, and μ^2 and is defined by Eq. (2.7) [or Eq. (A7)] *only when* $\gamma^2=0$. The determination of ν as a function of λ, γ^2, and μ^2 is obtained by the requirement of a nontrivial solution for the coefficients $a_{\nu,2r}^\mu(\gamma^2)$. This functional dependence is expressed by the condition

$$\Delta_{\mu,\nu}(\lambda,\gamma^2)=0, \quad (3.4)$$

where $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ is the infinite determinant defined in Eq. (A3). For sufficiently small γ^2, the expansion of λ as a series in ascending powers of γ^2, with coefficients depending on ν and μ, is given in Eq. (A5).

We consider even and odd solutions of Eq. (1.1) and consequently restrict our attention to the interval $[0,\infty)$. When $\gamma^2>0$, the solution associated with the characteristic exponent ν and possessing the correct behavior as $x\to+\infty$ is

$$y(x) = (-\tfrac{1}{2}\pi e^{(1/2)\pi i\nu})S_\nu^{\mu(3)}(ix;\gamma)$$
$$= (1+x^2)^{-(1/2)\mu}[A_\nu^\mu(\gamma^2)]^{-1}\left(\frac{\pi}{2\gamma x}\right)^{1/2} \times \sum_{r=-\infty}^{\infty}(-)^r a_{\nu,2r}^\mu(\gamma^2)K_{\nu+1/2+2r}(\gamma x) \quad (x>1), \quad (3.5)$$

since, for $\gamma x>0$, the Hankel function can be expressed in terms of the modified Bessel function of the second kind by

$$H_\nu^{(1)}(i\gamma x) = (2/\pi i)e^{-(1/2)\pi i\nu}K_\nu(\gamma x).$$

From Eq. (A11), this solution then has the leading asymptotic behavior as $x\to+\infty$

$$y(x)\sim\left(\frac{\pi}{2\gamma x}\right)^{1/2}K_{\nu+1/2}(\gamma x)\sim\frac{\pi}{2\gamma x}\exp(-\gamma x). \quad (3.6)$$

The continuation of this solution onto the interval [0,1] is given by Eq. (A15) as

$$y(x) = (-\tfrac{1}{2}\pi e^{\pi i\nu/2})\{APs_\nu^\mu(ix;\gamma)+BQs_\nu^\mu(ix;\gamma)\}, \quad (3.7)$$

where the constants A and B involve the joining factors and are defined in Eq. (A17). The even and odd eigenvalue relations are then obtained by requiring that $y(x)$ be, respectively, an even or odd function of x. This is equivalent to the conditions that the continuation of the spheroidal function $S_\nu^{\mu(3)}(z;\gamma)$ in Eq. (A15) be either an even or odd function of z.

The definition of the Legendre functions in terms of Gauss hypergeometric functions in $|z^2|<1$ [with $\mathrm{Im}(z)>0$ since $x>0$] is

$$P_{\nu+2r}^\mu(z) = 2^\mu\pi^{1/2}(z^2-1)^{-(1/2)\mu}\{aF_1(z^2)+bzF_2(z^2)\},$$
$$Q_{\nu+2r}^\mu(z) = 2^\mu\pi^{1/2}(z^2-1)^{-(1/2)\mu}e^{\pi\mu i}\{cF_1(z^2)+dzF_2(z^2)\},$$

where

$$F_1(z^2)\equiv{}_2F_1(-\tfrac{1}{2}\nu-\tfrac{1}{2}\mu-r,\tfrac{1}{2}+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r;\tfrac{1}{2};z^2),$$
$$F_2(z^2)\equiv{}_2F_1(\tfrac{1}{2}-\tfrac{1}{2}\nu-\tfrac{1}{2}\mu-r,1+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r;\tfrac{3}{2};z^2),$$

and

$$a=[\Gamma(\tfrac{1}{2}-\tfrac{1}{2}\nu-\tfrac{1}{2}\mu-r)\Gamma(1+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r)]^{-1},$$
$$b=-2[\Gamma(\tfrac{1}{2}+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r)\Gamma(-\tfrac{1}{2}\nu-\tfrac{1}{2}\mu-r)]^{-1},$$
$$c=\tfrac{1}{2}e^{\pi i(\mu-\nu-1-2r)/2}[\Gamma(\tfrac{1}{2}+\tfrac{1}{2}\nu+\tfrac{1}{2}\mu+r)/\Gamma(1+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r)],$$
$$d=e^{\pi i(\mu-\nu-2r)/2}[\Gamma(1+\tfrac{1}{2}\nu+\tfrac{1}{2}\mu+r)/\Gamma(\tfrac{1}{2}+\tfrac{1}{2}\nu-\tfrac{1}{2}\mu+r)].$$

The conditions for the right-hand side of Eq. (A16) to be an even or odd function of z, given by the vanishing of the coefficients of $zF_2(z^2)$ and $F_1(z^2)$ in the summand, are consequently $Ab+Bde^{\pi\mu i}=0$ and $Aa+Bce^{\pi\mu i}=0$, respectively; that is

$$A+\tfrac{1}{2}\pi\frac{e^{(1/2)\pi i(3\mu-\nu)}\{{}^1_i\}B}{\{{}^{\sin}_{\cos}\}\tfrac{1}{2}\pi(\mu+\nu)}=0, \quad \begin{cases}\text{even,}\\ \text{odd.}\end{cases}$$

Substitution of the values of A and B from Eq. (A17) then yields[20]

$$K_\nu^\mu(\gamma) + e^{\pi i(\nu - 1/2)}\left[\frac{1 \pm e^{\pi i(\mu+\nu)}}{1 \mp e^{\pi i(\mu-\nu)}}\right]K^\mu_{-\nu-1}(\gamma) = 0\,,$$

$$\mu \pm \nu \neq 0, \pm 1, \pm 2,\ldots, \quad \nu + \tfrac{1}{2} \neq 0, \pm 1, \pm 2,\ldots,$$

where the upper or lower signs correspond to the even or odd solution, respectively.

Equation (3.8) is the desired eigenvalue relation for the ballooning problem and describes the dependence of the characteristic exponent ν on μ and γ for the even and odd eigenvalues of Eq. (2.1). This dependence is expressed in terms of the joining factors $K_\nu^\mu(\gamma)$ and $K^\mu_{-\nu-1}(\gamma)$ for the spheroidal functions defined in Eqs. (A12) and (A13). We remark that, although the eigenvalue λ does not appear explicitly in Eq. (3.8), it is contained implicitly through the dependence of the characteristic exponent ν on λ given by the functional relation [Eq. (3.4)]. For most of the range of parameters, ν is found to be complex in Eq. (3.8) with $\mathrm{Re}(\nu) = -\frac{1}{2}$ [so that λ defined by Eq. (A5) is real in accordance with the fact that L_γ is self-adjoint]. As mentioned in Sec. II, this behavior may be contrasted with that found in the case of the spheroidal wave functions, where the poloidal angular function satisfies an equation of type Eq. (3.1) on the interval $z\in[-1,1]$, while the radial function satisfies an equation of type Eq. (2.1). The parameter μ in this latter case corresponds to the toroidal mode number and is consequently an integer m. The requirement that the angular solution be bounded at $z = \pm 1$ then necessarily restricts the characteristic exponent ν also to be an integer, n ($\geqslant m$) (see Refs. 11 and 17). Curves of λ as a function of γ^2 for different values of m and n for the wave functions have been obtained by Meixner.[21]

IV. SOLUTION OF THE EIGENVALUE RELATION

To express the eigenvalue relations Eq. (3.8) in a form more suitable for numerical computation, we employ the definition of the joining factor $K_\nu^\mu(\gamma)$ given in Eq. (A13) to find

$$\left(\frac{\gamma}{4}\right)^{2\nu+1}\frac{\Gamma(\mu+\nu+1)}{\Gamma(\mu-\nu)}\left[\frac{\Gamma(\frac{1}{2}-\nu)}{\Gamma(\frac{3}{2}+\nu)}\right]^2$$

$$\times H_{\mu,\nu}(\gamma) \pm \frac{\left\{\begin{smallmatrix}\cos\\ \sin\end{smallmatrix}\right\}\frac{1}{2}\pi(\mu-\nu)}{\left\{\begin{smallmatrix}\sin\\ \cos\end{smallmatrix}\right\}\frac{1}{2}\pi(\mu+\nu)} = 0\,, \qquad (4.1)$$

where we have defined

$$H_{\mu,\nu}(\gamma) \equiv \frac{S_-(\mu,-\nu-\frac{1}{2})S_-(-\mu,-\nu-\frac{1}{2})}{S_+(\mu,\nu+\frac{1}{2})S_+(-\mu,\nu+\frac{1}{2})}\,,$$

$$S_\pm(\mu,\alpha) \equiv \sum_{r=0}^{\infty} a^\mu_{\nu,\pm 2r}(\gamma^2)\frac{(\alpha)_r}{r!}\,, \qquad (4.2)$$

and employed the result $(1-\alpha)_{-r} = (-)^r/(\alpha)_r$ with $(\alpha)_r = \Gamma(\alpha + r)/\Gamma(\alpha)$. Employing the properties of the gamma function, we may finally express the eigenvalue relations in the more symmetrical form

$$F^{(e,o)}(\tilde\nu)$$

$$\equiv\left(\frac{1}{2}\gamma\right)^{2\tilde\nu}\frac{\Gamma(\frac{1}{2}\mp\delta+\frac{1}{2}\mu+\frac{1}{2}\tilde\nu)\Gamma(\frac{1}{2}\mp\delta-\frac{1}{2}\mu+\frac{1}{2}\tilde\nu)}{\Gamma(\frac{1}{2}\mp\delta+\frac{1}{2}\mu-\frac{1}{2}\tilde\nu)\Gamma(\frac{1}{2}\mp\delta-\frac{1}{2}\mu-\frac{1}{2}\tilde\nu)}$$

$$\times\left[\frac{\Gamma(1-\tilde\nu)}{\Gamma(1+\tilde\nu)}\right]^2 H_{\mu,\nu}(\gamma) - 1 = 0\,, \qquad (4.3)$$

where

$$\tilde\nu = \nu + \tfrac{1}{2}\,, \quad \delta = \tfrac{1}{4}\,,$$

and the upper and lower signs correspond to the even and odd eigenvalues, respectively.

From the definition of $H_{\mu,\nu}(\gamma)$ and the properties of the coefficients $a^\mu_{\nu,2r}(\gamma^2)$ in Eq. (A9), it is readily seen that

$$H_{\mu,\nu}(\gamma) = H_{-\mu,\nu}(\gamma), \quad H_{\mu,\nu}(\gamma)H_{\mu,-\nu-1}(\gamma) = 1\,.$$

The zeros of $F^{(e,o)}(\tilde\nu)$ are therefore invariant when μ is replaced by $-\mu$ and $\tilde\nu$ by $-\tilde\nu$ (i.e., $\nu \to -\nu - 1$). This confirms the independence of the eigenvalues on the sign of μ and whether the characteristic exponent is taken as ν or $-\nu - 1$ [see Eq. (A6)].

In the complex $\tilde\nu$ plane, symmetric pairs of zeros of $F^{(e,o)}(\tilde\nu)$ are found on the real and imaginary axes. The number of zeros on the imaginary axis is infinite, while, for sufficiently small γ^2, the number of real zeros is finite when $\mu^2 > \frac{1}{4}$ and null when $\mu^2 \leqslant \frac{1}{4}$. When $\gamma^2 = 0$ the real zeros are situated at the points $\tilde\nu = \pm(\mu - \frac{1}{2} - k)$, $k = 0,1,\ldots,[\mu - \frac{1}{2}]$ [even and odd k being associated with $F^{(e,o)}(\tilde\nu)$, respectively] and correspond to the eigenvalues in Eq. (2.4). As γ^2 steadily increases, these pairs of real zeros approach the origin and then successively move off the real axis as conjugate pairs on the imaginary axis.

We observe from Eq. (A5) that only the zeros on the real and imaginary axes can lead to real values of λ. This is evident when the zeros are real. To see this in the case of imaginary zeros, we write $\tilde\nu = iu$ ($u > 0$) so that

$$\nu = -\tfrac{1}{2} + iu, \quad u = u(\gamma,\mu)\,. \qquad (4.4)$$

Then, for arbitrary integer r and real values of γ^2 and μ^2, the coefficients $A_r(\mu,\nu)$, $B_r(\mu,\nu)$, and $C_r(\mu,\nu)$, which appear in the recurrence relation [Eq. (A1)], satisfy the conjugate relations

$$A_r(\mu,\nu) = B^*_{-r}(\mu,\nu) \quad (\mu^2 \geqslant 0)\,,$$

$$A_r(\mu,\nu) = B^*_{-r}(-\mu,\nu) \quad (\mu^2 < 0)\,,$$

with $C_r(\mu,\nu) = C^*_{-r}(\mu,\nu)$ in both cases. From the symmetry of the elements of the determinant [Eq. (A3)] with respect to the central element, $\Delta_{\mu,\nu}(\lambda,\gamma^2) = \Delta^*_{\mu,\nu}(\lambda^*,\gamma^2)$ and consequently $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ is a real function when λ is real. Considered as a function of λ, $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ is analytic except at its simple poles situated at $\lambda = C_0(\mu,\nu)$ and the conjugate points $\lambda = C_{\pm r}(\mu,\nu)$ ($r = 1,2,\ldots$). Since $\Delta_{\mu,\nu}(\lambda,\gamma^2) \to 1$ as $\lambda \to \pm\infty$ and the simple pole at $\lambda = C_0(\mu,\nu)$ lies on the real axis, it follows that there is always at least one real zero of $\Delta_{\mu,\nu}(\lambda,\gamma^2)$.

For sufficiently small values of γ^2 it is certain that $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ has one real zero near $C_0(\mu,\nu)$ (corresponding to the expansion [Eq. (A5)]), together with conjugate pairs of zeros located near the points $C_{\pm r}(\mu,\nu)$, $r = 1,2,\ldots$. It seems difficult to prove, however, that there is always just one real zero for all values of γ^2. There is strong evidence, both from numerical computations and analysis of the root loci for small γ^2, to suggest that as γ^2 increases the zero described by Eq. (A5) remains on the real axis, while the loci of the complex zeros move *away* from the real axis. In what follows we shall assume this to be the case.

