Deformation and contraction in Clifford algebras

R. Ablamowicz
Department of Mathematics, Gannon University, Erie, Pennsylvania 16541

(Received 6 February 1985; accepted for publication 3 October 1985)

Spin ideals of Clifford algebras over quadratic spaces of any rank are constructed through contractive limits of suitably deformed classical spin ideals of nondegenerate Clifford algebras. The deformation of the algebraic structures, including the standard Witt basis for the quadratic space, results only from a deformation of the underlying quadratic form. It is shown that a contractive limit of deformed twistor spaces, considered as spin ideals of the Dirac–Clifford algebra, provides a decomposable representation space for the Galilei–Clifford algebra. The limit spin ideals of degenerate Clifford algebras are then decomposed into indecomposable Clifford modules.

I. INTRODUCTION

In the theory of spin representations of Clifford algebras generators of spin (minimal left) ideals are usually constructed through a Witt decomposition of underlying quadratic space.[1–4] In particular, each spin ideal is generated by a primitive idempotent.[5,6] The spin (left regular) representations of Clifford, Pin, and Spin groups can then be naturally considered in such ideals viewed as Clifford modules.[7,8] However, the existing theory is limited to algebras over vector spaces endowed with nondegenerate quadratic forms and cannot be directly applied to algebras (called here degenerate) when the quadratic form is degenerate.[9,10] The degenerate Clifford algebras studied in this paper provide yet another generalization of classical Clifford algebras enjoying a revived interest among physicists.[11–14]

It is well known that in the category of Z_2-graded algebras a degenerate Clifford algebra is isomorphic to a graded tensor product of a Clifford algebra and an exterior algebra associated with the orthogonal complement of the underlying quadratic space.[15] Instead of adopting this point of view we rather consider a two sided nilpotent ideal (the Jacobson radical) contained in such algebra and generated by the complement. We construct the spin representation and spin ideal for the Clifford algebra of quadratic form of arbitrary rank. This construction seems to be novel in that it utilizes suitably defined (see Sec. II) deformation and contraction of a Witt basis in a nondegenerate quadratic space. In fact, we have a continuous family of isomorphic Clifford algebras and associated algebraic structures.

In Sec. III we study deformed classical spin ideals and spin representations. It is shown that every generator of a deformed spin ideal has a nonzero contractive limit and that the set of these limits provides a basis for a limit spin ideal (space). In an example, a spin ideal of the Galilei–Clifford algebra is viewed as the limit of a family of deformed twistor spaces. Using well-known structure theorems on Clifford algebras[1,4,15] the spin representation is then decomposed into indecomposable components.

In Sec. IV we examine structure of the limit ideals of degenerate Clifford algebras for arbitrary rank. It is shown that such ideals can be decomposed into direct sums of simple Clifford modules invariant under nondegenerate Clifford subalgebras.

II. DEFORMED WITT BASIS AND ITS CONTRACTION

We review first several definitions and state Witt's decomposition theorem for an arbitrary quadratic space.[15] Let (V,Q) be a finite-dimensional quadratic vector space over a field K (char $K \neq 2$) and let B be the symmetric bilinear form associated with Q. Form Q is said to be *nondegenerate* (or *regular*) (resp. *degenerate*), if $B(x,y) = 0$ for every y in V implies $x = 0$ (resp. if there exists a nonzero vector x orthogonal to the entire space). We say that a nonzero vector x in V is *isotropic* if $Q(x) = 0$, and say that x is *anisotropic* otherwise. The quadratic space (V,Q) is said to be *isotropic* if it contains a (nonzero) isotropic vector, and is said to be *anisotropic* otherwise. It is *totally isotropic* if all vectors are isotropic. A direct sum of totally isotropic spaces orthogonal to each other is again a totally isotropic space. Thus a set of totally isotropic subspaces of (V,Q) ordered by inclusion contains a *maximal totally isotropic* subspace. It is well known that all maximal totally isotropic subspaces of a nondegenerate quadratic space have the same dimension called the *Witt index* of Q. Finally, a two-dimensional nondegenerate isotropic quadratic space is called the *hyperbolic plane* and an orthogonal sum of hyperbolic planes is called a *hyperbolic space*.

Theorem 2.1 (Witt Decomposition): Any quadratic space (V,Q) splits into an orthogonal sum, $(V_t, Q_t) \perp (V_h, Q_h) \perp (V_a, Q_a)$, where V_t is totally isotropic, V_h is hyperbolic (or zero), and V_a is anisotropic. Furthermore, the isometry types of V_t, V_h, and V_a are all uniquely determined.

Corollary 2.2: Let F be a maximal totally isotropic subspace of (V,Q) of dimension r.

(i) There exists a maximal totally isotropic subspace F' such that $F \cap F' = \{0\}$, and V_h is isomorphic to the direct sum $F \oplus F'$.

(ii) For every basis $\{f_i\}$ of F there exists a basis $\{f'_j\}$ of F' such that $B(f_i, f'_j) = \delta_{ij}$ (Kronecker delta), $B(f_i, f_j) = B(f'_i, f'_j) = 0, i,j = 1,...,r$.

The basis elements $\{f_i, f'_j\}$ of $F \oplus F'$ together with an orthogonal basis of V_a form a *Witt basis* of $V/V^\perp$ associated with its decomposition into the hyperbolic and anisotropic components only. [Here, $V^\perp$ denotes the orthogonal complement of (V,Q) itself.]

We limit our attention to real and complex vector spaces. Throughout this work $R^{d,p,k}$ denotes a real space of dimension $n = d + p + k$ endowed with a quadratic form Q

 0022-2488/86/020423-05$02.50 © 1986 American Institute of Physics

of rank $p+k$ and signature (d,p,k) (that is, $\dim\ker Q = \dim R_t^{d,p,k} = d$, and the diagonalized form of Q contains k plus 1's and p minus 1's), whereas the universal Clifford algebra over $R^{d,p,k}$ is denoted by $R_{d,p,k}$. We call them both *degenerate* when $d \neq 0$. When $d=0$, the Witt index of Q is equal to $\min\{p,k\}$ and we write $R^{p,k}$ and $R_{p,k}$ instead of $R^{0,p,k}$ and $R_{0,p,k}$, respectively. (It is also assumed that $k \leqslant p$.)