Inspection of the recurrence relation [Eq. (A1)] then shows that, when ν is defined by Eq. (4.4), the coefficients $a^{\mu}_{\nu,2r}(\gamma^2)$ satisfy the conjugate relations

$$\begin{aligned} a^{\pm\mu}_{\nu,-2r}(\gamma^2) &= (a^{\pm\mu}_{\nu,2r}(\gamma^2))^* \quad (\mu^2 \geqslant 0), \\ a^{\pm\mu}_{\nu,-2r}(\gamma^2) &= (a^{\mp\mu}_{\nu,2r}(\gamma^2))^* \quad (\mu^2 < 0), \end{aligned} \tag{4.5}$$

and hence that $|H_{\mu,\nu}(\gamma)| = 1$. Introducing the real phase function $\phi = \phi(\gamma,\mu,u)$ by

$$H_{\mu,\nu}(\gamma) = \exp(2i\phi), \quad \nu = -\tfrac{1}{2} + iu, \tag{4.6}$$

we can write the even and odd eigenvalue relations in the form

$$G^{(e,o)}(u) \equiv u \ln \tfrac{1}{2}\gamma + \phi + \Theta^{(e,o)} + (n+1)\pi = 0, \quad n = 0,1,2,\ldots, \tag{4.7}$$

where

$$\Theta^{(e,o)} = \arg[\Gamma(\tfrac{1}{2} \mp \delta + \tfrac{1}{2}\mu + \tfrac{1}{2}iu)\Gamma(\tfrac{1}{2} \mp \delta - \tfrac{1}{2}\mu + \tfrac{1}{2}iu)] - 2\arg\Gamma(1+iu). \tag{4.8}$$

The restriction of the integer n in Eq. (4.7) to non-negative values can be argued as follows. From Eq. (2.12) and Fig. 2, we observe that as $-\lambda \to \frac{1}{4}$ the admissible values of γ^2 must also approach zero when $\mu^2 < \frac{1}{4}$. The functional relation [Eq. (A5)] then shows that under these circumstances u also approaches zero. The eigenvalue relations in Eq. (4.7) yield $\gamma^2 = O[\exp(-2\pi(n+1)/u)]$ as $u \to 0$ when $\mu^2 < \frac{1}{4}$, since, for $\mu^2 < \frac{1}{4}$, $\Theta^{(e,o)}$ and ϕ are both $O(u)$ in this limit. Hence n must take the values 0,1,2,... and Eq. (4.7) then defines the dependence of $\nu = -\frac{1}{2} + iu$ on the parameters γ and μ corresponding to the even and odd eigenvalues $\lambda^{(e,o)}_n$, $n = 0,1,2,\ldots$. The point $-\lambda = \frac{1}{4}$, $\gamma^2 = 0$ is therefore an accumulation point of the spectrum. This has been shown previously in the case $\mu = 0$ by Kulsrud.[22]

An alternative representation of $\Theta^{(e,o)}$ suitable when $\mu > \frac{1}{2}$ and $\mu > \frac{3}{2}$, respectively, which does not involve a gamma function whose argument has a negative real part, is given by

$$\Theta^{(e,o)} = -u \ln 2 + \arg\Gamma(\tfrac{1}{2} + \mu + iu) - 2\arg\Gamma(1+iu) - \theta^{(e,o)} \tag{4.9}$$

with

$$\begin{aligned} \theta^{(e,o)} &= \arg\cos\pi(\tfrac{1}{2}\mu \pm \tfrac{1}{4} - \tfrac{1}{2}iu) \\ &= \arctan[\tanh\tfrac{1}{2}\pi u \tan\tfrac{1}{2}\pi(\mu \pm \tfrac{1}{2})] + m(\mu)\pi, \end{aligned} \tag{4.10}$$

respectively, where the inverse tangent takes its principal value. The integer $m(\mu)$ is the "winding" number defined by

$$\begin{aligned} m(\mu) &= 0, \quad \tfrac{1}{2}\mu - \tfrac{1}{2} \pm \delta \leqslant 0, \\ m(\mu) &= k, \quad k-1 < \tfrac{1}{2}\mu - \tfrac{1}{2} \pm \delta \leqslant k \quad (k = 1,2,\ldots), \end{aligned} \tag{4.11}$$

where the upper or lower sign corresponds, respectively, to the even or odd mode and, as before, $\delta = \frac{1}{4}$. We observe that as $u \to 0$ the phase angles $\Theta^{(e)}$ ($\mu > \frac{1}{2}$) and $\Theta^{(o)}$ ($\mu > \frac{3}{2}$) are no longer $O(u)$ but approach the values $m(\mu)\pi$.

V. RESULTS AND DISCUSSION

The eigenvalue relations [Eqs. (4.3) and (4.7)] have been solved numerically for different values of γ^2 with μ^2 as the parameter. For given values of μ^2 and γ^2, we guess an initial value of the characteristic exponent ν; for small γ^2, this initial guess is made by employing the approximate analytical formulas developed in Sec. VI. The value of λ corresponding to these values of ν,μ, and γ^2 is first estimated by means of the expansion [Eq. (A5)] and subsequently refined by iterative solution of the functional equation (A2). For this latter calculation, we truncate the infinite determinant $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ in Eq. (A3) and work with a determinant of dimension $2M+1$ ($M = 2,3,\ldots$) based on the central element. The coefficients $a^{\mu}_{\nu,2r}(\gamma^2)$ ($-M \leqslant r \leqslant M$) are then obtained by solution of the recurrence relation [Eq. (A1)] expressed as truncated continued fractions (cf. Refs. 10 and 17, pp. 102 and 136). The normalizing factors $A^{\pm\mu}_{\nu}(\gamma^2)$ and the joining factors $K^{\mu}_{\nu}(\gamma), K^{\mu}_{-\nu-1}(\gamma)$, defined in Eqs. (A10) and (A13), are then evaluated as truncated sums, with the precision of the calculation being verified by means of the identity [Eq. (A14)]. The function $H_{\mu,\nu}(\gamma)$ in Eq. (4.2) is likewise determined and, according as ν is real or complex, the values of $F^{(e,o)}(\tilde{\nu})$ and $G^{(e,o)}(u)$ in Eqs. (4.3) and (4.7) are calculated.

These calculations are repeated for a set of neighboring values of ν until the corresponding functions $F^{(e,o)}$ or $G^{(e,o)}$ change sign, with the zero subsequently found by iteration. The eigenvalue λ corresponding to this value of ν is then obtained as described above. This procedure is repeated using successively larger values of M until there is no further change in the value of λ to within the prescribed accuracy.

This procedure is found to work satisfactorily for values of γ^2 not too large. It is clear that as γ^2 increases the greater must be the value of M for a given accuracy. For $\gamma^2 \gg 1$, this process becomes unwieldy and the eigenvalues in this limit have been calculated by direct numerical solution of Eq. (1.1). Solution in this case is facilitated by making the change of variable

$$t = x/(1+x)$$

in Eq. (1.1), so that the interval $[0,\infty)$ is mapped onto the finite interval $[0,1)$ and application of the boundary condition at infinity becomes much easier to apply. The differential equation (1.1) is then written in terms of the new variable t in the form

$$\frac{d}{dt}Y(t) = A(t)Y(t), \quad 0 \leqslant t < 1,$$

where $Y(t) = (y(t), y'(t))^T$ and $A(t)$ denotes the 2×2 ma-

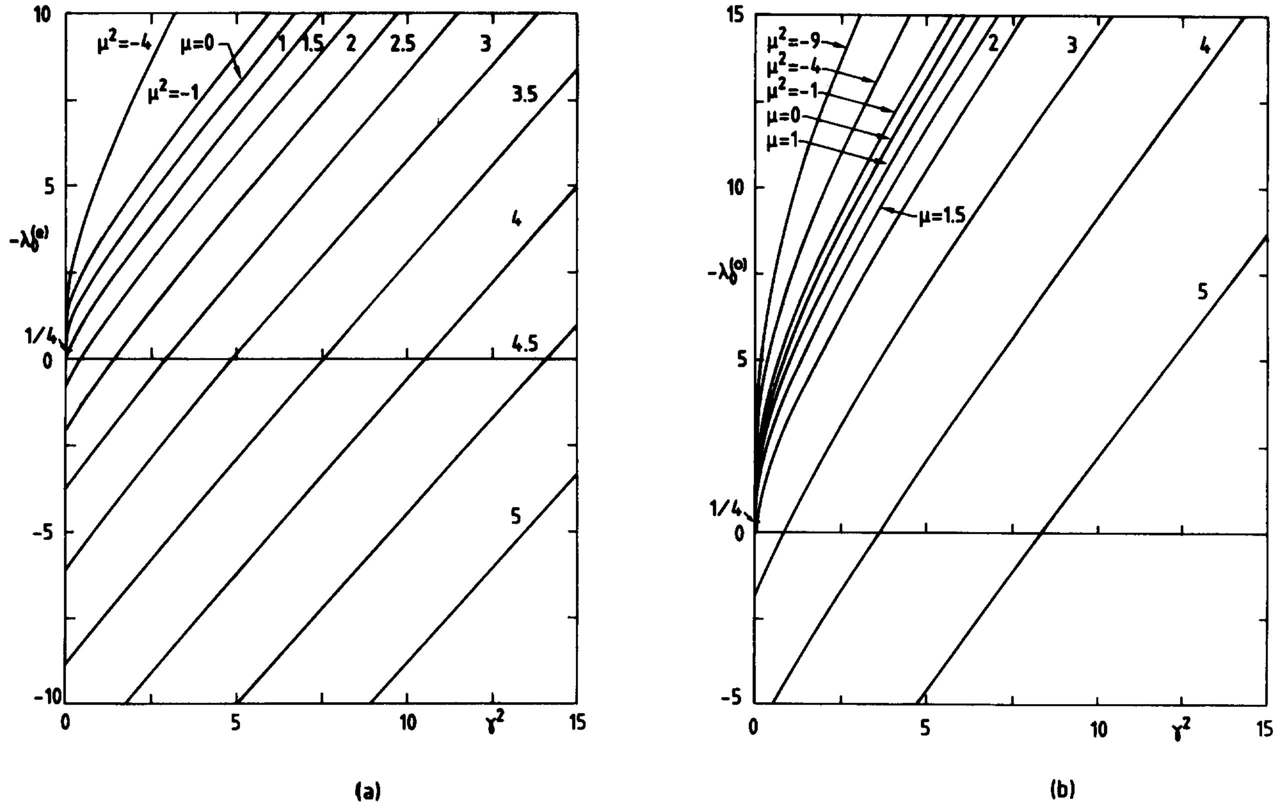


FIG. 3. The behavior of the lowest ($n=0$) (a) even and (b) odd eigenvalues as a function of γ^2 for different values of μ.

trix with coefficients

$$A_{11}=0, \quad A_{12}=1\,,$$

$$A_{21}=\frac{\gamma^2}{(1-t)^4}+\frac{\lambda}{(1-t)^2[t^2+(1-t)^2]} - \frac{\mu^2}{[t^2+(1-t)^2]^2}\,,$$

$$A_{22}=2(1-2t)/[t^2+(1-t)^2]\,.$$

The above equation, subject to the even and odd boundary conditions, $y(0)=1, y'(0)=0$ and $y(0)=0, y'(0)=1$, respectively, together with $y(1)=0$, then constitutes a two-point boundary value problem and is solved by standard matrix inversion methods.

The results of such calculations for the lowest even and odd eigenvalues $\lambda_0^{(e)}$ and $\lambda_0^{(o)}$, are shown plotted in Fig. 3 as functions of γ^2 with μ^2 as a parameter. The detailed behavior of these eigenvalues for small γ^2 is presented in Fig. 4. As

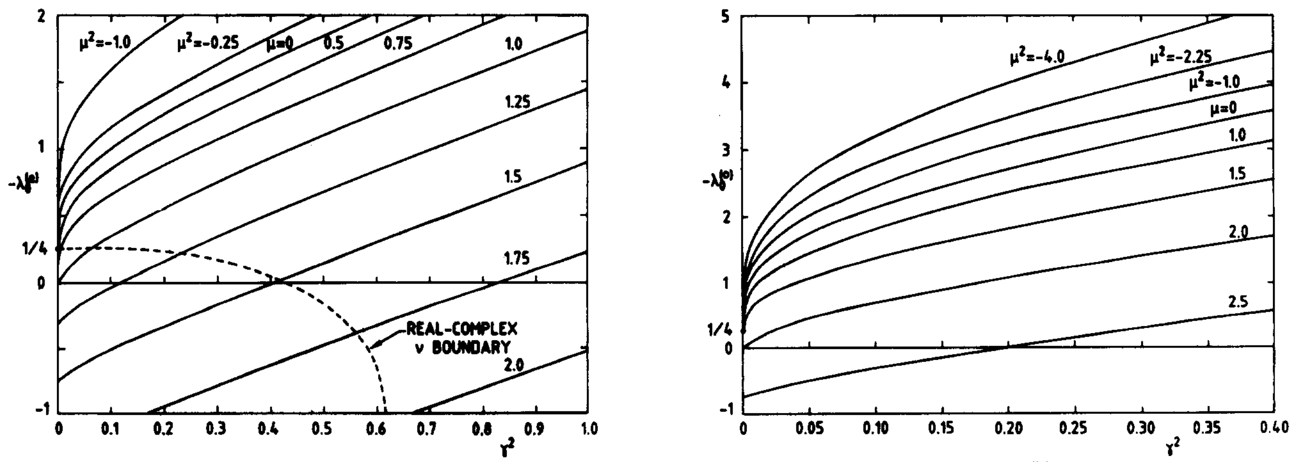


FIG. 4. The behavior of the lowest ($n=0$) (a) even and (b) odd eigenvalues for small values of γ^2 for different values of μ. All the curves with (a) $\mu<\frac{1}{2}$ and (b) $\mu<\frac{3}{2}$ pass through the accumulation point at $-\lambda=\frac{1}{4}, \gamma^2=0$. The dashed curve in (a) represents the boundary on which the characteristic exponent ν changes from real to complex values as described in Eq. (4.4).

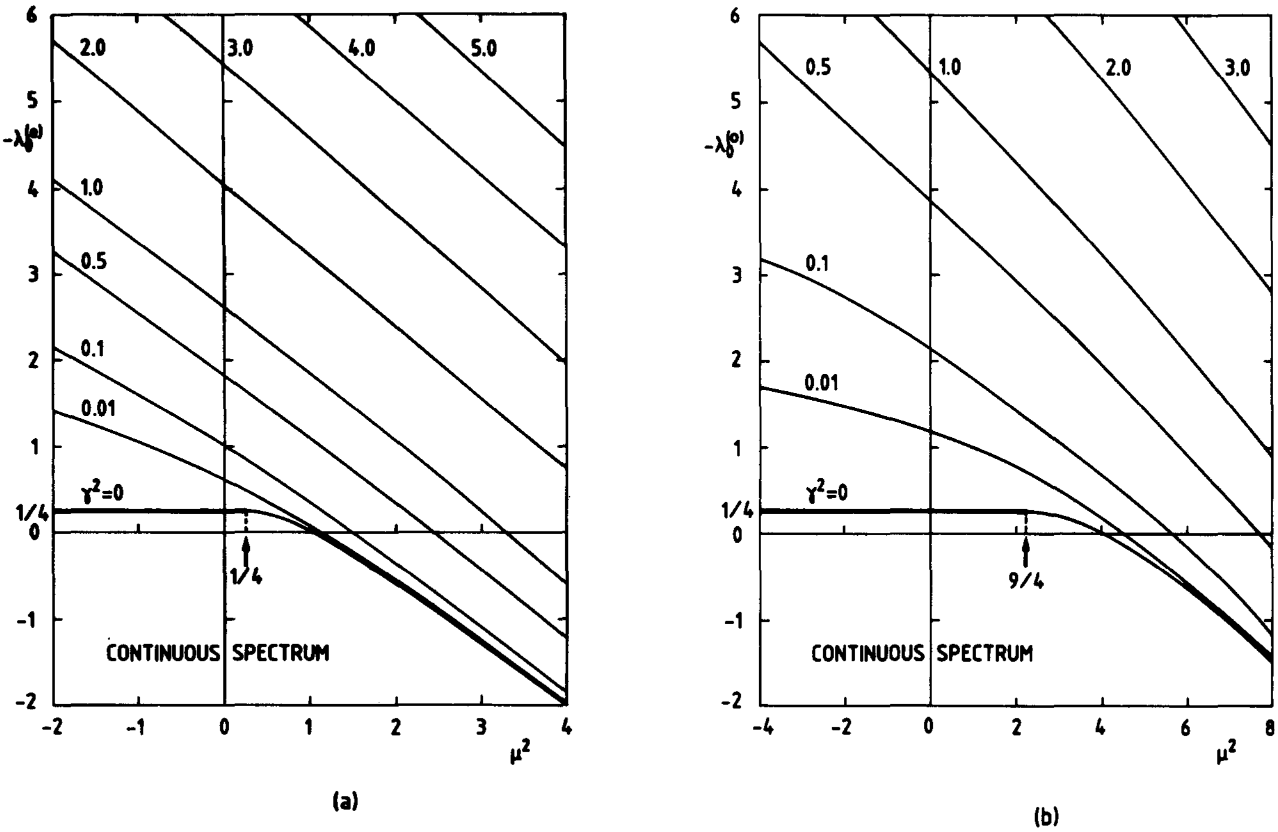


FIG. 5. The behavior of the lowest ($n = 0$) (a) even and (b) odd eigenvalues as a function of μ^2 for different values of γ^2. The solid line belonging to $\gamma^2 = 0$ denotes the continuous spectrum of non-$L^2(-\infty,\infty)$ eigenfunctions. Below the curve $\gamma^2 = 0$, the spectrum is continuous with $L^2(-\infty,\infty)$ eigenfunctions.

discussed in Sec. II, it is seen that, for $\mu^2<\frac{1}{4}$, all the $\lambda_0^{(e)}(\gamma^2)$ curves pass through the point $-\lambda=\frac{1}{4}$ as $\gamma^2\to 0$, whereas when $\mu>\frac{1}{2}$, $\lambda_0^{(e)}(\gamma^2)$ approaches the value $\mu(\mu-1)$ as $\gamma^2\to 0$ [cf. Eq. (2.10)]. A similar behavior is exhibited for the lowest odd eigenvalue curves $\lambda_0^{(o)}(\gamma^2)$: as $\gamma^2\to 0$ the curves all pass through the point $-\lambda=\frac{1}{4}$ when $\mu^2<\frac{9}{4}$, while $\lambda_0^{(o)}(\gamma^2)$ approaches the value $(\mu-1)(\mu-2)$ when $\mu>\frac{3}{2}$. In Fig. 5 we show the variation of $\lambda_0^{(e,o)}$ as functions of μ^2 with γ^2 as parameter.

The computed boundary upon which the characteristic exponent ν changes from the complex value in Eq. (4.4) to real values as γ^2 decreases along an eigenvalue curve is shown in the case of $\lambda_0^{(e)}$ by the dashed curve in Fig. 4(a). From Eq. (A5) in the limit $\nu\to -\frac{1}{2}$, this curve is approximately represented in the neighborhood of $\gamma^2=0$ by the envelope of the points of intersection of the family of curves

$$-\lambda=\tfrac{1}{4}-\tfrac{1}{2}\gamma^2(\mu^2-\tfrac{5}{4})-\tfrac{1}{16}\gamma^4(\mu^2-\tfrac{3}{4})+\cdots,$$

for different μ^2 and the corresponding eigenvalue curves. Above this dashed curve, the characteristic exponent is of the form in Eq. (4.4), while below it ν is real.

In Fig. 6 we show the variation of the higher eigenvalues $\lambda_n^{(e,o)}$, for $n=0,1,2$, and 3, as functions of γ^2 for the case $\mu=0$. All the eigenvalue curves for this value of μ accumulate at the point $-\lambda=\frac{1}{4}$ as $\gamma^2\to 0$ [cf. Eq. (6.8)]. This result confirms the conjecture made in Ref. 23 that Eq. (1.1) for positive γ^2 and $\mu=0$ admits solutions satisfying the boundary conditions only when $-\lambda>\frac{1}{4}$. The qualitative behavior of the lowest even eigenvalues near $\gamma^2=0$ is depicted in Fig. 7 for $\mu=1$ and $\mu=3$. The quantitative representation of these eigenvalue curves is difficult on account of their disparate functional dependence on γ^2 in the neighborhood of $\gamma^2=0$ (see Sec. VI). From Eq. (2.4), $N=0$ when $\mu=1$ and as $\gamma^2\to 0$ the eigenvalue curve $\lambda_0^{(e)}(\gamma^2)$ approaches the

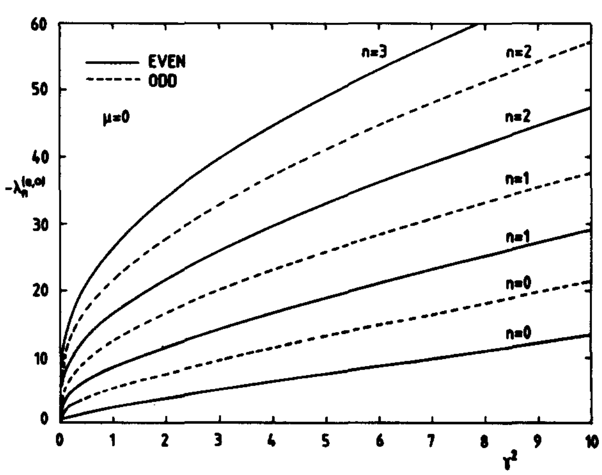


FIG. 6. The behavior of the higher even and odd eigenvalues $\lambda_n^{(e,o)}$ as a function of γ^2 when $\mu=0$.

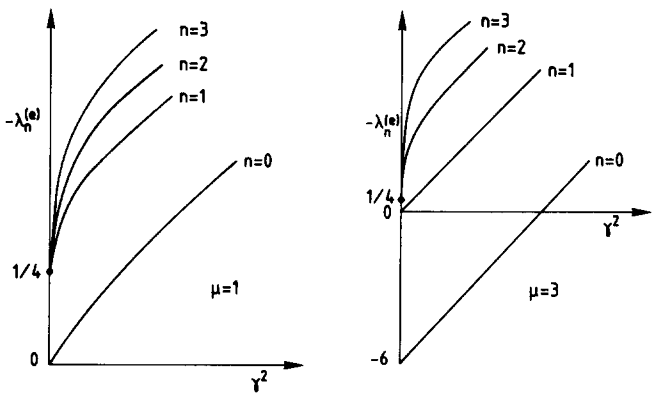


FIG. 7. The qualitative behavior of the first four even eigenvalues $\lambda_n^{(e)}$ for small γ^2 when $\mu = 1$ and $\mu = 3$.

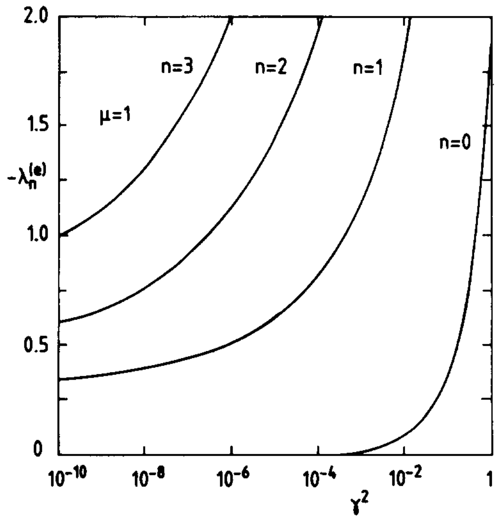


FIG. 8. The behavior of the first four even eigenvalues in the neighborhood of $\gamma^2 = 0$ in the case $\mu = 1$. The lowest eigenvalue $n = 0$ approaches zero algebraically while the higher eigenvalues with $n = 1,2,3$ approach the accumulation point $-\lambda = \frac{1}{4}$, $\gamma^2 = 0$ logarithmically.

value $\mu(\mu - 1)$ belonging to the finite spectrum when $\gamma^2 = 0$. The higher eigenvalue curves $\lambda_n^{(e)}(\gamma^2)$ $(n = 1,2,...)$ all pass through the accumulation point $-\lambda = \frac{1}{4}$, $\gamma^2 = 0$. When $\mu = 3$, $N = 2$, and the eigenvalue curves $\lambda_n^{(e)}(\gamma^2)$, $n = 0,1$, approach the values $\mu(\mu - 1)$ and $(\mu - 2)(\mu - 3)$ of the finite spectrum in Eq. (2.4) as $\gamma^2 \to 0$, with the higher eigenvalue curves for $n = 2,3,...$ all accumulating at $-\lambda = \frac{1}{4}$. A similar behavior is found for the odd eigenvalue curves when $\mu = 2$ and $\mu = 4$.

In general, when $\mu^2 \leqslant \frac{1}{4}$ the eigenvalue curves $\lambda_n^{(e,o)}(\gamma^2)$ $(n = 0,1,2,...)$ all pass through the accumulation point as $\gamma^2 \to 0$. When $\mu > \frac{1}{2}$, the even eigenvalue curves for

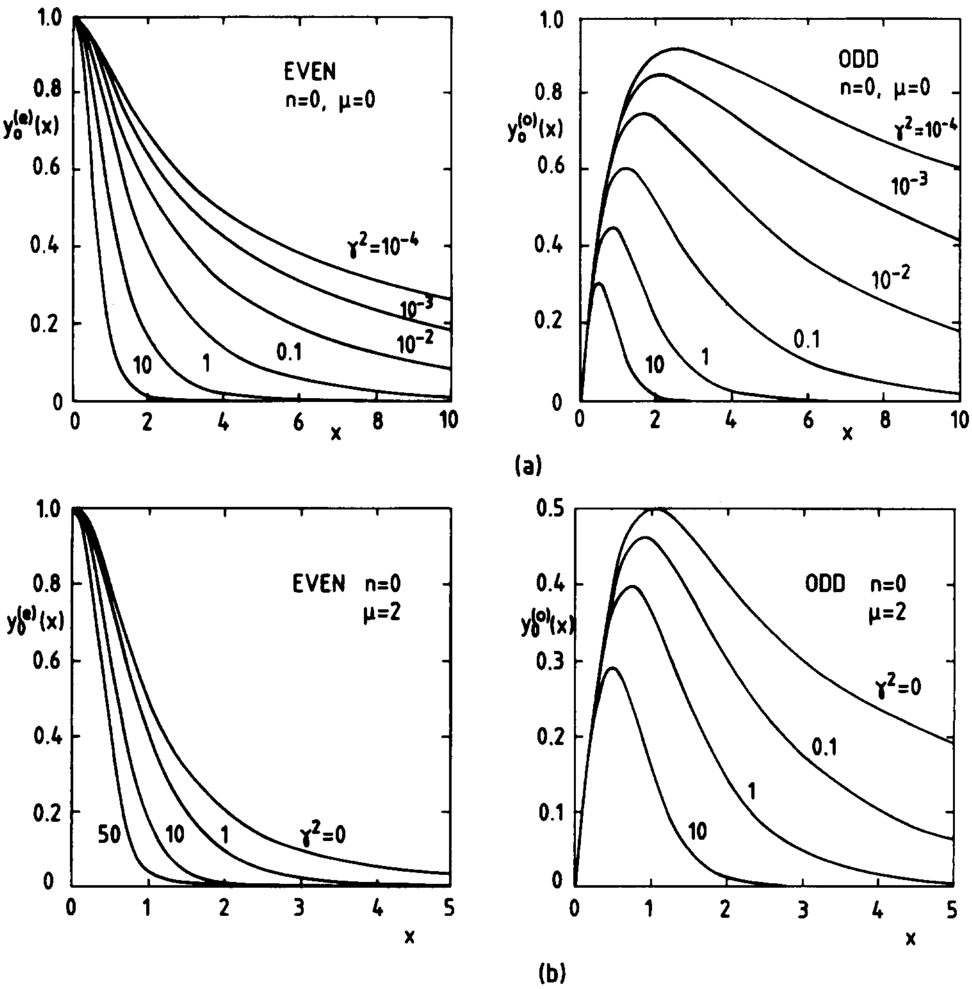


FIG. 9. Examples of the lowest even and odd eigenfunctions $y_0^{(e,o)}(x)$ for different values of γ^2 when (a) $\mu = 0$ and (b) $\mu = 2$. Note that when $\gamma^2 = 0$ there are no $L^2(-\infty,\infty)$ eigenfunctions for $\mu = 0$, while when $\mu = 2$ the eigenfunctions are given by Eq. (2.10).

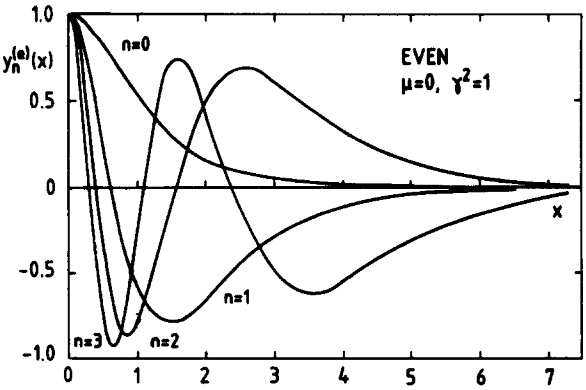


FIG. 10. The first four even eigenfunctions $y_n^{(e)}(x)$ $(n=0,1,2,3)$ for $\mu=0$ and $\gamma^2=1$. The nth even eigenfunction is associated with $2n$ zeros on the interval $(-\infty,\infty)$.

$n=0,1,\ldots,N^e$, where $N^{(e)}=[\frac12(\mu-\frac12)]$, approach their limiting values $(\mu-2n)(\mu-2n-1)$ in Eq. (2.8) as $\gamma^2\to 0$, while for the curves with $n>N^{(e)}$ the point $-\lambda=\frac14$, $\gamma^2=0$ is an accumulation point of the spectrum. Similarly, when $\mu>\frac32$ the odd eigenvalue curves for $n=0,1,\ldots,N^{(o)}$, where $N^{(o)}=[\frac12(\mu-\frac32)]$, approach the values $(\mu-2n-1)(\mu-2n-2)$ in Eq. (2.9) as $\gamma^2\to 0$, while the curves with $n>N^{(o)}$ all pass through the accumulation point in this limit. Figure 8 shows the detailed behavior when $\mu=1$ of the first four even eigenvalues as $\gamma^2\to 0$. The dependence of the eigenvalue curves for $n=1,2,3$ is seen to be logarithmic, in accordance with Eq. (6.8), while that of the eigenvalue curve for $n=0$ is algebraic [cf. Fig. 4(a)].

In Fig. 9 we present examples of the lowest even and odd eigenfunctions $y_0^{(e,o)}(x)$ [normalized so that $y_0^{(e)}(0)=1$ and $y_0^{(o)\prime}(0)=1$] for different values of γ^2 when $\mu=0$ and $\mu=2$. The eigenfunctions are seen to decay more rapidly as γ^2 increases [in accordance with Eq. (3.6)] at fixed μ, and also as μ^2 increases at fixed γ^2. In the case $\mu=0$, there is no discrete eigenvalue and the solutions spread about $x=0$ progressively as $\gamma^2\to 0$. When $\mu=2$, the even and odd eigenfunctions for $\gamma^2=0$ are, from Eq. (2.10), given by $y_0^{(e,o)}(x)=1+x^2$ and $x/(1+x^2)$, respectively; these curves represent upper limits on the amount of spread about $x=0$ of the solutions for $\gamma^2>0$.