Let $\{e_1,...,e_n\}$ be an orthogonal basis in $V = R^{p,k}$ and let $V_h \perp V_a$ be its Witt decomposition. Then, following Refs. 1 and 2, a Witt basis for V can be represented, for example, as

$$\begin{aligned} F &= \mathrm{span}_R\{x_i = \lambda(e_i + e_{n-i+1})\},\\ F' &= \mathrm{span}_R\{y_i = \lambda(e_i - e_{n-i+1})\},\\ V_a &= \mathrm{span}_R\{e_{k+1},...,e_{n-k}\}, \end{aligned} \tag{2.1}$$

where $i = 1,...,k$ and λ is a normalizing factor $(2\lambda^2 = 1)$.

Definition 2.3: A (d_1,d_2)-*deformation* of a quadratic form Q of signature $(0,p,k)$ is a quadratic form Q^ϵ such that

$$\begin{aligned} &Q^\epsilon(e_i) = \epsilon^2, \quad i = 1,...,d_2,\\ &Q^\epsilon(e_i) = 1, \quad i = d_2+1,...,k,\\ &Q^\epsilon(e_i) = -1, \quad i = k+1,...,k+p',\\ &Q^\epsilon(e_i) = -\epsilon^2, \quad i = k+p'+1,...,n, \end{aligned} \tag{2.2}$$

where ϵ is a deformation parameter $0 < \epsilon \leqslant 1$ and $p' + d_1 = p$, $k' + d_2 = k$. We denote by $R^\epsilon_{p,k}$ the (d_1,d_2)-deformation of the Clifford algebra $R_{p,k}$ associated with the quadratic space $V^\epsilon = (R^{p,k}, Q^\epsilon)$. Obviously, since all deformations of Q are equivalent (that is, belong to the equivalence class containing Q), the quadratic spaces V^ϵ are isometric and the Clifford algebras $R^\epsilon_{p,k}$ are isomorphic.

Definition 2.4: A (d_1,d_2)-deformation of a Witt basis for V is obtained by replacing in (2.1) every vector e_i such that $Q^\epsilon(e_i) = \pm\epsilon^2$ with e_i/ϵ.

Lemma 2.5: A (d_1,d_2)-deformation of a Witt basis for V provides a Witt basis for the (d_1,d_2)-deformation V^ϵ.

Proof: By the remark above and Theorem 2.1 the isometry types of F^ϵ, F'^ϵ, and V_a^ϵ are the same as that of F, F', and V_a. Denoting now the basis elements of $F^\epsilon \oplus F'^\epsilon$ by $\{x_i^\epsilon, y_i^\epsilon\}$ we can easily verify that they provide a Witt basis for the hyperbolic component of V orthogonal to V_a^ϵ. □

Definition 2.6: A *contraction* of Q is the limit form Q^0 obtained from a (d_1,d_2) deformation Q^ϵ of Q when $\epsilon \to 0$ in (2.2).

Notice that the signature of Q^0 is (d,p',k'), $d = d_1 + d_2$, that is, the form Q^0 is degenerate. We will formally write $Q^\epsilon \to Q^0$, $V^\epsilon \to V^0$, and $R^\epsilon_{p,k} \to R_{d,p',k'}$ as $\epsilon \to 0$. In Lemma 2.7 we show how to contract a deformed Witt basis for V^ϵ to a Witt basis for V^0.

Lemma 2.7: Let V^0 be the contractive limit of a (d_1,d_2) deformation V^ϵ and let $(F^\epsilon \oplus F'^\epsilon)\perp V_a^\epsilon$ be a Witt decomposition of V^ϵ. Then there exists a contractive limit of a (d_1,d_2) deformation of a Witt basis of V^ϵ that provides a Witt basis for V^0.

Proof: We consider only the case when $k \leqslant d_1$, since other cases may be treated similarly. Then a Witt basis of V^ϵ is given by a (d_1,d_2) deformation of (2.1). We list the x^ϵ part and the anisotropic part below:

$$\begin{aligned} &\{x_1^\epsilon = \lambda(e_1+e_n)/\epsilon,...,x_{d_2}^\epsilon = \lambda(e_{d_2}+e_{p+k'+1})/\epsilon,\\ &\quad x_{d_2+1}^\epsilon = \lambda(e_{d_2+1}+e_{p+k'}/\epsilon,\\ &\quad ...,x_k^\epsilon = \lambda(e_k+e_{p+1}/\epsilon)\},\\ &\{e_{k+1}/\epsilon,...,e_{d_1}/\epsilon, e_{d_1+1},...,e_p\}. \end{aligned} \tag{2.3}$$