Finally, in Fig. 10 we show the behavior of the higher normalized even eigenfunctions $y_n^{(e)}(x)$ for the particular case $\mu=0$ when $\gamma^2=1$. Application of the Sonine–Polyá theorem[24] to Eq. (1.1) shows that for a given eigenfunction (with eigenvalue λ_n) the moduli of the successive maxima and minima form a decreasing or increasing sequence according as $(-\lambda_n/2\gamma^2)\gtrless 1$, respectively. Reference to Fig. 6 shows that for the eigenfunctions depicted in Fig. 10 (as well as those with $n\geqslant 4$) $(-\lambda_n/2\gamma^2)$ is indeed greater than unity. In the case of large μ, where $(-\lambda_n/2\gamma^2)$ can be less than unity for certain values of n, the corresponding eigenfunctions are found to possess an increasing sequence of moduli of the successive maxima and minima [cf. also the eigenfunctions in Eqs. (2.8) and (2.9) when $\gamma^2=0$].

VI. APPROXIMATE EXPRESSIONS FOR THE EIGENVALUES WHEN $\gamma^2 \ll 1$

In the final two sections we derive approximate fomulas for the eigenvalues in the limits of small and large values of γ^2. Such expressions are of considerable interest not only in physical applications but also in explaining the main qualitative features of the eigenvalue curves given in the preceding section.

When $\gamma^2\ll 1$, we find from Eq. (A8) that

$$H_{\mu,\nu}(\gamma)=1+O(\gamma^2)\,. \tag{6.1}$$

The eigenvalue relation Eq. (4.3) in the small γ^2 limit then becomes

$$F^{(e)}(\tilde\nu)\simeq\left(\frac12\gamma\right)^{2\tilde\nu}\frac{\Gamma(\frac14+\frac12\mu+\frac12\tilde\nu)\Gamma(\frac14-\frac12\mu+\frac12\tilde\nu)}{\Gamma(\frac14+\frac12\mu-\frac12\tilde\nu)\Gamma(\frac14-\frac12\mu-\frac12\tilde\nu)}\times\left[\frac{\Gamma(1-\tilde\nu)}{\Gamma(1+\tilde\nu)}\right]^2-1=0\,. \tag{6.2}$$

Consider first real values of $\tilde\nu=\nu+\frac12$. We recall from Sec. IV that it suffices to consider $\tilde\nu>0$ only, since the zeros of $F^{(e)}(\tilde\nu)$ occur in symmetric pairs with respect to the point $\tilde\nu=0$. When $\mu>\frac12$, the eigenvalues correspond to those values of $\tilde\nu$ situated near the poles $\tilde\nu=\mu-\frac12-2n$, $n=0,1,\ldots,N^{(e)}$, where $N^{(e)}=[\frac12(\mu-\frac12)]$ [cf. Eq. (2.8)]. If we denote these zeros by $\tilde\nu=\mu-\frac12-2n-2\epsilon_n$, where we suppose $|\epsilon_n|\ll 1$ for small γ^2, we find from Eq. (6.2)

$$\epsilon_n\simeq-\left(\frac14\gamma^2\right)^{\mu-1/2-2n}\times\frac{\pi\Gamma(\mu-n)\Gamma(\mu+\frac12-n)\sec\pi\mu}{n!\,\Gamma(n+\frac12)\{\Gamma(\mu+\frac12-2n)\Gamma(\mu-\frac12-2n)\}^2},\quad n=0,1,\ldots,N^{(e)}\,. \tag{6.3}$$

It should be noticed that this approximation is not uniformly valid in μ in the neighborhood of $\mu=\frac12,\frac32,\ldots$; for these values of μ more delicate approximations of the gamma functions would be required.

Restricting our attention to the lowest eigenvalue with $n=0$, we have

$$\nu\simeq\mu-1-\frac{\pi^{1/2}\Gamma(\mu)\sec\pi\mu}{\Gamma^2(\mu-\frac12)\Gamma(\mu+\frac12)}\left(\frac14\gamma^2\right)^{\mu-1/2}.$$

From Eq. (A5) the eigenvalue $\lambda_0^{(e)}$ for $\gamma^2\ll 1$ is then given by

$$\lambda_0^{(e)}\simeq\mu(\mu-1)+4\pi^{1/2}\frac{(\mu-\frac12)\Gamma(\mu)\sec\pi\mu}{\Gamma^2(\mu-\frac12)\Gamma(\mu+\frac12)}\times(\tfrac14\gamma^2)^{\mu-1/2},\quad \tfrac12<\mu<\tfrac32, \tag{6.4}$$
$$\simeq\mu(\mu-1)-[1+1/(2\mu-3)]\gamma^2,\quad \mu>\tfrac32\,.$$

A similar procedure for the odd eigenvalue relation in Eq. (4.3) shows that when $\mu>\frac32$ the real zeros of $F^{(o)}(\tilde\nu)$ are situated at $\tilde\nu=\mu-\frac32-2n-2\epsilon_n$, $n=0,1,\ldots,N^{(o)}$, where $N^{(o)}=[\frac12(\mu-\frac32)]$. For small γ^2, ϵ_n is now given by

$$\epsilon_n\simeq\left(\frac14\gamma^2\right)^{\mu-3/2-2n}\times\frac{\pi\Gamma(\mu-n)\Gamma(\mu-\frac12-n)\sec\pi\mu}{n!\,\Gamma(n+\frac32)\{\Gamma(\mu-\frac12-2n)\Gamma(\mu-\frac32-2n)\}^2},\quad n=0,1,\ldots,N^{(o)}, \tag{6.5}$$

provided μ is not in the neighborhood of the values $\mu = \frac{3}{2}, \frac{5}{2},\ldots$. The lowest odd eigenvalue is then approximately

$$\lambda_0^{(o)} \simeq (\mu-1)(\mu-2) + 8\pi^{1/2}\frac{(\mu-\frac{3}{2})\Gamma(\mu)\sec\pi\mu}{\Gamma^2(\mu-\frac{3}{2})\Gamma(\mu-\frac{1}{2})} \times (\tfrac{1}{4}\gamma^2)^{\mu-3/2}, \quad \tfrac{3}{2}<\mu<\tfrac{5}{2}, \tag{6.6}$$
$$\simeq (\mu-1)(\mu-2) - [1+3/(2\mu-5)]\gamma^2, \quad \mu>\tfrac{5}{2}.$$

The expression in Eq. (6.4) at once reveals that for $\mu > \frac{3}{2}$ the eigenvalue curves $\lambda_0^{(e)}(\gamma^2)$ are linear as $\gamma^2 \to 0$ with slope

$$-\frac{\partial\lambda_0^{(e)}}{\partial\gamma^2}\bigg|_{\gamma^2=0} = 1 + \frac{1}{2\mu-3} \quad (\mu>\tfrac{3}{2}).$$

Thus, as μ steadily increases the slopes of the eigenvalue curves steadily approach unity near the ordinate axis. When $\frac{1}{2}<\mu<\frac{3}{2}$, however, the eigenvalue curves are no longer linear but are tangential to the ordinate axis as $\gamma^2 \to 0$ (cf. Figs. 3 and 4). The odd eigenvalue curves $\lambda_0^{(o)}(\gamma^2)$ are similarly seen to be linear as $\gamma^2 \to 0$ for $\mu > \frac{5}{2}$ and to approach the ordinate axis tangentially for $\frac{3}{2}<\mu<\frac{5}{2}$.

We now consider the case when the parameters are such that the characteristic exponent ν is complex as given in Eq. (4.4) [cf. Fig. 3(a)]. In this case it becomes more expedient to consider γ^2 as the eigenvalue, with λ and μ^2 as parameters. In this domian we obtain from Eq. (4.7) the approximate eigenvalue relations for $\gamma^2 \ll 1$,

$$\gamma_n^{(e,o)} \simeq 2\exp[-\{\Theta^{(e,o)} + (n+1)\pi\}/u], \quad n=0,1,2,\ldots,$$

where, from Eq. (A5), the value of λ is related to u by

$$-\lambda = \tfrac{1}{4} + u^2 + \tfrac{1}{2}\gamma^2[1-(\mu^2-\tfrac{1}{4})/(1+u^2)] + O(\gamma^4). \tag{6.7}$$

In the limit $u \to 0$, we find, from Eqs. (4.9) and (4.10) when μ is real,

$$\Theta^{(e,o)} \simeq u[\psi(\mu+\tfrac{1}{2}) - 2\psi(1) - \ln 2] - \theta^{(e,o)},$$

where ψ denotes the logarithmic derivative of the gamma function and

$$\theta^{(e,o)} \simeq \tfrac{1}{2}\pi u \tan\tfrac{1}{2}\pi(\mu \pm \tfrac{1}{2}) + m(\mu)\pi(\mu \pm \tfrac{1}{2} \neq 1,3,5,\ldots),$$
$$= \tfrac{1}{2}\pi(\mu \pm \tfrac{1}{2}) \quad (\mu \pm \tfrac{1}{2} = 1,3,5,\ldots),$$

with $m(\mu)$ defined in Eq. (4.11). We then have the approximations

$$\gamma_n^{(e,o)} \simeq 4\exp[\tfrac{1}{2}\pi\tan\tfrac{1}{2}\pi(\mu\pm\tfrac{1}{2}) - 2\bar{\gamma} - \psi(\mu+\tfrac{1}{2})] \times e^{(m(\mu)-n-1)\pi/u} \quad (\mu \pm \tfrac{1}{2} \neq 1,3,5,\ldots), \tag{6.8}$$
$$\simeq 4\exp[-2\bar{\gamma} - \psi(\mu+\tfrac{1}{2})]e^{((1/2)(\mu\pm\frac{1}{2})-n-1)\pi/u} \quad (\mu \pm \tfrac{1}{2} = 1,3,5,\ldots) \quad n=0,1,2,\ldots,$$

respectively, where $\bar{\gamma} = 0.5772\ldots$ is Euler's constant.

The first approximation in Eq. (6.8) is not uniformly valid in μ in the neighborhood of the values $\mu \pm \frac{1}{2} = 2k+1$, $k = 0,1,2,\ldots$, since the corresponding approximation for $\theta^{(e,o)}$ breaks down. As $\mu \pm \frac{1}{2}$ passes through these values, the phase angles $\theta^{(e,o)}$, respectively, can be seen to vary extremely rapidly when $u \ll 1$. The description of $\gamma_n^{(e,o)}$ in the neighborhood of $\mu \pm \frac{1}{2} = 2k+1$ would require more refined approximations to $\theta^{(e,o)}$ depending on the relative magnitude of u and $\mu \pm \frac{1}{2} - 2k - 1$.

The formulas in Eq. (6.8) are valid as $u \to 0$ only for those values of μ^2 and n for which $\gamma^2 \ll 1$. The larger the value of n, the greater is the range of values of μ^2 that is covered. Inspection of Fig. 3 reveals that the range of μ^2 for which the lowest eigenvalues $\gamma_0^{(e,o)}$ are adequately described by Eq. (6.8) is limited above by roughly $\mu^2 = 2$ and $\mu^2 = 4$, respectively. For fixed u and μ^2, the eigenvalues $\gamma_n^{(e,o)}$ depend exponentially on u and form a decreasing sequence of values with zero as limit point. It is clear from Eqs. (6.7) and (6.8) that there is an infinite number of eigenvalues for any given μ^2 that accumulate at the point $-\lambda = \frac{1}{4}, \gamma^2 = 0$.

For a given value of $u \ll 1$, the ratio of the even and odd eigenvalues in the small γ^2 limit is given by $\gamma_n^{(e)}/\gamma_n^{(o)} \simeq \exp[(\theta^{(e)} - \theta^{(o)})/u]$. This ratio is always greater than unity and is periodic in μ with period 2. In the particular case $\mu = 0$, we have from Eq. (6.7), as $u \to 0$,

$$\gamma_n^{(e,o)} \simeq 16e^{-\bar{\gamma}\pm(1/2)\pi}\exp[-(n+1)\pi/u], \quad n=0,1,2,\ldots \quad (\mu=0), \tag{6.9}$$

so that in the neighborhood of the accumulation point $\gamma_n^{(e)}/\gamma_n^{(o)} \approx e^{\pi}$.

When the parameters are such that u is somewhat larger than unity (but with γ^2 still small), it can be shown from Eqs. (4.9) or (4.8) (according as μ is real or imaginary) by Stirling's formula that

$$-\Theta^{(e,o)} - \tfrac{1}{2}\pi \simeq u\ln 2u - u \pm \tfrac{1}{4}\pi + (1/2u)(\mu^2 - \tfrac{5}{12}),$$

when μ^2 is not too large. In this limit the eigenvalues are approximated by

$$\gamma_n^{(e,o)} \simeq \frac{4u}{e}\exp\left[\pm\frac{\pi}{4u} + \frac{1}{2u^2}\left(\mu^2 - \frac{5}{12}\right)\right]e^{-(n+1/2)\pi/u}, \quad n=0,1,2,\ldots, \tag{6.10}$$

whence $\gamma_n^{(e)}/\gamma_n^{(o)} \simeq e^{\pi/2u}$ independent of μ^2.

We emphasize that the approximations in Eqs. (6.4) and (6.6)–(6.10) can only be applied for those values of u, μ^2, and n that correspond to $\gamma^2 \ll 1$. It is found that there is satisfactory agreement with the numerically computed values shown in Figs. 3 and 4 in their respective domains of validity.

The even eigenvalue relation Eq. (6.7) for $\gamma^2 \ll 1$ has been previously obtained in the case $\mu = 0$ by Kulsrud[22] and Stringer.[25] By approximate solution of Eq. (1.1) in terms of Bessel and Legendre functions in two overlapping regions, $x \gg 1$ and $\gamma x \ll |\lambda|^{1/2}$, these authors employed the method of matched asymptotic expansions to obtain Eq. (6.7). A similar procedure has been employed by Antonsen *et al.*[7] and Paris[26] for the general case with μ arbitrary. The matched asymptotic expansion approach is equivalent to the small γ^2 limit employed in Eq. (6.1), since this consists of taking only the first term in the expansion of $S_\nu^{\mu(3)}(ix;\gamma)$ in Eq. (3.5) and its continuation involving Legendre functions in Eq. (3.7).

VII. APPROXIMATE EXPRESSIONS FOR THE EIGENVALUES WHEN $\gamma^2 \to +\infty$

The determination of the eigenvalues from Eq. (3.8) and the functional equation (A2) in terms of the characteristic exponent ν is not suitable as $\gamma^2 \to +\infty$ on account of the increasing number of terms that have to be retained in the determinant. To determine the dependence of λ on γ^2 and μ^2 in this limit, we adopt a different procedure that does not require knowledge of the characteristic exponent. It is well known (e.g., Ref. 27, p. 521) that the density of the eigenvalues of Eq. (2.2) for large γ^2 is described by the Weyl formula

$$\int_{-A}^{A} [\Lambda - q(\zeta)]^{1/2}\, d\zeta = \left(n + \frac{1}{2}\right)\pi + O\left(\frac{1}{\gamma}\right), \quad n = 0,1,2,\dots,$$

where Λ and $q(\zeta)$ are defined in Eq. (2.3) and $\pm$ A denote the roots of $\Lambda = q(\zeta)$ given by

$$A = \operatorname{arccosh}[\{\Lambda + [\Lambda^2 + 4\gamma^2(\mu^2 - \tfrac{1}{4})]^{1/2}\}/2\gamma^2]^{1/2}.$$

This leads to the estimate

$$-\lambda_n = \gamma^2 + (2n+1)\gamma + O(1), \quad \gamma^2 \to +\infty, \quad n = 0,1,2,\dots,$$

where even and odd values of n correspond to the even and odd eigenvalues, respectively.