Vectors $\{y_i^\epsilon\}$ differ from $\{x_i^\epsilon\}$ only by the sign, that is, $y_1 = \lambda(e_1 - e_n)/\epsilon$, etc. where $n = p+k$. Then we have the following *contractive limits*:

$$\begin{aligned} &\epsilon x_i^\epsilon \to \tilde{x}_i = \lambda(\tilde{e}_i + \tilde{e}_{n-i+1}),\\ &\epsilon y_i^\epsilon \to \tilde{y}_i = \lambda(\tilde{e}_i - \tilde{e}_{n-i+1}), \quad i = 1,...,d_2,\\ &\epsilon(x_i^\epsilon + y_i^\epsilon)/2 \to \lambda e_i,\\ &\epsilon(x_i^\epsilon - y_i^\epsilon)/2 \to \lambda\tilde{e}_{n-i+1}, \quad i = d_2+1,...,k,\\ &\epsilon(e_{k+i}/\epsilon) \to \tilde{e}_{k+i}, \quad i = 1,...,d_1-k,\\ &e_i \to e_i, \quad i = d_1+1,...,p, \end{aligned} \tag{2.4}$$

where $\sim$ is to remind us that certain basis vectors $\{\tilde{e}_i\}$ belong now to the kernel of Q^0. There are then d isotropic basis vectors spanning $\ker Q^0$ in V^0. Also, we see that the remaining vectors span a $(k'+p')$-dimensional nondegenerate subspace of V^0 orthogonal to $\ker Q^0$ and isomorphic to $R^{p',k'}$. Applying now Theorem 2.1 and Eqs. (2.1) we get $R^{p',k'} = (\tilde{F}\oplus\tilde{F}')\perp V_a^0$, where $\tilde{F}$ and $\tilde{F}'$ are maximal totally isotropic subspaces of dimension $\min\{p',k'\}$ and V_a^0 is anisotropic of dimension $|p'-k'|$. Finally, $V^0 = \ker Q^0 \perp (\tilde{F}\oplus\tilde{F}')\perp V_a^0$ and a Witt basis of V^0 is the contractive limit of (2.3). □

Now we present an example of (0,1)-deformation of the real Minkowski space-time $R^{3,1}$ of special relativity, its Witt decomposition, and contraction leading to the Galilei space-time $R^{1,3,0}$ of classical mechanics. (For a comparison of the Galileian and Lorentzian structures of space-time see Ref. 16.)

Example 2.8: Let $\{e_1,e_2,e_3,e_4\}$ be an orthogonal basis in $R^{3,1}$ such that $Q(e_1) = 1$ and $Q(e_i) = -1$, $i = 2,3,4$. Then Q has index 1 and $R^{3,1}$ has a Witt decomposition $(F\oplus F')\perp G$,

$$\begin{aligned} F &= \mathrm{span}_R\{x_1 = \lambda(e_1+e_4)\},\\ F' &= \mathrm{span}_R\{y_1 = \lambda(e_1-e_4)\},\\ G &= \mathrm{span}_R\{e_2,e_3\}. \end{aligned} \tag{2.5}$$

Then the (0,1)-deformation of (2.5) gives a Witt basis for $V^\epsilon = (R^{3,1}, Q^\epsilon)$ and a Witt decomposition $(F^\epsilon \oplus F'^\epsilon)\perp G^\epsilon$,

$$\begin{aligned} F^\epsilon &= \mathrm{span}_R\{x_1^\epsilon = \lambda(e_1/\epsilon + e_4)\},\\ F'^\epsilon &= \mathrm{span}_R\{y_1 = \lambda(e_1/\epsilon - e_4)\},\\ G^\epsilon &= \mathrm{span}_R\{e_2,e_3\}, \end{aligned} \tag{2.6}$$

while $Q^\epsilon(e_1) = \epsilon^2$ and $Q^\epsilon(e_i) = -1$, $i = 2,3,4$. Then the following contractive limits exist as $\epsilon \to 0$:

$$\begin{aligned} &\epsilon(x_1^\epsilon + y_1^\epsilon)/2 \to \lambda\tilde{e}_1, \quad (x_1^\epsilon - y_1^\epsilon)/2 \to \lambda e_4,\\ &\quad e_2, e_3 \to e_2, e_3, \end{aligned} \tag{2.7}$$

where $Q^0(\tilde{e}_1) = 0$, $Q^0(e_i) = -1$. Moreover, $V^\epsilon \to V^0 = (R^{1,3,0}, Q^0)$ as $\epsilon \to 0$ and a Witt decomposition of V^0 is provided by the orthogonal sum $\ker Q^0 \perp G^0$,

$$\ker Q^0 = \mathrm{span}_R\{\tilde{e}_1\}, \text{ and } G^0 = \mathrm{span}_R\{e_2,e_3,e_4\}. \tag{2.8}$$

Example 2.9: Consider now the complexification V^c of $(R^{3,1}, Q)$ from example 2.8. Then the complexified quadratic

form Q^c is neutral (that is, of maximum index 2) and $V^c = \mathrm{F} \oplus \mathrm{F}'$,

$$F = \mathrm{span}_C\{x_1 = \lambda(e_1 + e_4), x_2 = \lambda(ie_2 + e_3)\},$$
$$F' = \mathrm{span}_C\{y_1 = \lambda(e_1 - e_4), y_2 = \lambda(ie_2 - e_3)\}. \tag{2.9}$$

Applying now a (0,1)-deformation of V^c we get $V^{\epsilon c} = F^\epsilon \oplus F'^\epsilon$, where

$$F^\epsilon = \mathrm{span}_C\{x_1^\epsilon = \lambda(e_1/\epsilon + e_4),\ x_2^\epsilon = \lambda(ie_2 + e_3)\},$$
$$F'^\epsilon = \mathrm{span}_C\{y_1^\epsilon = \lambda(e_1/\epsilon - e_4),\ y_2^\epsilon = \lambda(ie_2 - e_3)\}, \tag{2.10}$$

and $Q^{\epsilon c}(e_1) = \epsilon^2$, $Q^{\epsilon c}(e_i) = -1$, $i = 2,3,4$. Then the following contractive limits exist as $\epsilon \to 0$:

$$\epsilon(x_1^\epsilon + y_1^\epsilon)/2 \to \lambda\tilde{e}_1, \quad (x_1^\epsilon - y_1^\epsilon)/2 \to \lambda e_4, \quad x_2^\epsilon,\ y_2^\epsilon \to x_2,\ y_2, \tag{2.11}$$

where $Q^{0c}(\tilde{e}_1) = 0$ and $Q^{0c}(e_i) = -1$. Moreover, $V^{\epsilon c}$ contracts to V^{0c}, the complexification of $(R^{1,3,0}, Q^0)$ and a Witt decomposition of V^{0c} is provided by the orthogonal sum $\ker Q^{0c} \perp (\tilde{F} \oplus \tilde{F}') \perp G^0$,

$$\ker Q^{0c} = \mathrm{span}_C\{\tilde{e}_1\}, \quad G_0 = \mathrm{span}_C\{e_4\},$$
$$\tilde{F} = \mathrm{span}_C\{x_2 = \lambda(ie_2 + e_3)\}, \tag{2.12}$$
$$\tilde{F}' = \mathrm{span}_C\{y_2 = \lambda(ie_2 - e_3)\}.$$