To obtain the constant implied in the $O(1)$ term and higher-order correction terms, we put $y(x) = (1 + x^2)^{\mu/2} v(\xi)$, $\xi = (2\gamma)^{1/2}x$ $(\gamma > 0)$ in Eq. (1.1) to obtain

$$\left(1 + \frac{\xi^2}{2\gamma}\right)\frac{d^2v}{d\xi^2} + \frac{(1+\mu)}{\gamma}\xi\frac{dv}{d\xi} + \left(\widetilde{\Lambda} - \frac{1}{4}\xi^2\right)v = 0, \tag{7.1}$$

$$\widetilde{\Lambda} = [-\lambda - \gamma^2 + \mu(1+\mu)]/2\gamma,$$

subject to the conditions that $v(\xi) \to 0$ as $\xi \to \pm\infty$. Since the limiting form of Eq. (7.1) as $\gamma^2 \to +\infty$ is Weber's equation, we expand $v(\xi)$ in terms of normalized Hermite functions

$$v(\xi) = \sum_{n=0}^{\infty} \alpha_n u_n(\xi). \tag{7.2}$$

The coefficients α_n are independent of ξ and $u_n(\xi)$ is defined in terms of the parabolic cylinder function $D_n(\xi)$ by

$$u_n(\xi) = [(2\pi)^{-1/4}/(n!)^{1/2}]D_n(\xi) = \frac{(-)^n}{(2\pi)^{1/4}(n!)^{1/2}} e^{(1/4)\xi^2}\frac{d^n}{d\xi^n}e^{-(1/2)\xi^2} \quad (n = 0,1,2,\dots)$$

and satisfy the differential equation

$$u_n''(\xi) + (n + \tfrac{1}{2} - \tfrac{1}{4}\xi^2)u_n(\xi) = 0.$$

Formal substitution of Eq. (7.2) into Eq. (7.1) and application of the orthogonality property of the Hermite functions yields

$$\left(\widetilde{\Lambda} - n - \frac{1}{2}\right)\alpha_n - \frac{1}{2\gamma}\sum_{m=0}^{\infty}\alpha_m I_{nm} = 0, \tag{7.3}$$

where

$$I_{nm} = \int_{-\infty}^{\infty}\left[\xi^2\left(n + \frac{1}{2} - \frac{1}{4}\xi^2\right)u_m(\xi) - 2(1+\mu)\xi\, u_m'(\xi)\right]u_n(\xi)\,d\xi. \tag{7.4}$$

The integrals I_{nm} involve moments of products of Hermite functions and have been evaluated in Ref. 10. It is found that I_{nm} possesses a banded structure in n and m, taking nonzero values only when $m = n$, $n \pm 1$, and $n \pm 2$. It is this banded structure that makes the sum in Eq. (7.3) a sum over a finite number of terms, thus simplifying considerably the asymptotic evaluation of λ.

The infinite secular determinant resulting from Eq. (7.3) is

$$\|(\widetilde{\Lambda} - n - \tfrac{1}{2})\delta_{nm} - (1/2\gamma)I_{nm}\| = 0, \tag{7.5}$$

where δ denotes the Kronecker delta. Since I_{nm} vanishes for n and m of different parity, we can uncouple the even and odd solutions. A first approximation to the even and odd eigenvalues may be obtained by equating the diagonal elements to zero, to find

$$\widetilde{\Lambda}_n^{(e)} - 2n - \tfrac{1}{2} - (1/2\gamma)I_{2n,2n} = 0,$$
$$\widetilde{\Lambda}_n^{(o)} - 2n - \tfrac{3}{2} - (1/2\gamma)I_{2n+1,2n+1} = 0,$$
$$n = 0,1,2,\dots.$$

From Ref. 10, we have the integrals

$$I_{2n,2n} = \mu + \tfrac{3}{4} + n(2n+1),$$
$$I_{2n+1,2n+1} = \mu + \tfrac{7}{4} + n(2n+3),$$

so that for $\gamma^2 \to +\infty$ the even and odd eigenvalues are given by

$$\begin{aligned} -\lambda_n^{(e)} &= \gamma^2 + (4n+1)\gamma - \mu^2 + \tfrac{3}{4} + n(2n+1) + O(\gamma^{-1}), \\ -\lambda_n^{(o)} &= \gamma^2 + (4n+3)\gamma - \mu^2 + \tfrac{7}{4} + n(2n+3) + O(\gamma^{-1}), \\ n &= 0,1,2,\dots. \end{aligned} \tag{7.6}$$

The higher-order terms in the expansion of $\lambda_n^{(e,o)}$ can be obtained by requiring that the determinant in Eq. (7.5) vanish order by order in inverse powers of γ. For the lowest eigenvalues with $n = 0$ it is found[10] that, for $\gamma^2 \to +\infty$,

$$\begin{aligned} -\lambda_0^{(e)} &= \gamma^2 + \gamma - \mu^2 + \frac{3}{4} + \frac{1}{2\gamma}\left(\mu^2 - \frac{3}{8}\right) - \frac{3}{4\gamma^2}\left(\mu^2 - \frac{5}{16}\right) + O(\gamma^{-3}), \\ -\lambda_0^{(o)} &= \gamma^2 + 3\gamma - \mu^2 + \frac{7}{4} + \frac{3}{2\gamma}\left(\mu^2 - \frac{5}{8}\right) - \frac{15}{4\gamma^2}\left(\mu^2 - \frac{7}{16}\right) + O(\gamma^{-3}). \end{aligned} \tag{7.7}$$

These asymptotic approximations are found to agree well with the numerically computed values for large γ^2 shown in Figs. 4 and 6 as illustrated in Table I. It is clear from Eq. (7.6) that the eigenvalue curves $\lambda_n^{(e,o)}(\gamma^2)$ are linear in γ^2 to leading order and that for $n = 1,2,\dots$ this approximate linear

TABLE I. Comparison of the lowest even and odd eigenvalues with the asymptotic values in Eq. (7.7) for $\mu=0$ and $\mu=1$.

$\mu=0$ γ^2	$-\lambda_0^{(e)}$ Exact	Asymptotic	$-\lambda_0^{(o)}$ Exact	Asymptotic
5	7.929 12	7.9491	13.203 95	13.3671
10	13.868 23	13.8764	21.035 99	21.1044
25	30.719 45	30.7219	41.607 28	41.6281
50	57.798 30	57.7992	72.855 20	72.8634

$\mu=1$ γ^2	$-\lambda_0^{(e)}$ Exact	Asymptotic	$-\lambda_0^{(o)}$ Exact	Asymptotic
5	7.069 35	7.0227	12.541 06	12.7098
10	12.978 71	12.9595	20.312 36	20.4147
25	29.797 56	29.7919	40.812 04	40.7781
50	56.857 15	56.8549	72.014 17	72.0006

dependence is obtained for progressively larger values of γ^2 (cf. Fig. 6).

The asymptotic expansion of λ as $\gamma^2\to+\infty$ for the spheroidal differential equation (3.1), when μ is an integer, has been given in Refs. 11, 18 (p. 243), and 28 as far as the term in γ^{-5}. With $y(z)=(1-z^2)^{(1/2)\mu}v(\xi)$, $\xi=(2\gamma)^{1/2}z$ the corresponding transformed equation has the same form as Eq. (7.1) when γ is replaced by $-\gamma$. The requirement that $y(z)$ be finite at $z=\pm 1$ is then equivalent, in the limit $\gamma^2\to+\infty$, to the requirement $v(\xi)\to 0$ as $\xi\to\pm\infty$ in Eq. (7.1). Consequently, the expansion of the eigenvalues for Eq. (1.1) as $\gamma^2\to+\infty$ for arbitrary real μ^2 may be obtained from that given by Meixner and Schäfke[18] and Meixner[28] for Eq. (3.1) in the same limit, provided that in their expansion we replace γ by $-\gamma$ and their definition of q by $q=4n+1$ and $q=4n+3$ ($n=0,1,2,\ldots$) for the even and odd solutions, respectively.

The corresponding eigenfunctions $y_n^{(e,o)}(x)$ in the large γ^2 limit may be approximated by solving for the coefficients α_n in Eq. (7.3) in descending powers of γ. The lowest even and odd eigenfunctions corresponding to Eq. (7.7) are found (to within arbitrary constant multiples) to be given by[10]

$$y_0^{(e)}(x)=(1+x^2)^{(1/2)\mu}\sum_{n=0}^{\infty}\frac{(-)^n[(2n)!]^{1/2}}{2^{2n}}\times C_{2n}u_{2n}[(2\gamma)^{1/2}x], \tag{7.8}$$

$$y_0^{(o)}(x)=(1+x^2)^{(1/2)\mu}\sum_{n=0}^{\infty}\frac{(-)^n[(2n+1)!]^{1/2}}{2^{2n}}\times C_{2n+1}u_{2n+1}[(2\gamma)^{1/2}x],$$

where the coefficients C_n up to $O(\gamma^{-3})$ are given by

$$C_0=C_1=1,$$

$$C_2=\frac{\mu}{\gamma}\left[1-\frac{9}{8\gamma}+\frac{1}{32\gamma^2}(63-8\mu^2)+\cdots\right],$$

$$C_3=\frac{\mu}{\gamma}\left[1-\frac{15}{8\gamma}+\frac{1}{32\gamma^2}(155-8\mu^2)+\cdots\right],$$

$$C_4=\frac{1}{2\gamma}\left[1+\frac{1}{2\gamma}(4\mu^2-15)+\frac{1}{16\gamma^2}\left(\frac{151}{4}-52\mu^2\right)+\cdots\right],$$

$$C_5=\frac{1}{2\gamma}\left[1+\frac{1}{4\gamma}(4\mu^2-7)+\frac{1}{16\gamma^2}\left(\frac{283}{4}-76\mu^2\right)+\cdots\right],$$

$$C_6=\frac{\mu}{2\gamma^2}\left[1+\frac{1}{3\gamma}\left(\mu^2-\frac{77}{8}\right)+\cdots\right],$$

$$C_7=\frac{\mu}{2\gamma^2}\left[1+\frac{1}{3\gamma}\left(\mu^2-\frac{107}{8}\right)+\cdots\right],$$

$$C_8=\frac{1}{8\gamma^2}\left[1+\frac{1}{\gamma}\left(2\mu^2-\frac{7}{2}\right)+\cdots\right],$$

$$C_9=\frac{1}{8\gamma^2}\left[1+\frac{1}{\gamma}\left(2\mu^2-\frac{9}{2}\right)+\cdots\right],$$

$$C_{10}=\mu/8\gamma^3+\cdots,\quad C_{11}=\mu/8\gamma^3+\cdots,$$

$$C_{12}=1/48\gamma^3+\cdots,$$

$$C_{13}=1/48\gamma^3+\cdots.$$

When due account is taken of the normalization employed in Eq. (7.2), the above expansions for $\gamma^2\to+\infty$ of the coefficients for arbitrary real μ^2 may be seen to be the same as those given in Ref. 28 for the eigenfunctions of Eq. (3.1) for integer μ on the interval $[-1,+1]$, provided γ is again replaced by $-\gamma$ with $q=1,3,\ldots$.

ACKNOWLEDGMENT

One of the authors (R.B.P.) is grateful to Professor A. D. Wood for helpful discussions on the nature of the spectrum discussed in Sec. II.

APPENDIX: PRINCIPAL PROPERTIES OF THE SPHEROIDAL FUNCTIONS

The oblate spheroidal differential equation (3.1) has two limiting cases for which simple solutions may be given. These are $\gamma^2=0$, when the solution may be expressed in terms of the Legendre functions $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ and $|z^2|\gg 1$, when Eq. (3.1) reduces to Bessel's equation with solution $(\gamma z)^{-1/2}\mathscr{C}_{\nu+1/2}(\gamma z)$, where $\mathscr{C}$ denotes any Bessel function and ν (in these cases) is defined in Eq. (2.7). Based on these limiting forms, solutions of Eq. (3.1) can be found in series of Bessel and Legendre functions. The spheroidal functions of the first and second groups are defined in Eqs. (3.2) and (3.3). In addition to these solutions, there are also the solutions associated with the parameter $-\mu$, namely $Ps_\nu^{-\mu}(z;\gamma)$, $Qs_\nu^{-\mu}(z;\gamma)$, and $S_\nu^{-\mu(j)}(z;\gamma)$, together with the solutions $Ps_{-\nu-1}^{\pm\mu}(z;\gamma)$, $Qs_{-\nu-1}^{\pm\mu}(z;\gamma)$, and $S_{-\nu-1}^{\pm\mu(j)}(z;\gamma)$ ($j=1,2,3,4$).

The parameter ν is known as the characteristic exponent of Eq. (3.1) and, for general values of γ^2, is determined by the condition of a nontrivial solution for the coefficients $a_{\nu,2r}^\mu(\gamma^2)$ appearing in Eqs. (3.2) and (3.3). Use of the recurrence relations for either the Legendre or Bessel functions shows that the coefficients $a_{2r}\equiv a_{\nu,2r}^\mu(\gamma^2)$ satisfy the three-term recurrence relation

$$a_{2r}+\theta\left[\frac{A_r(\mu,\nu)}{\lambda-C_r(\mu,\nu)}a_{2r-2}+\frac{B_r(\mu,\nu)}{\lambda-C_r(\mu,\nu)}a_{2r+2}\right]=0,$$
$$\theta=\tfrac{1}{4}\gamma^2,\quad r=0,\pm 1,\pm 2,\ldots$$
$$(\nu\pm\tfrac{1}{2}\neq 0,\pm 1,\pm 2,\ldots),\tag{A1}$$

where

$$A_r(\mu,\nu)=\frac{(\nu+2r-\mu)(\nu+2r-\mu-1)}{(\nu+2r-\frac{1}{2})(\nu+2r-\frac{3}{2})},$$

$$B_r(\mu,\nu)=\frac{(\nu+2r+\mu+1)(\nu+2r+\mu+2)}{(\nu+2r+\frac{3}{2})(\nu+2r+\frac{5}{2})},$$

$$C_r(\mu,\nu)=(\nu+2r)(\nu+2r+1)-2\theta\times\frac{[(\nu+2r)(\nu+2r+1)+\mu^2-1]}{(\nu+2r-\frac{1}{2})(\nu+2r+\frac{3}{2})}.$$

Equation (A1) then possesses a nontrivial solution provided

$$\Delta_{\mu,\nu}(\lambda,\gamma^2)=0,\tag{A2}$$

where $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ is the absolutely convergent infinite determinant (with diagonal elements all equal to unity) defined by

$$\Delta_{\mu,\nu}(\lambda,\gamma^2)=\begin{vmatrix} & & & \ddots & & & & & \\ \cdots & 0 & \theta A_{-1}/(\lambda-C_{-1}) & 1 & \theta B_{-1}/(\lambda-C_{-1}) & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & \theta A_0/(\lambda-C_0) & 1 & \theta B_0/(\lambda-C_0) & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \theta A_1/(\lambda-C_1) & 1 & \theta B_1/(\lambda-C_1) & 0 & \cdots \\ & & & & & & \ddots & & \end{vmatrix}.\tag{A3}$$

From the properties of the coefficients $A_r(\mu,\nu)$, $B_r(\mu,\nu)$, and $C_r(\mu,\nu)$, it is readily demonstrated that $\Delta_{\mu,\nu}(\lambda,\gamma^2)$ is an even function of μ and is invariant under the transformation $\nu\to-\nu-1$; that is,

$$\Delta_{-\mu,\nu}(\lambda,\gamma^2)=\Delta_{\mu,\nu}(\lambda,\gamma^2),\quad \Delta_{\mu,\nu}(\lambda,\gamma^2)=\Delta_{\mu,-\nu-1}(\lambda,\gamma^2).\tag{A4}$$

The determinantal equation (A2) describes the *functional* relationship between λ,γ^2,μ, and the characteristic exponent ν. The functional dependence of λ on γ^2, μ, and ν is represented by the notation $\lambda\equiv\lambda_\nu^\mu(\gamma^2)$. For sufficiently small γ^2, the expansion of λ in ascending powers of γ^2 can be obtained by setting the central 3×3 determinant in Eq. (A.3) to zero to find

$$\lambda\equiv\lambda_\nu^\mu(\gamma^2)=\nu(\nu+1)-\frac{1}{2}\left[1+\frac{4\mu^2-1}{(2\nu-1)(2\nu+3)}\right]\gamma^2+\frac{1}{2}\left[\frac{(\nu^2-\mu^2)\{(\nu-1)^2-\mu^2\}}{(2\nu-3)(2\nu-1)^3(2\nu+1)}-\frac{\{(\nu+1)^2-\mu^2\}\{(\nu+2)^2-\mu^2\}}{(2\nu+1)(2\nu+3)^3(2\nu+5)}\right]\gamma^4+O(\gamma^6).\tag{A5}$$

By retaining more central rows and columns, the functional relation for $\lambda_\nu^\mu(\gamma^2)$ has been obtained as far as the term in γ^8 in Ref. 18 (p. 269). The value of $\lambda_\nu^\mu(\gamma^2)$ is invariant under the transformation $\mu\to-\mu$ and $\nu\to-\nu-1$, so that

$$\lambda_\nu^\mu(\gamma^2)=\lambda_\nu^{-\mu}(\gamma^2)=\lambda_{-\nu-1}^{\pm\mu}(\gamma^2)\tag{A6}$$

and [cf. Eq. (2.7)]

$$\lambda_\nu^\mu(0)=\lambda_{-\nu-1}^\mu(0)=\nu(\nu+1).\tag{A7}$$

The coefficients $a_{\nu,2r}^{\pm\mu}(\gamma^2)$ are normalized such that $a_{\nu,0}^{\pm\mu}(0)=1$, and as $\gamma^2\to 0$ satisfy

$$a_{\nu,2r}^{\pm\mu}(\gamma^2)=O(\gamma^{|2r|})\quad(r\neq 0).\tag{A8}$$

From the recurrence relation [Eq. (A1)], it is readily seen that

$$a_{\nu,2r}^{\pm\mu}(\gamma^2)=a_{-\nu-1,-2r}^{\pm\mu}(\gamma^2),\tag{A9}$$

$$a_{\nu,2r}^{-\mu}(\gamma^2)=\frac{\Gamma(\nu-\mu+1)\Gamma(\nu+\mu+2r+1)}{\Gamma(\nu+\mu+1)\Gamma(\nu-\mu+2r+1)}a_{\nu,2r}^\mu(\gamma^2).$$

The quantities $A_\nu^{\pm\mu}(\gamma^2)$ are normalizing factors chosen to be

$$A_\nu^{\pm\mu}(\gamma^2)=\sum_{r=-\infty}^{\infty}(-)^r a_{\nu,2r}^{\pm\mu}(\gamma^2)=A_{-\nu-1}^{\pm\mu}(\gamma^2),\tag{A10}$$

so that as $|z|\to\infty$ in $|\arg(\gamma z)|<\pi$ [cf. Eq. (3.3)],

$$S_\nu^{\pm\mu(j)}(z;\gamma)\sim(\pi/2\gamma z)^{1/2}\,\mathscr{C}_{\nu+1/2}^{(j)}(\gamma z)\quad(j=1,2,3,4).\tag{A11}$$

The different spheroidal functions must be related since there can only be two linearly independent solutions of Eq. (3.1). The pair of solutions $S_\nu^{\mu(1)}(z;\gamma)$ and $Qs_{-\nu-1}^\mu(z;\gamma)$ are found to be multiples of each other, expressed by the relation[10,17,18]

$$S_\nu^{\mu(1)}(z;\gamma)=(1/\pi)\sin\pi(\nu-\mu)e^{-\pi i(\nu+\mu+1)}\times K_\nu^\mu(\gamma)Qs_{-\nu-1}^\mu(z;\gamma).\tag{A12}$$

The right-hand side of Eq. (A12) represents the analytic continuation of the solution $S_\nu^{\mu(1)}(z;\gamma)$ into the unit circle $|z|<1$. The quantity $K_\nu^\mu(\gamma)$ is a constant (dependent on ν, μ, and γ) known as the joining factor that links the $S_\nu^{\mu(1)}(z;\gamma)$ and $Qs_{-\nu-1}^\mu(z;\gamma)$ functions in their common domain of convergence. The representation of $K_\nu^\mu(\gamma)$ may be determined by expanding both sides of Eq. (A12) in a Laurent double series in powers of z^2 and equating like powers of

z^2 to find[10,18]

$$K_\nu^\mu(\gamma) = \frac{1}{2}\left(\frac{\gamma}{4}\right)^\nu \frac{\Gamma(1+\nu-\mu)e^{\pi\nu i}}{A_\nu^{-\mu}(\gamma^2)} \times \frac{\Sigma_{r=0}^\infty \left[(-)^r a_{\nu,-2r}^{-\mu}(\gamma^2)/r!\Gamma(\nu-r+\frac{3}{2})\right]}{\Sigma_{r=0}^\infty \left[(-)^r a_{\nu,2r}^{\mu}(\gamma^2)/r!\Gamma(\frac{1}{2}-\nu-r)\right]} . \tag{A13}$$

The joining factor $K_{-\nu-1}^\mu(\gamma)$, which links the solutions $S_{-\nu-1}^{\mu(1)}(z;\gamma)$ and $Qs_\nu^\mu(z;\gamma)$, is obtained from Eq. (A13) by replacing ν by $-\nu-1$. Use of the properties of the coefficients in Eq. (A9) then leads to the relation between the joining factors associated with the exponents ν and $-\nu-1$ and the normalizing constants associated with $\pm\mu$

$$K_\nu^\mu(\gamma)K_{-\nu-1}^\mu(\gamma)A_\nu^\mu(\gamma^2)A_\nu^{-\mu}(\gamma^2) = -(1/\gamma)\Gamma(1+\nu-\mu)\Gamma(-\nu-\mu) . \tag{A14}$$

Similar continuations exist for the other Bessel function expansion solutions. In Sec. III we require the continuation formula corresponding to Eq. (A12) for the Hankel function expansion solution $S_\nu^{\mu(3)}(z;\gamma)$. To obtain this, we use[10,17,18]

$$\cos\pi\nu\, S_\nu^{\mu(3)}(z;\gamma) = e^{-\pi\nu i}S_\nu^{\mu(1)}(z;\gamma) - i\,S_{-\nu-1}^{\mu(1)}(z;\gamma) ,$$

$$\sin\pi(\nu-\mu)Qs_{-\nu-1}^\mu(z;\gamma) = \sin\pi(\nu+\mu)Qs_\nu^\mu(z;\gamma) - \pi e^{\pi i\mu}\cos\pi\nu\, Ps_\nu^\mu(z;\gamma) ,$$

which follow from the standard properties of the Bessel and Legendre functions. We then deduce from these relations, together with Eqs. (A12) and (3.2), that

$$S_\nu^{\mu(3)}(z;\gamma) = APs_\nu^\mu(z;\gamma) + BQs_\nu^\mu(z;\gamma) \tag{A15}$$

$$= \sum_{r=-\infty}^{\infty} (-)^r a_{\nu,2r}^\mu(\gamma^2) \times \left[AP_{\nu+2r}^\mu(z) + BQ_{\nu+2r}^\mu(z)\right] , \tag{A16}$$

where

$$A = e^{-2\pi\nu i}\, K_\nu^\mu(\gamma) ,$$

$$B = -e^{-\pi\mu i}\frac{\sin\pi(\nu+\mu)}{\pi\cos\pi\nu} \times\left[e^{-2\pi\nu i}K_\nu^\mu(\gamma) + ie^{\pi\nu i}K_{-\nu-1}^\mu(\gamma)\right] . \tag{A17}$$

[1]B. Coppi, Phys. Rev. Lett. **39**, 939 (1977).
[2]D. Dobrott, D. B. Nelson, J. M. Greene, A. H. Glasser, M. S. Chance, and E. A. Frieman, Phys. Rev. Lett. **39**, 943 (1977).
[3]J. W. Connor, R. J. Hastie, and J. B. Taylor, Phys. Rev. Lett. **40**, 396 (1978).
[4]C. Mercier, Plasma Phys. **21**, 589 (1979).
[5]M. S. Chance, R. L. Dewar, E. A. Frieman, A. H. Glasser, J. M. Greene, R. C. Grimm, S. C. Jardin, J. L. Johnson, J. Manickam, M. Okabayashi, and A. M. Todd, in *Plasma Physics and Controlled Nuclear Fusion Research*, Vol. 1 (I.A.E.A., Vienna, 1979), p. 677.
[6]B. Coppi, A. Ferriera, and J. J. Ramos, Phys. Rev. Lett. **44**, 990 (1980).
[7]T. M. Antonsen, A. Ferriera, and J. J. Ramos, Plasma Phys. **24**, 197 (1982).
[8]A. B. Mikhailovskii and E. I. Yurchenko, Plasma Phys. **24**, 977 (1982).
[9]R. Y. Dagazian and R. B. Paris, Plasma Phys. **24**, 661 (1982).
[10]R. B. Paris, Euratom—CEA Report EUR-CEA-FC 1155, 1983.
[11]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1964), Chap. 21.
[12]E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., Oxford, 1962), Vol. 1, Chap. 5.
[13]N. I. Akhiezer, and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Pitman and Scottish Academic, London, 1981), Vol. II, p. 523.
[14]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, p. 768.
[15]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966), pp. 208–216.
[16]E. C. Titchmarsh, Proc. R. Soc. London Ser. A **207**, 321 (1951).
[17]*Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1955), Vol. III, pp. 134–145.
[18]J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer, Berlin, 1954), Chap. 3.
[19]E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1965), p. 412ff.
[20]R. Y. Dagazian and R. B. Paris, Los Alamos National Laboratory Report LA-9055-MS, 1982.
[21]J. Meixner, Arch. Math. **5**, 492 (1954).
[22]R. M. Kulsrud, Phys. Fluids **6**, 904 (1963).
[23]S. Yoshikawa and R. B. White, Phys. Fluids **23**, 791 (1980).
[24]L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge U.P., Cambridge, 1960), p. 119.
[25]T. E. Stringer, Nucl. Fusion **15**, 125 (1975).
[26]R. B. Paris, Ann. Phys. (Paris) **9**, 347 (1984).
[27]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978), p. 521.
[28]J. Meixner, Z. Angew. Math. Mech. **28**, 304 (1948).

The resonant thermonuclear reaction rate

H. J. Haubold
Zentralinstitut für Astrophysik, Akademie der Wissenschaften der DDR, DDR-1502 Potsdam-Babelsberg, German Democratic Republic

A. M. Mathai
Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke Street West, Montreal, Province of Quebec, Canada H3A 2K6

(Received 29 January 1986; accepted for publication 9 April 1986)

Basic physical principles for the resonant and nonresonant thermonuclear reaction rates are applied to find their standard representations for nuclear astrophysics. Closed-form representations for the resonant reaction rate are derived in terms of Meijer's G-function. Analytic representations of the resonant and nonresonant nuclear reaction rates are compared and the appearance of Meijer's G-function is discussed in physical terms.

I. INTRODUCTION

At present it is believed that nuclear reactions govern all aspects of the chemical evolution of the universe. Any deep understanding of nuclear reactions in the cosmological and stellar nucleosynthesis must be based on a sound theory of nuclear reaction dynamics.[1,2] The rate of nuclear reactions is usually expressed in terms of a quantum-mechanical quantity known as the reaction cross section. In describing a reaction involving neutrons, protons, alpha particles, or heavier atomic nuclei it is convenient to introduce an effective cross section of the target nuclei that indicates what area of the incident beam is affected. If the relative velocity of the reacting particles is v, the nuclear cross section for reactions is σ, and the number densities of incident and target particles are n_m $(m = i,j,k,l)$, then the number of two-particle reactions, abbreviated as $i + j \to k + l$, in unit volume and in unit time is

$$R_{ij} = n_i n_j \sigma v . \tag{1}$$

However, under the physical conditions realized in the cosmological and stellar nucleosynthesis there is no distinction between incident and target particles. This does not matter as far as the relative velocity of the reacting particles is concerned. In addition, not all the relative velocities of the reacting particles are the same and the nuclear cross section depends strongly on the relative velocities. Thus, formula (1) for the number of two-particle reactions has to be generalized by integrating over all values of the relative velocity of the particles. If the corrections due to quantum-mechanical effects (long-range interactions, many-body collisions, high-energy fluctuations) and relativity are negligible, for both the incident and target particles, then we can assume a Maxwell–Boltzmann distribution of the relative velocity,

$$dv\, f(v) = \left(\frac{\mu}{2\pi kT}\right)^{3/2} \exp\left\{-\frac{\mu v^2}{2kT}\right\} 4\pi v^2\, dv , \tag{2}$$

where $\mu = m_i m_j/(m_i + m_j)$ denotes the reduced mass of the two particles under consideration, T is the temperature, and k is the Boltzmann constant. The assumption of the Maxwell–Boltzmann distribution also implies that the nuclear reaction proceeds under the physical conditions of thermodynamic equilibrium.[3] In thermodynamic equilibrium all of the physical properties of the system can be calculated in terms of its density, temperature, and chemical composition alone.

In terms of the relative kinetic energy $E = \mu v^2/2$, we obtain, with (1) and (2),

$$r_{ij} = n_i n_j \langle \sigma v \rangle , \tag{3}$$

where

$$\langle \sigma v \rangle = \left(\frac{8}{\pi\mu}\right)^{1/2} \left(\frac{1}{kT}\right)^{3/2} \int_0^\infty dE\, \sigma(E) E \exp\left\{-\frac{E}{kT}\right\} \tag{4}$$

is the probability per unit time that two particles, confined to a unit volume, will react with each other.[3]

In the last few years there has been considerable interest in the development of mathematical methods for analytic representations of the thermonuclear reaction rate r_{ij} in (3) with (4) depending on the specific analytic structure of the cross section $\sigma(E)$ in (4).[2–5] A review of early work on the closed-form representation of nuclear reaction rates has been given by Haubold and John.[6]

In Sec. II we summarize the results for the closed-form representation of the nonresonant thermonuclear reaction rate obtained by us recently.[3] In Sec. III we specify the nuclear cross section $\sigma(E)$ of (4) for the resonant reaction rate by means of the one-level Breit–Wigner dispersion formula. A most general parametrization for the resonant nuclear reaction rate integral will be given that will be suitable for deducing all special cases of physical interest. By taking advantage of the theory of generalized hypergeometric functions known as Meijer's G-function we obtain the closed-form representation of the resonant nuclear reaction rate. The type of results obtained will be discussed and the connections between the analytic representation of the nonresonant reaction rate given in Sec. II and the resonant reaction rate will be pointed out. Finally the conclusions will be given in Sec. IV.