Notice that $(\tilde{F} \oplus \tilde{F}') \perp G^0$ is a Witt decomposition of the complexification of $R^{3,0}$ spanned by $\{e_2, e_3, e_4\}$ and considered as a nondegenerate part of V^{0c}.

III. CONTRACTION OF SPIN REPRESENTATIONS

In this section we discuss spin representations of a family of deformed complex Clifford algebras $\{C(Q^\epsilon)\}$ parameterized by the deformation parameter ϵ, $0 < \epsilon \leqslant 1$. It is useful to continue the coordinate dependent approach developed in Sec. II to study the contractions of spin bases induced by the contractions of the associated Witt bases. We do not consider here any global properties of such contractive limits in the way it is presented in Ref. 17 for space-times depending on a free parameter.

The theory of spin representations of nondegenerate Clifford algebras can be found in Refs. 1, 2, 4, and 7. Recall that a *spin representation* ρ is defined as the regular representation in a minimal left (or right) ideal S called *spin ideal* (or *spin space*). The latter can then be considered as a Clifford module and, in the following, this approach will be extended to degenerate Clifford algebras. Since $R_{p,k}$ is a finite-dimensional semisimple algebra, there exists a primitive idempotent e so that $S = R_{p,k}e$ (see Refs. 5, 6, and 18). The construction of e in terms of an orthogonal basis of $R^{p,k}$ possessing an isotropic subspace of maximal index is due to Cartan and Chevalley and its generalizations can be found in Refs. 7 and 8. However, we utilize the Witt basis for the complexification of $R^{\epsilon p,k}$ to construct a nilpotent generator f^ϵ of S^ϵ in $C(Q^\epsilon)$ (see Refs. 2 and 4 when $\epsilon = 1$).

It is well known that Pauli and Dirac spinors can be considered abstractly as elements of the spin ideals of $R^c_{3,0}$ and $R^c_{3,1}$, respectively. In particular, the spin ideal of the latter provides also a representation space for the covering group of the 15-parameter group of conformal transformations in the Minkowski space-time (see Ref. 19 for the relation between the conformal Lie algebra and the Clifford algebra), in which case it is then called *twistor space*.[20,21] The relation between the regular action of the covering group on twistors and induced projected conformal action on the Minkowski space-time can be found in Ref. 22.

In the following, we write V^ϵ [resp. $C(Q^\epsilon)$] for the complexification of $R^{\epsilon p,k}$ (resp. $R^\epsilon_{p,k}$). First, let the dimension n of V^ϵ be even, $n = p + k = 2r$. Then $S^\epsilon = C(Q^\epsilon)\, f^\epsilon$ is a family of deformed spin ideals, where $f^\epsilon = y_1^\epsilon \cdots y_r^\epsilon$ is the volume element in the exterior algebra over F'^ϵ in the Witt decomposition of V^ϵ (see example 2.9 when $n = 4$). Each S^ϵ provides a representation space for the spin representation ρ^ϵ of $C(Q^\epsilon)$ defined as the left regular action.

When n is odd, the spin space is defined as a left $C^+(Q^\epsilon)$ module of the even subalgebra of $C(Q^\epsilon)$ with the generator f^ϵ factored as above. It is shown in Ref. 2 that there are only two inequivalent ways to extend the spin representation of the even subalgebra to irreducible representations of the entire algebra.

Let $\{x_i^\epsilon, y_i^\epsilon\}$, $i = 1,\ldots,r$ be the deformed Witt basis in V^ϵ. Then $C(Q^\epsilon)$ is spanned by the monomials $\{x_J^\epsilon y_K^\epsilon\}$, $0 \leqslant |J| \leqslant r$, where J and K are multi-indices, $x_J^\epsilon = x_{j_1}^\epsilon \cdots x_{j_s}^\epsilon$, $|J| = s$, $y_K^\epsilon = y_{k_1}^\epsilon \cdots y_{k_t}^\epsilon$, $|K| = t$, $x_\varnothing = y_\varnothing = 1$. Since f^ϵ is a multivector of length r and all its factors are isotropic, the spin space S^ϵ is spanned by 2^r linearly independent elements $\{x_J^\epsilon f^\epsilon\}$. It can also be shown that when n is odd the spin space $C^+(Q^\epsilon) f^\epsilon$ again has dimension 2^r (see Ref. 2).

Lemma 3.1: The dimension of S^ϵ is 2^r and the set $\{x_J^\epsilon f^\epsilon\}$, $0 \leqslant |J| \leqslant r$, provides a basis for S^ϵ, $r = [n/2]$.

Using (2.4) we can find a contractive limit of every generator of the spin space $\epsilon^{z(J)} x_J^\epsilon f^\epsilon \to \tilde{x}_J F$ as $\epsilon \to 0$, where $z(J)$ is the smallest positive integer depending on J such that a nonzero limit exists. For example, $\epsilon^z f^\epsilon \to F$, $z = d/2$ (resp. $[d/2] + 1$) when d is even (resp. odd) for some nilpotent element F in the radical of $C(Q^0)$ ($d = \dim V^{0\perp}$).