II. ANALYTIC RESULTS FOR THE NONRESONANT THERMONUCLEAR REACTION RATE

Here we shall compute the quantity $\langle \sigma v \rangle$ in (4) for the standard case of a low-energy nuclear reaction far from any resonance as discussed by Frank-Kameneckij[7] and Clayton.[2] For reactions induced by charged particles such as pro-

 0022-2488/86/082203-05$02.50

tons, alpha particles, or heavier nuclei the rapid variation of the cross section $\sigma(E)$ in (4) is tackled by eliminating the exponentially varying term of the Gamow penetration factor governing transmission through the Coulomb barrier

$$\sigma(E) = [S(E)/E]\exp\{-2\pi\eta(E)\}\,, \tag{5}$$

$$\eta(E) = (\mu/2)^{1/2}(Z_i Z_j e^2/\hbar E^{1/2})\,, \tag{6}$$

where Z_i and Z_j are the charge numbers of the interacting nuclei, $\hbar$ is Planck's quantum of action, and e is the quantum of electric charge. The dependence of $\sigma(E)$ to the inverse of the relative kinetic energy E goes back to the quantum-mechanical interaction between two particles which is always proportional to a geometrical factor, $\pi\lambda^2 \propto E^{-1}$, where λ is the de Broglie wavelength. Equation (5) defines the cross section factor $S(E)$, representing the intrinsically nuclear parts of the probability for the occurrence of a nuclear reaction. The cross section factor $S(E)$ is often found to be constant or a slowly varying function of energy over a limited energy range and it may be conveniently expressed in terms of the power series expansion[1,2,5,6]

$$S(E) = S(0) + \frac{dS(0)}{dE}E + \frac{1}{2}\frac{d^2S(0)}{dE^2}E^2\,. \tag{7}$$

If we use relations (5)–(7) we may write (4) with the substitution $y = E/kT$ in the form

$$\langle\sigma v\rangle = \left(\frac{8}{\pi\mu}\right)^{1/2}\sum_{\nu=0}^{2}\frac{1}{(kT)^{-\nu+1/2}}\frac{S^{(\nu)}(0)}{\nu!} \times \int_0^\infty dy\, e^{-y}y^{\nu}e^{-z/y^{1/2}}\,, \tag{8}$$

where

$$z = 2\pi(\mu/2kT)^{1/2}(Z_i Z_j e^2/\hbar)\,. \tag{9}$$

Equation (8) is the basic parametrized representation of the nonresonant thermonuclear reaction probability per unit time that two charged particles, confined to a unit volume, will react with each other.

The closed-form representation of the parameter-dependent integral in (8),

$$Ni_\nu(z) = \int_0^\infty dy\, e^{-y}y^{\nu}e^{-z/y^{1/2}}\,, \tag{10}$$

can be obtained by means of the integration theory of generalized hypergeometric functions as discussed by Mathai and Saxena.[8] Here we quote the final result derived by Haubold and Mathai[3]:

$$Ni_\nu(z) = \frac{1}{\pi^{1/2}}G_{0,3}^{3,0}\left(\frac{z^2}{4}\middle|_{0,1/2,1+\nu}\right), \tag{11}$$

where z is defined in (9); $G_{p,q}^{m,n}(z|_{b_q}^{a_p})$ denotes Meijer's G-function (see Mathai and Saxena[8]). Representations of the right-hand side of (11), which are suitable for the numerical evaluation of the collision probability integral for all classes of the parameter ν, are given in the Appendix (cf. also Haubold and Mathai[3]). The result (11) generalizes approximated representations of the integral (8) contained in the papers of Bahcall[5] and Critchfield[4] as discussed in detail by Haubold and John.[6]

III. ANALYTIC RESULTS FOR THE RESONANT THERMONUCLEAR REACTION RATE

In the following we will illustrate the approach for the closed-form evaluation of the reaction probability (4) when a strong resonance occurs in the low-energy range under consideration in cosmological and stellar nuclear reactions. Resonant cross sections are many orders of magnitude greater than nonresonant cross sections (5) at energies near the resonance, so that a resonance can dominate the reaction probability (4) in spite of the required integration over the particle velocity distribution function given in (2).[1,2,7]

A. Breit–Wigner resonance cross section

For a single resonance at $E = E_r$, where E_r is a resonance energy, the cross section $\sigma(E)$ of the nuclear reaction proceeding via the formation of a compound nucleus can be represented as a function of energy in terms of the Breit–Wigner formula[2,7]

$$\sigma(E) = \pi\bar{\lambda}^2\omega\Gamma_{ij}\Gamma_{kl}/[(E_r - E)^2 + (\tfrac{1}{2}\Gamma)^2]\,, \tag{12}$$

where $\bar{\lambda} = \hbar/(2\mu E)^{1/2}$ is the reduced de Broglie wavelength corresponding to the relative motion between particles i and j when far apart. The first factor in (12) is the maximum possible cross section for a partial wave with zero angular momentum. The second statistical factor, $\omega = (2J+1)/[(2J_i+1)(2J_j+1)]$, is of order unity and accounts for the spins or internal angular momenta of the interacting particles, where J is the angular momentum of the resonant state, and J_i and J_j are the angular momenta of particles of the type i and j, respectively. The total width Γ of the resonance state is given by $\Gamma = \hbar/\tau = \Gamma_{ij} + \Gamma_{kl} + \cdots$, where τ is the effective lifetime of the state. The partial width Γ_{ij} is the width for reemission of particles i with j, and Γ_{kl} is the partial width for emission of particles k and l.

The partial width Γ_{ij} for the absorption of a certain particle by the compound nucleus is a strongly energy-dependent function and can be written[2,7] as

$$\Gamma_{ij} = (2^{3/2}\mu^{1/2}R_0 D/\hbar)E^{1/2}P(E)\,, \tag{13}$$

where R_0 is the characteristic wavelength of nucleons inside the nucleus, D is the energy spacing between neighboring states of the compound nucleus, and $P(E)$ denotes the Coulomb barrier penetration factor. At low energy the zero angular momentum interactions dominate and the barrier penetration factor $P(E)$ can be written as

$$P(E) = 2\pi\eta(E)\exp\{-2\pi\eta(E)\}\,, \tag{14}$$

where $\eta(E)$ is defined in (6).

For the total width, which increases with increasing excitation energy of the compound nucleus, we assume an *ad hoc* linear energy dependence in the form

$$\Gamma(E) = \Gamma_0 + \Gamma_1 E\,, \tag{15}$$

where Γ_0 and Γ_1 are empirical constants measured in nuclear experiments.

Inserting (13)–(15) into (12) we obtain the parametrized form of the Breit–Wigner one-resonance-level formula

$$\sigma(E) = \left[\frac{\pi^2\hbar^2}{2\mu E}\right]\left[\frac{2^{3/2}\mu^{1/2}R_0E^{1/2}}{\hbar}\frac{2\pi\mu^{1/2}Z_iZ_j e^2}{2^{1/2}\hbar E^{1/2}}\right.$$
$$\left.\times \exp - \left\{\frac{2\pi\mu^{1/2}Z_iZ_j e^2}{2^{1/2}\hbar E^{1/2}}\right\}\right]$$
$$\times\left[\omega\frac{\Gamma_{kl}D}{(E_r-E)^2+(\frac{1}{2}[\Gamma_0+\Gamma_1E])^2}\right]. \quad (16)$$

B. Resonant reaction rate integral

We put the resonant cross section (16) into (4) and get the representation of the reaction probability for resonant thermonuclear reaction as

$$\langle\sigma v\rangle = (2\pi)^{5/2}\frac{Z_iZ_j e^2R_0\omega\Gamma_{kl}D}{\mu^{1/2}(kT)^{3/2}}$$
$$\times\int_0^\infty dE\frac{\exp\{-E/kT-\bar{q}/E^{1/2}\}}{(E_r-E)^2+(\frac{1}{2}[\Gamma_0+\Gamma_1E])^2}, \quad (17)$$

where $\bar{q}$ is given by

$$\bar{q} = 2\pi(\mu/2)^{1/2}(Z_iZ_j e^2/\hbar) = z(kT)^{1/2}, \quad (18)$$

and z is defined in (9). From (17) we remove the integral,

$$R = \int_0^\infty dE\frac{\exp\{-E/kT-\bar{q}/E^{1/2}\}}{(E_r-E)^2+(\frac{1}{2}[\Gamma_0+\Gamma_1E])^2}, \quad (19)$$

which may be rewritten more conveniently as

$$R = \frac{1}{1+(\frac{1}{2}\Gamma_1)^2}\int_0^\infty dE\frac{\exp\{-E/kT-\bar{q}/E^{1/2}\}}{(\widetilde{E}_r-E)^2+(\frac{1}{2}\widetilde{\Gamma})^2}, \quad (20)$$

where $\widetilde{E}_r$ denotes a modified resonance energy,

$$\widetilde{E}_r = (E_r-\tfrac{1}{4}\Gamma_0\Gamma_1)/[1+(\tfrac{1}{2}\Gamma_1)^2], \quad (21)$$

and $\widetilde{\Gamma}$ is a modified total width,

$$\widetilde{\Gamma} = (\Gamma_0+E_r\Gamma_1)/[1+(\tfrac{1}{2}\Gamma_1)^2]. \quad (22)$$

The form of the resonance denominator with Γ_0 is conserved if one transforms $\Gamma_0\to\Gamma_0+\Gamma_1E$. Choosing the variable $E=y/(1+[\frac{1}{2}\Gamma_1]^2)$ leads to

$$R(q,a,b,g) = \int_0^\infty dy\frac{\exp\{-ay-qy^{-1/2}\}}{(b-y)^2+g^2}, \quad (23)$$

where

$$a = 1/kT(1+[\tfrac{1}{2}\Gamma_1]^2), \quad b = E_r-\tfrac{1}{4}\Gamma_0\Gamma_1,$$
$$g = \tfrac{1}{2}(\Gamma_0+E_r\Gamma_1), \quad q = \bar{q}(1+[\tfrac{1}{2}\Gamma_1]^2)^{1/2}. \quad (24)$$

Now we can rewrite (17) with (23) in a more convenient form

$$\langle\sigma v\rangle = (2\pi)^{5/2}\frac{Z_iZ_j e^2R_0\omega\Gamma_{kl}D}{\mu^{1/2}(kT)^{3/2}}\frac{1}{1+(\frac{1}{2}\Gamma_1)^2}R(q,a,b,g). \quad (25)$$

C. Closed-form evaluation of the resonant reaction rate integral

To comprehend cases in which a cross section factor $S(E)$ according to (7) will be introduced in the resonant cross section (16), we consider in the following the more general integral

$$R_1(q,a,b,g;\nu,n,m) = \int_0^\infty dt\, t^\nu\frac{\exp\{-at-qt^{-n/m}\}}{(b-t)^2+g^2}, \quad (26)$$

which includes (23) as a special case for $\nu=0$, $n=1$, and $m=2$. We may replace the denominator $[(b-t)^2+g^2]^{-1}$ in (26) by an equivalent integral for $g^2>0$. That is,

$$\frac{1}{(b-t)^2+g^2} = \int_0^\infty dx\exp\{-[(b-t)^2+g^2]x\}. \quad (27)$$

But, we can also write

$$e^{-x(b-t)^2} = \sum_{k=0}^\infty\frac{(-1)^k}{k!}x^k$$
$$\times\sum_{k_1=0}^{2k}\binom{2k}{k_1}(-1)^{k_1}b^{2k-k_1}t^{k_1}, \quad (28)$$

where, for example,

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}, \quad 0! = 1.$$

From (26)–(28) we have

$$R_1(q,a,b,g;\nu,n,m) = \int_0^\infty dx\, e^{-g^2x}\sum_{k=0}^\infty\frac{(-1)^k}{k!}x^k\sum_{k_1=0}^{2k}\binom{2k}{k_1}(-1)^{k_1}b^{2k-k_1}\int_0^\infty dt\, t^{\nu+k_1}e^{-at}e^{-qt^{-n/m}}. \quad (29)$$

Note that

$$\int_0^\infty dx\, x^ke^{-g^2x} = \int_0^\infty dx\, x^{(k+1)-1}e^{-g^2x} = \frac{k!}{g^2(g^2)^k}, \quad (30)$$

and according to Haubold and Mathai[3] we have

$$\int_0^\infty dt\, t^{\nu+k_1}e^{-at}e^{-qt^{-n/m}} = a^{-(\nu+k_1+1)}(2\pi)^{(1/2)(2-n-m)}m^{1/2}n^{1/2+\nu+k_1}$$
$$\times G^{m+n,0}_{0,m+n}\left(\frac{q^ma^n}{m^mn^n}\Big|_{0,1/m,\ldots,(m-1)/m,(1+\nu+k_1)/n,\ldots,(n+\nu+k_1)/n}\right). \quad (31)$$

It may be noted that the result (31) contains (10) and (11) as a special case ($k_1=0$, $a=1$, $q=z$, $n=1$, $m=2$). Substituting (30) and (31) in (29) we have the following:

$$R_1(q,a,b,g;\nu,n,m) = \sum_{k=0}^{\infty} \frac{(-1)^k}{g^2(g^2)^k} \sum_{k_1=0}^{2k} \binom{2k}{k_1} (-1)^{k_1} b^{2k-k_1} a^{-(\nu+k_1+1)} (2\pi)^{(1/2)(2-n-m)} m^{1/2} n^{1/2+\nu+k_1}$$

$$\times G_{0,m+n}^{m+n,0}\left(\frac{q^m a^n}{m^m n^n} \bigg|_{0,1/m,\ldots,(m-1)/m,(1+\nu+k_1)/n,\ldots,(n+\nu+k_1)/n} \right). \tag{32}$$

Put $n = 1$, $m = 2$, $\nu = 0$, to get (23)

$$R_1(q,a,b,g;0,1,2)$$

$$= R(q,a,b,g) = \frac{1}{g^2 a} \sum_{k=0}^{\infty} \frac{(-1)^k}{(g^2)^k} \times \sum_{k_1=0}^{2k} \binom{2k}{k_1} \frac{(-1)^k}{a^{k_1}} b^{2k-k_1} \frac{1}{\pi^{1/2}} \times G_{0,3}^{3,0}\left(\frac{q^2 a}{4} \bigg|_{0,1/2,1+k_1} \right) \tag{33}$$

for $(b - v/a)^2/g^2 < 1$, where $v = (q^2 a/4)^{1/3}$.

IV. CONCLUSIONS

The central quantity for the description of cosmological and stellar nucleosynthesis is the thermonuclear reaction rate. There are a few basic physical principles that are common to the evaluation of all thermonuclear reaction rates, as discussed by Fowler[1] (Table I on p. 154 and Table II on p. 155). The aim of the present paper is to take into account these basic physical principles for deriving closed-form representations for the two fundamental cases of nuclear reaction rates: the nonresonant and the resonant ones. We can make the following conclusions.