Definition 3.2: The ideal $C(Q^0)F$ is called the *limit spin ideal* and the limit regular representation ρ^0 of $C(Q^0)$ in $C(Q^0)F$ is called the *limit spin representation*.

It should be noted, however, that the limit spin ideals defined above cease to be minimal and are, in general, decomposable.

In the following example it is demonstrated that the spin ideal of the Galilei–Clifford algebra can be viewed as the limit of twistor spaces.

Example 3.3: Let V be the quadratic space from example 2.9. Then the Witt basis of the (0,1)-deformation V^ϵ was given in (2.10). Thus $f^\epsilon = y_1^\epsilon y_2^\epsilon$ and $T^\epsilon = R^{\epsilon c}_{3,1} f^\epsilon$ can be considered as a left Clifford module.[21,22] The standard basis in T^ϵ is given by $\{f^\epsilon, x_1^\epsilon f^\epsilon, x_2^\epsilon f^\epsilon, x_1^\epsilon x_2^\epsilon f^\epsilon\}$ and we have the following limits as $\epsilon \to 0$:

$$\epsilon f^\epsilon \to F = \lambda^2 \tilde{e}_1(ie_2 - e_3), \quad \epsilon x_1^\epsilon f^\epsilon \to 2\lambda e_4 F,$$
$$\epsilon x_2^\epsilon f^\epsilon \to x_2 F, \quad \epsilon x_1^\epsilon x_2^\epsilon f^\epsilon \to 2\lambda e_4 x_2 F, \tag{3.1}$$

where $\{\tilde{e}_1, e_2, e_3, e_4\}$ is an orthogonal basis in the complexified

Galilei space-time V_0 endowed with the degenerate quadratic form Q^0, $Q^0(\tilde{e}_1)=0$, $Q^0(e_i)=-1$, $i=2,3,4$. We then have $R^{\epsilon c}_{1,3}\to R^{c}_{1,3,0}$ as $\epsilon\to 0$. Let $T^0=\mathrm{span}_C\{F, e_4F, x_2F, e_4x_2F\}$, and we may write formally $T^\epsilon\to T^0$. Since every element of the radical of $R^{c}_{1,3,0}$ (of the form $m\tilde{e}_1$ with m in $R^{c}_{3,0}$) annihilates F, $T^0=R^{c}_{1,3,0}F=R^{c}_{3,0}F$, where $R^{c}_{3,0}$ is the Clifford subalgebra generated by e_2, e_3, and e_4. If ρ^ϵ denotes the spin representation of $R^{\epsilon c}_{1,3}$ in T^ϵ, then $\rho^\epsilon\to\rho^0$, where ρ^0 is the spin representation of the Galilei–Clifford algebra.

Before we summarize this section let $C(Q^\epsilon)$ [resp. $C(Q^0)$, $C(Q')$] denote $R^{\epsilon c}_{3,1}$ (resp. $R^{c}_{1,3,0}$, $R^{c}_{3,0}$). Also, let f_i, $i=1,2,3,4$, be mutually annihilating primitive idempotents in $C(Q')$ [hence in $C(Q^0)$ by the Lifting Idempotents Theorem of Ref. 18]. For example,

$$f_1=\tfrac{1}{4}(1+ie_{23})(1+ie_4),\quad f_2=\tfrac{1}{4}(1+ie_{23})(1-ie_4),$$
$$f_3=\tfrac{1}{4}(1-ie_{23})(1+ie_4),\quad f_4=\tfrac{1}{4}(1-ie_{23})(1-ie_4),$$

where $e_{23}=e_2e_3$ (see Ref. 23). Then $C(Q')$ has a direct sum decomposition, which can be lifted to $C(Q^0)$, into minimal (hence indecomposable) left ideals $C(Q')f_i$. Thus $T^0=C(Q^0)F=C(Q')F$ can be written as a direct sum of $C(Q^0)$-modules $C(Q^0)f_iF$. In fact, we can show that two of these modules are indecomposable, whereas the remaining two are trivial.

Proposition 3.4: Let $T^\epsilon=C(Q^\epsilon)f^\epsilon$ be the twistor space and let ρ^ϵ be the irreducible spin representation of $C(Q^\epsilon)$ in T^ϵ, $0<\epsilon\leqslant 1$.

(i) There exist limits $T^\epsilon\to T^0$, $\rho^\epsilon\to\rho^0$, where $T^0=C(Q^0)F$ is a left regular representation space of $C(Q^0)$, $\epsilon f^\epsilon\to F$. The representation ρ^0 is unfaithful.

(ii) ρ^0 is decomposable into a direct sum of two indecomposable subrepresentations.

(iii) T^0 is a complex four-dimensional vector space spanned by the contractive limits $\{F, e_4F, x_2F, e_4x_2F\}$ of the canonical basis in T^ϵ.

IV. SPIN SPACE AS A CLIFFORD MODULE

In this section we examine the module structure of the spin ideal of a degenerate Clifford algebra. Let V^ϵ denote the complexification of a deformed real quadratic space $R^{p,k}$ of even dimension $n=p+k=2r$ (the case when dimension is odd will be considered later) and let $\{x_i^\epsilon, y_i^\epsilon\}$, $i=1,\ldots,r$ be a deformed Witt basis in V^ϵ. If, for example, we want to consider a contraction of the Clifford algebra $C(Q^\epsilon)$ over V^ϵ to the Clifford algebra $C(Q^0)$ over V^0, dim $V^{0\perp}=d=2q$, we may modify (2.3) as follows

$$\{x_1^\epsilon=\lambda(e_1+e_n)/\epsilon,\ldots,x_q^\epsilon=\lambda(e_q+e_{n-q+1})/\epsilon,$$
$$x_{q+1}^\epsilon=\lambda(e_{q+1}+e_{n-q}),\ldots,x_r^\epsilon=\lambda(e_r+e_{r+1})\},\qquad(4.1)$$

and similarly for the y^ϵ part. Then the contractive limits can be found as in (2.4). In particular, $\epsilon^q f^\epsilon\to F=\tilde{y}_1\cdots\tilde{y}_q y_{q+1}\cdots y_r$ is the limit of the generator f^ϵ of the spin ideal S^ϵ of $C(Q^\epsilon)$, where $\tilde{y}_i$ are isotropic vectors in the radical of $C(Q^0)$. Since a typical generator of $C(Q^0)$ has the form $x_K\tilde{x}_Jy_M\tilde{y}_L$, $0\leqslant|J|,|L|\leqslant q$, $0\leqslant|M|\leqslant r-q$, the limit spin space $S^0=C(Q^0)F$ is spanned by 2^r linearly independent monomials $\{x_K\tilde{x}_JF\}$.