(i) The main energy dependence of the nonresonant nuclear cross section (5) comes from the Gamow penetration factor, which is based on the solution of the Schrödinger equation for the Coulomb wave functions. The average of the nonresonant nuclear cross section (5) over the Maxwell–Boltzmann distribution (2) leads to a parameter-dependent integral (10), which is expressible in closed form by a Meijer's G-function of the type $G_{0,3}^{3,0}((z^2/4)|_{0,1/2,1+\nu})$. We can conclude that always all nonresonant thermonuclear reaction rates find their closed form representation in the Meijer's G-function considered. As can be shown the asymptotic representations of the nonresonant thermonuclear reaction rate for small and large values of the characteristic parameter (9) often used in nuclear astrophysics [1,4,5] follow immediately from asymptotic considerations of the G-function under consideration.[6,8]

(ii) For deriving the closed-form representation of the resonant thermonuclear reaction rate we took into account the full energy dependence of the one-resonant level formula of Breit–Wigner (16). Comparing the resonant result (33) with the nonresonant one (11) we observe that the former is an infinite sum over nonresonant contributions (note that $q^2 a = z^2$). Thus, the direct connection of the closed-form representation of resonant and nonresonant thermonuclear reaction rates is obtained via the Meijer's G-function of the type $G_{0,p}^{p,0}(x|_{b_1,\ldots,b_p})$. For a detailed mathematical discussion of that type of Meijer's G-function see Mathai and Saxena.[8]

(iii) The appearance of Meijer's G-function of the type $G_{0,p}^{p,0}(x|_{b_1,\ldots,b_p})$ in the closed-form representation of nonresonant (11) and resonant (33) thermonuclear reaction rates can be discussed in physical terms. As can be seen from (31) the characteristic numbers of Meijer's G-function under discussion ($p = 1 + 2 = 3$) goes back to the well-known energy dependence of Gamow's barrier penetration factor [cf. (5) and (6)]. The third term of the parameters (b_1,b_2,b_3) of the G-function contains directly the exponent ν of the power series expansion of the cross section factor (7) if it is necessary to take it into account for the evaluation of the nonresonant or resonant thermonuclear reaction rate [cf. (11) and (32)]. That parameter is of some importance for the asymptotic behavior of the Meijer's G-function.[6,8] As will be shown in a forthcoming paper, a modification of the Maxwell–Boltzmann distribution function (2) for averaging the nuclear cross sections will lead to a change in the G-function of the type $G_{0,p}^{p,0}(x|_{b_1,\ldots,b_p})$ giving a G-function of the type $G_{q,p}^{p,0}(z|_{b_1,\ldots,b_p}^{a_1,\ldots,a_q})$.

(iv) The advantage of the representation (33) of the resonant thermonuclear reaction rate in comparison to other representations published earlier[9] is that it is accessible to numerical computation via the series representation given in the Appendix. These computations are needed for special physical problems (cf., e.g., Fetisov and Kopysov[10]).

ACKNOWLEDGMENTS

One of the authors (H.J.H.) is very grateful to Professor Dr. W. A. Fowler and Professor Dr. C. L. Critchfield for their interest in this work.

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for financial assistance.

APPENDIX: SERIES REPRESENTATIONS FOR MEIJER'S G-FUNCTION

Here we give representations of the Meijer's G-function contained in (11) and (33) that are termwise integrable over any finite range of the variable and that can be used for the numerical computation of the nonresonant and resonant thermonuclear reaction rates. In the following ${}_pF_q(a_p;b_q;z)$ denotes the generalized hypergeometric function, $\psi(z)$ is a psi function or digamma function, and

$$(\beta)_r = \frac{\Gamma(\beta + r)}{\Gamma(\beta)} = \beta(\beta+1)\cdots(\beta+r-1), \quad (\beta)_0 = 1$$

is the Pochammer symbol.

Case (i): For $\nu \neq \pm \lambda/2, \lambda = 0,1,2,...,$

$$\frac{1}{\pi^{1/2}} G_{0,3}^{3,0}\left(\frac{z^2}{4}\bigg|_{0,1/2,1+\nu}\right) = \Gamma(1+\nu)\,{}_0F_2\left(-;\frac{1}{2},-\nu;-\frac{z^2}{4}\right) - 2\Gamma\left(\frac{1}{2}+\nu\right)\left(\frac{z^2}{4}\right)^{1/2}{}_0F_2\left(-;\frac{3}{2},\frac{1}{2}-\nu;-\frac{z^2}{4}\right)$$
$$+\frac{\Gamma(-1-\nu)\Gamma(-\frac{1}{2}-\nu)}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{1+\nu}{}_0F_2\left(-;\nu+2,\nu+\frac{3}{2};-\frac{z^2}{4}\right). \tag{A1}$$

Case (ii): For ν a positive integer,

$$\frac{1}{\pi^{1/2}} G_{0,3}^{3,0}\left(\frac{z^2}{4}\bigg|_{0,1/2,1+\nu}\right) = \Gamma(1+\nu)\sum_{r=0}^{\nu}\frac{1}{(\frac{1}{2})_r(-\nu)_r\, r!}\left(\frac{z^2}{4}\right)^r - 2\Gamma\left(\frac{1}{2}+\nu\right)\left(\frac{z^2}{4}\right)^{1/2}{}_0F_2\left(-;\frac{3}{2},\frac{1}{2}-\nu;-\frac{z^2}{4}\right)$$
$$+\frac{1}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{1+\nu}\sum_{r=0}^{\infty}\left(\frac{z^2}{4}\right)^r\left\{-\ln\left(\frac{z^2}{4}\right)+A_r\right\}B_r, \tag{A2}$$

where

$$A_r = \psi(r+1)+\psi(r+\nu+2)+\psi(-\tfrac{1}{2}-\nu-r),\quad B_r = (-1)^{1+\nu+r}\Gamma(-\tfrac{1}{2}-\nu)/r!(r+\nu+1)!(\tfrac{3}{2}+\nu)_r.$$

Case (iii): For ν a negative integer,

$$\frac{1}{\pi^{1/2}} G_{0,3}^{3,0}\left(\frac{z^2}{4}\bigg|_{0,1/2,1+\nu}\right) = \frac{\Gamma(-\nu-1)\Gamma(-\frac{1}{2}-\nu)}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{1+\nu}\sum_{r=0}^{-\nu-2}\frac{(-1)^r}{r!}\left(\frac{z^2}{4}\right)^r\frac{1}{(\nu+2)_r(\nu+\frac{3}{2})_r}$$
$$-2\Gamma\left(\frac{1}{2}+\nu\right)\left(\frac{z^2}{4}\right)^{1/2}{}_0F_2\left(-;\frac{3}{2},\frac{1}{2}-\nu,-\frac{z^2}{4}\right)+\sum_{r=0}^{\infty}\frac{1}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^r\left\{-\ln\left(\frac{z^2}{4}\right)+A'_r\right\}B'_r, \tag{A3}$$

where

$$A'_r = \psi(r+1)+\psi(r-\nu)+\psi(\tfrac{1}{2}-r),\quad B'_r = \Gamma(\tfrac{1}{2}-r)(-1)^{\nu+1}/r!(r-\nu-1)!.$$

Case (iv): For ν a positive half-integer, $\nu = m+\frac{1}{2}$, $m = 0,1,2,...,$

$$\frac{1}{\pi^{1/2}} G_{0,3}^{3,0}\left(\frac{z^2}{4}\bigg|_{0,1/2,1+\nu}\right) = \Gamma\left(m+\frac{3}{2}\right){}_0F_2\left(-;-\frac{1}{2},-m-\frac{1}{2};-\frac{z^2}{4}\right) - 2\Gamma(m+1)\left(\frac{z^2}{4}\right)^{1/2}\sum_{r=0}^{\infty}\frac{(-1)^r}{r!}$$
$$\times\left(\frac{z^2}{4}\right)^r\frac{1}{(\frac{3}{2})_r(-m)_r}+\frac{1}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{m+3/2}\sum_{r=0}^{\infty}\left(\frac{z^2}{4}\right)^r\left\{-\ln\left(\frac{z^2}{4}\right)+D_r\right\}C_r, \tag{A4}$$

where

$$D_r = \psi(r+1)+\psi(m+r+2)+\psi(-m-\tfrac{3}{2}-r),\quad C_r = (-1)^{m+1+r}\Gamma(-m-\tfrac{3}{2})/r!(m+1+r)!(m+\tfrac{5}{2})_r.$$

Case (v): For ν a negative half-integer, $\nu = m-\frac{1}{2}$, $m = 0,1,2,...,$

$$\frac{1}{\pi^{1/2}} G_{0,3}^{3,0}\left(\frac{z^2}{4}\bigg|_{0,1/2,1+\nu}\right) = \Gamma\left(\frac{1}{2}-m\right){}_0F_2\left(-;\frac{1}{2},m+\frac{1}{2};-\frac{z^2}{4}\right)+\frac{\Gamma(m)\Gamma(m-\frac{1}{2})}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{-m+1/2}$$
$$\times\sum_{r=0}^{m-1}\frac{(-1)^r}{r!}\left(\frac{z^2}{4}\right)^r\frac{1}{(-m+1)_r(-m+\frac{3}{2})_r}+\frac{1}{\Gamma(\frac{1}{2})}\left(\frac{z^2}{4}\right)^{1/2}\sum_{r=0}^{\infty}\left(\frac{z^2}{4}\right)^r$$
$$\times\left\{-\ln\left(\frac{z^2}{4}\right)+D'_r\right\}C'_r, \tag{A5}$$

where

$$D'_r = \psi(1)+\psi(m+r+1)+\psi(-\tfrac{1}{2}-r),\quad C'_r = (-1)^{m+r}\Gamma(-\tfrac{1}{2})/r!(m+r)!(\tfrac{3}{2})_r.$$

[1]W. A. Fowler, Rev. Mod. Phys. **56**, 149 (1984).
[2]D. D. Clayton, *Principles of Stellar Evolution and Nucleosynthesis* (McGraw-Hill, New York, 1968).
[3]H. J. Haubold and A. M. Mathai, Ann. Phys. (Leipzig) **41**, 380 (1984).
[4]C. L. Critchfield, in *Cosmology, Fusion and Other Matters*, George Gamow Memorial Volume, edited by F. Reines (Univ. Colorado P., Boulder, 1972).
[5]J. N. Bahcall, Astrophys. J. **143**, 259 (1966).
[6]H. J. Haubold and R. W. John, Astron. Nachr. **303**, 161 (1982).
[7]D. A. Frank-Kameneckij, *Physikalische Prozesse in Inneren von Sternen* (Fizmatgiz, Moskow, 1959) (in Russian).
[8]A. M. Mathai and R. K. Saxena, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences, Lecture Notes in Mathematics*, Vol. 348 (Springer, Berlin, 1973).
[9]H. J. Haubold and R. W. John, Astron. Nachr. **300**, 63 (1979); **303**, 217 (1982).
[10]V. N. Fetisov and Yu. S. Kopysov, Nucl. Phys. A **239**, 511 (1975).

ERRATA

Erratum: The relation of a theory of countable sets to the field equations of physics [J. Math. Phys. 26, 585 (1985)]

D. J. BenDaniel
Malott Hall, Cornell University, Ithaca, New York 14853-4201

(Received 20 March 1986; accepted for publication 30 April 1986)

For reasons of both typographical and author-responsible errors, the reader is asked to replace the paragraph beginning "Let us now construct sets which have physical relevance..." by the paragraph following.

Let us now construct sets which have physical relevance, that is, functions having a range and domain of continuous real of complex variables. We first state the simple result that the range of a constant function is given by AC. To show the range of nonconstant functions of a real variable, we start by looking at bounded continuous bijective mappings from the real numbers into the real numbers. In theory T, since we have only a countable set of reals, we obtain the basic result that an interval of the real line can have nonzero length if and only if its points map one-to-one with the reals. It follows from this result that a mapping $\phi(u,x)$ from all the reals in $(0,1)$ into the reals in $(0,1)$ can be bijective only if it is constructed of pieces such that for each piece

$$\forall x \in (0,1) \exists u \in (0,1) \ [\phi(x,u) \wedge \forall x' \in (0,1) \forall u' \in (0,1) \exists C_{xu} \exists C_{ux} \\ (C_{xu} > 0 \wedge C_{ux} > 0 \wedge \phi(x',u') \rightarrow |x - x'| \geqslant C_{xu} |u - u'| \\ \wedge |u - u'| \geqslant C_{ux} |x - x'|)].$$

The range (set of all u) exists by ABR. This statement describes a continuous, strictly monotonic piece. It follows that any change in one variable must be accompanied by a change in the other and, applied to physical variables, may be considered a concise description of universal tight coupling. Furthermore, since the reciprocals of C_{xu} and C_{ux} each have a least upper bound at every set of points x, u, x', u', we obtain

$$\forall x \in (0,1) \exists u \in (0,1) [\phi(x,u) \wedge \exists C_1 \exists C_2 \forall x' \in (0,1) \forall u' \in (0,1) \\ (1/C_1 > 0 \wedge 1/C_2 > 0 \wedge \phi(x',u') \rightarrow |u - u'| \leqslant C_1 |x - x'| \\ \wedge |x - x'| \leqslant C_2 |u - u'|)].$$

This result is a bi-Lipschitz condition, thus $u(x)$ and $x(u)$ are both Lipschitz continuous.

Also, in the paragraph following, please replace the word "bi-Lipschitz" by "Lipschitz continuous."

Erratum: Scalar formalism for quantum electrodynamics [J. Math. Phys. 26, 1348 (1985)]

Levere C. Hostler
Wilkes College, Wilkes Barre, Pennsylvania 18766

(Received 23 September 1985; accepted for publication 30 April 1986)

Since publication of "Scalar formalism for quantum electrodynamics," a number of related references, listed below, have been pointed out to me. Feynman rules equivalent to our Table II, p. 1352, were obtained before in the article by Brown[1] using the c-number formalism, and were obtained again by Tonin,[2] who second-quantized a special Lagrangian that was developed in Ref. 1 for the second-order Dirac equation.

The additional references are listed below.

[1] L. M. Brown, Phys. Rev. **111**, 957 (1958).
[2] M. Tonin, Nuovo Cimento **14**, 1108 (1959).
[3] W. R. Theis, Fortschr. Phys. **7**, 559 (1959).
[4] H. Pietschmann, Acta Phys. Austriaca **14**, 63 (1961).
[5] L. M. Brown, "Two-component fermion theory," in *Lectures in Theoretical Physics*, Vol. IV (Interscience, New York, 1962).
[6] L. M. Brown, "Quantum electrodynamics at high energy," in *Topics in Theoretical Physics, Proceedings of the Liperi Summer School in Theoretical Physics 1967*, edited by Christofer Cronstrom (Gordon and Breach, New York, 1969), p. 113.

 0022-2488/86/082208-01$02.50

Erratum: Petrov type *D* perfect-fluid solutions in generalized Kerr–Schild form [J. Math. Phys. 27, 265 (1986)]

J. Martín and J. M. M. Senovilla
Departamento de Física Teórica, Universidad de Salamanca, 37008-Salamanca, Spain

(Received 10 April 1986; accepted for publication 7 May 1986)

There is a mistake in Eq. (5.20). The correct version of this equation reads as follows:

$$\rho\Delta U+\bar{\delta}\delta U-3\bar{\alpha}\bar{\delta}U+U\,[3\alpha\bar{\alpha}-\rho(5\mu+3\bar{\gamma}+\gamma)]=0.$$

Consequently, Eqs. (5.27) and (5.28) become, respectively,

$$\rho\Delta f-\rho f(\mu+\bar{\gamma}-\gamma)+f(\bar{\delta}f+\alpha f)=0,$$

$$\Delta\bar{\alpha}=\bar{\alpha}(\mu+\bar{\gamma}-\gamma).$$

The authors are grateful to F. Martín-Pascual for pointing out the mistake in question.

Erratum: Imparting to a Bianchi type II space-time [J. Math. Phys. 27, 417 (1986)]

M. J. Rebouças
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E14NS, England
J. B. S. d'Olival
Departamento de Física, Universidade Federal Fluminense, 24210 Niteroi RJ, Brazil

(Received 1 May 1986; accepted for publication 13 May 1986)

Equation (3.9) should read

$$(\dot{B}/B)^{\cdot}-(2B)^{-2}=0. \tag{3.9}$$

Equation (3.11) should read

$$B=(1/2\beta)\cosh\beta(t-t_0). \tag{3.11}$$