When the dimension of $V^{0\perp}$ is odd, $d=2q+1$, vector x_{q+1}^ϵ in (4.1) is replaced with $x_{q+1}^\epsilon=\lambda(e_{q+1}/\epsilon+e_{n-q})$ (and similarly for y_{q+1}^ϵ), while the other basis vectors remain unchanged. Then,

$$\epsilon^{q+1}f^\epsilon\to F=\tilde{y}_1\cdots\tilde{y}_q\tilde{e}_{q+1}y_{q+2}\cdots y_r,$$

and a basis for S^0 is again formed by 2^r elements $\{x_K\tilde{x}_JF\}$ and $\{x_Ke_{n-q}\tilde{x}_JF\}$, $0\leqslant|K|\leqslant r-q-1$. Notice that the monomials $\{x_Ky_M\}$, $0\leqslant|K|$, $|M|\leqslant r-q$ (resp. $\{x_Ky_M\}$ and $\{x_Ke_{n-q}y_M\}$, $0\leqslant|K|$, $|M|\leqslant r-q-1$) provide a canonical basis for a maximal nondegenerate Clifford subalgebra $C(Q')$ of $C(Q^0)$ when d is even (resp. odd).

Proposition 4.1: Let $S^0=C(Q^0)F$ be the spin space of a complex degenerate Clifford algebra $C(Q^0)$ over V^0, dim $V^0=2r$ considered as the contractive limit of spin spaces $S^\epsilon=C(Q^\epsilon)f^\epsilon$ and let dim $V^{0\perp}=d$, $q=[d/2]$.

(i) The dimension of S^0 is 2^r.

(ii) A basis in S^0 is provided by the monomials $\{x_K\tilde{x}_JF\}$ (resp. $\{x_K\tilde{x}_JF\}$ and $\{x_Ke_{n-q}\tilde{x}_JF\}$) when d is even (resp. odd), $0\leqslant|J|\leqslant q$, $0\leqslant|K|\leqslant r-q$.

(iii) S^0 is a direct sum of 2^q left simple $C(Q')$-modules $M_J=C(Q')\tilde{x}_JF$, $0\leqslant|J|\leqslant q$.

(iv) The decomposition in (iii) is not $C(Q^0)$ invariant.

Proof: (iii) We only need to prove that the direct summands M_J are simple $C(Q')$-modules. It is enough to notice that the generator F of S^0 contains a factor $f_1=y_{q+1}\cdots y_r$ (resp. $f_1=y_{q+2}\cdots y_r$) when d is even (resp. odd). In any case, f_1 is an $(r-q)$ vector generating a spin ideal $C(Q')f_1$ in $C(Q')$. Following the arguments contained in Ref. 2 and concerning the irreducibility of spin representations of $C(Q')$ in even and odd dimensions, we conclude that M_J are simple $C(Q')$-modules. (iv) Notice that $M_{J,|J|}=q$, is the only $C(Q^0)$-invariant summand in the decomposition of S^0. □

Corollary 4.2:

(i) The dimension of M_J is 2^{r-q}.

(ii) The spin representation ρ^0 of $C(Q^0)$ in S^0, when restricted to the nondegenerate Clifford subalgebra $C(Q')$, is decomposable into 2^q irreducible 2^{r-q}-dimensional subrepresentations.

(iii) The further restriction ρ^0_+ of ρ^0 from part (ii) to the even subalgebra $C^+(Q')$ is decomposable into 2^{q+1} irreducible 2^{r-q-1}-dimensional subrepresentations when d is even. When d is odd, part (ii) applies also to ρ^0_+.

Proof: See Refs. 1, 2, and 4 for the structure theorems on Clifford algebras. □

Finally, we briefly consider the case when the dimension n of V^ϵ is odd, $n=2r+1$. Let $\{x_i^\epsilon, y_i^\epsilon, e_n\}$, $i=1,\ldots,r$ be a deformed Witt basis in V^ϵ. It is assumed first that the vector e_n spanning one-dimensional anisotropic component of V^ϵ (see Theorem 2.1) is not deformed. We can make use of (4.1) to represent the isotropic vectors in the Witt basis.

Proposition 4.3: Let $S^0=C^+(Q^0)F$ be the spin space of a complex degenerate Clifford algebra $C(Q^0)$ over V^0, dim $V^0=2r+1$, considered as the contractive limit of spin spaces $S^\epsilon=C^+(Q^\epsilon)f^\epsilon$ and let dim $V^{0\perp}=d$, $q=[d/2]$.

(i) The dimension of S^0 is 2^r.

(ii) A basis in S^0 is provided by the monomials $\{x_Kz_N\tilde{x}_JF\}$, $|J|+|K|+|N|$ even, when d is even, and by

$\{e_L x_K z_N \tilde{x}_J F\}$, $|L| + |J| + |K| + |N|$ even, where d is odd, $0 \leqslant |N|$, $|L| \leqslant 1$, $0 \leqslant |J| \leqslant q$, $0 \leqslant |K| \leqslant r - q$ ($z_N = e_n$, when $|N| = 1$ and $e_L = e_{2r-q}$, when $|L| = 1$).

(iii) S^0 is a direct sum of 2^q left simple $C^+(Q')$ modules $M_J^+ = C^+(Q')\tilde{x}_J F, |J|$ even, and $M_J^- = C^-(Q')\tilde{x}_J F, |J|$ odd, $0 \leqslant |J| \leqslant q$ [here $C^-(Q')$ denotes the odd part of $C(Q')$].

Proof: Follow the arguments presented above and in Ref. 2. □

Notice that in the case of maximum degeneracy the contractive limit of $C(Q^\epsilon)$ is the exterior algebra Λ over $V^0 = V^{01}$, dim $V^0 = n$. The ideal ΛF is then spanned by 2^r linearly independent elements $\{\tilde{x}_J F\}$, $0 \leqslant |J| \leqslant r$, $r = [n/2]$. However, ΛF is not minimal since it contains a one-dimensional space of n-vectors.

V. SUMMARY

The method of deformation and contraction in Clifford algebras developed in this paper was primarily applied to construct spin ideals in degenerate Clifford algebras of any rank. Through the classical approach of the Witt decomposition to irreducible spin spaces, the spin bases were deformed and contracted to obtain a spin basis for the limit spin space. Considering the latter as a left Clifford module over a degenerate Clifford algebra, it was later decomposed into indecomposable components. The limit spin ideals, which provide natural representation spaces for Clifford, Pin, and Spin groups associated with degenerate Clifford algebras, are further studied in Ref. 24.

[1] N. Bourbaki, *Elements de Mathématique* (Hermann, Paris, 1967), Livre II.
[2] A. Crumeyrolle, *Algèbres de Clifford et Spineurs*, Cours de Seminaires du Department de Mathématiques de l'Université de Toulouse (Université Paul Sabatier, Toulouse, 1974).
[3] O. Laborde, Bull. Sci. Math. **96**, 199 (1972).
[4] J. Komorowski, *Od Liczb Zespolonych do Tensorów, Spinorów, Algebr Liego i Kwadryk* (PWN–Polish Scientific, Warsaw, 1978).
[5] I. N. Herstein, *Noncommutative Rings* (Mathematical Association of America, Chicago, 1968).
[6] J. Lambek, *Lectures on Rings and Modules* (Blaisdell, Waltham, 1966).
[7] P. Lounesto, Ann. Inst. H. Poincaré, A **33**, 53 (1980).
[8] P. Budinich and K. Bugajska, J. Math. Phys. **26**, 588 (1985).
[9] J. A. Brooke, J. Math. Phys. **19**, 952 (1978), **21**, 617 (1980).
[10] A. Crumeyrolle, Ann. Inst. H. Poincaré, A **33**, 235 (1980).
[11] N. Salingaros, J. Math. Phys. **22**, 226 (1981), **23**, 1231 (E) (1982).
[12] R. T. Jantzen, J. Math. Phys. **23**, 1741 (1982).
[13] K. R. Greider, Found. Phys. **14**, 467 (1984).
[14] Z. Hasiewicz, A. K. Kwasniewski, and P. Morawiec, J. Math. Phys. **25**, 2031 (1984).
[15] T. Y. Lam, *The Algebraic Theory of Quadratic Forms* (Benjamin, Reading, MA, 1980).
[16] R. Cecchini, Nuovo Cimento A **43**, 181 (1978); H. P. Kunzle, Ann. Inst. H. Poincaré A **17**, 337 (1972); J. M. Lévy-Leblond, Commun. Math. Phys. **6**, 286 (1967).
[17] F. Bampi and R. Cianci, Int. J. Theor. Phys. **19**, 211 (1980).
[18] Ch. W. Curtis and I. Reiner, *Methods of Representation Theory with Applications to Finite Groups and Orders* Vol. I. (Wiley Interscience, New York, 1981).
[19] P. Lounesto and E. Latvamaa, Proc. Am. Math. Soc. **79**, 533 (1980).
[20] R. Penrose and M. MacCallum, Phys. Rep. C **6**, 243 (1972).
[21] R. Ablamowicz and N. Salingaros, Lett. Math. Phys. **9**, 149 (1985).
[22] R. Ablamowicz, Z. Oziewicz, and J. Rzewuski, J. Math. Phys. **23**, 231 (1982).
[23] P. Lounesto, "*On primitive idempotents of Clifford algebras*," Report HTKK-MAT-A113, 1977.
[24] R. Ablamowicz, "Structure of spin groups associated with degenerate Clifford algebras," J. Math. Phys. **27**, 1 (1986).

Coherent states of the real symplectic group in a complex analytic parametrization. I. Unitary-operator coherent states

C. Quesne[a)]
Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Bd. du Triomphe, B 1050 Brussels, Belgium

(Received 28 May 1985; accepted for publication 19 September 1985)

In the present series of papers, the coherent states of $Sp(2d,R)$, corresponding to the positive discrete series irreducible representations $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$ encountered in physical applications, are analyzed in detail with special emphasis on those of $Sp(4,R)$ and $Sp(6,R)$. The present paper discusses the unitary-operator coherent states, as defined by Klauder, Perelomov, and Gilmore. These states are parametrized by the points of the coset space $Sp(2d,R)/H$, where H is the stability group of the $Sp(2d,R)$ irreducible representation lowest weight state, chosen as the reference state, and depends upon the relative values of $\lambda_1,...,\lambda_d$, subject to the conditions $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d \geqslant 0$. A parametrization of $Sp(2d,R)/H$ corresponding to a factorization of the latter into a product of coset spaces $Sp(2d,R)/U(d)$ and $U(d)/H$ is chosen. The overlap of two coherent states is calculated, the action of the $Sp(2d,R)$ generators on the coherent states is determined, and the explicit form of the unity resolution relation satisfied by the coherent states in the representation space of the irreducible representation is obtained. The Hilbert space of analytic functions arising from the coherent state representation is studied in detail. Finally, some applications of the formalism developed in the present paper are outlined. In particular, its relevance to the study of boson realizations of the $Sp(2d,R)$ algebra is stressed.

I. INTRODUCTION

The harmonic oscillator coherent states (CS), also referred to in the literature as Glauber's standard CS, or the CS associated with the Heisenberg–Weyl group $N(1)$,[1,2] are known to be endowed with a host of properties making the CS suitable to various interesting applications (for a recent review on the standard CS and their extensions see Ref. 3). They may be defined in many different, but essentially equivalent ways. We shall mention herein only three of them: (i) as unitary-operator CS, they are obtained by applying a unitary transformation to the oscillator ground state; (ii) as annihilation-operator CS, they are the eigenstates of the oscillator annihilation operator, corresponding to complex eigenvalues; and (iii) as minimum-uncertainty CS, they form the set of states minimizing the position-momentum uncertainty relation subject to the restriction that the oscillator ground state be in the set.

In recent years, many works have been devoted to the generalization of the standard CS (see Ref. 3). In extending the notion of CS from the harmonic oscillator to other physical systems, the various CS definitions, which were equivalent for the former system, lose their equivalency for the latter. Consequently there are many different generalized CS for a given system.

In the present series of papers, we shall be concerned with the group theoretical generalizations of both the unitary-operator and the annihilation-operator CS, respectively proposed by Klauder,[4] Perelomov,[5] and Gilmore[6] and by Barut and Girardello.[7] In both extensions, the CS are associated with a given Lie group, assumed to be a dynamical group of the considered physical system. There exists, however, an essential distinction between them as regards the Lie groups which may be treated: if the Klauder–Perelomov–Gilmore generalization of CS is valid for any Lie group, either compact or noncompact, that of Barut and Girardello can only be applied to noncompact Lie groups, and was actually developed by these authors only for $SO(2,1)$, and its locally isomorphic groups $SU(1,1)$, $Sl(2,R)$, and $Sp(2,R)$.

Up to now, most efforts have concentrated on the CS associated with compact Lie groups, for which all unitary irreducible representations (irreps) are finite dimensional.[8] Little work has been devoted to the CS associated with noncompact Lie groups, whose unitary irreps are all infinite dimensional, except for the trivial identity representation. Among noncompact groups, however, the real symplectic group in $2d$ dimensions, $Sp(2d,R)$, plays an outstanding role in many physical problems. Let us mention, for instance, the relevance of $Sp(4,R)$ to a class of generalized helium Hamiltonians, as shown by Mlodinow and Papanicolaou,[9] and the importance of $Sp(6,R)$ in the microscopic nuclear collective model, as pointed out by Rosensteel and Rowe,[10] and studied by various authors from different viewpoints.[11–28] The $Sp(2d,R)$ irreps encountered in all such physical applications are positive discrete series,[29,30] characterized by their lowest weight $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$, where $[\lambda_1 \cdots \lambda_d]$ is some partition, and n is an integer greater than or equal to $2d$.

The $Sp(2,R)$ CS were derived by Perelomov[5] and Gilmore[6] on one hand, and by Barut and Girardello[7] on the other hand. For higher-dimensional symplectic groups, a full analysis of the CS was only carried out for the irreps $\langle (\lambda + n/2)^d\rangle$, which in the nuclear collective model (wherein $d = 3$) are relevant to closed-shell nuclei. For such irreps, the unitary-operator CS were studied by Kramer,[21] while the annihilation-operator CS were determined by Deenen

[a)] Maître de recherches F.N.R.S.

 0022-2488/86/020428-14$02.50 © 1986 American Institute of Physics

and Quesne.[25] The Sp(2d,R) irreps $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$, for which the λ_i's are not all equal, play, however, an important part in the description of open-shell nuclei in the nuclear collective model. For such irreps, some partial results for the unitary-operator CS were recently obtained by Rowe,[15] who also studied their relation with boson realizations of Sp(2d,R). By the same time, Deenen and Quesne had given a full analysis of partially coherent states (PCS) for the same irreps.[27] Contrary to the CS, which are specified by some continuous parameters, the PCS are characterized by a set of continuous labels as well as by some discrete indices. Both PCS generalizing either the unitary-operator or the annihilation-operator CS were studied, and then used to analyze the properties of the Sp(2d,R) boson realizations.[27,28]

Since the CS corresponding to the Sp(2d,R) irreps $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$, for which $\lambda_1,...,\lambda_d$ are not all equal, are of considerable interest for physical applications,[14,16,22] it is the purpose of the present series of two papers (henceforth referred to as I and II) to analyze them in full detail with special emphasis on those of Sp(4,R) and Sp(6,R). Paper I deals with the unitary-operator CS, while paper II will be devoted to the annihilation-operator CS. If a complex parametrization necessarily arises for the latter, for the former we may choose between a real or complex one, both of them having some respective advantages. Real parameters can be more easily given a physical meaning, whereas complex parameters are useful to connect CS representations to boson realizations. A complex parametrization has been chosen in the present paper because it is easier to deal with than a real one.

In Sec. II of this paper, the positive discrete series irreps of Sp(2d,R) are briefly reviewed. The corresponding CS are defined in Sec. III, and their overlap is calculated in Sec. IV. The action of the Sp(2d,R) generators on the CS is determined in Sec. V, and used in Sec. VI to find the explicit form of the unity resolution satisfied by the set of CS in the representation space of the irrep $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$. Finally, in Sec. VIII, some applications of the formalism developed in the previous sections are discussed.

After completion of the present work, it came to the knowledge of the author that some results similar to those contained in this paper were independently obtained by Kramer.[31]

II. THE Sp(2d,R) POSITIVE DISCRETE SERIES IRREPS

As usual, let us realize the Sp(2d,R) generators in terms of dn boson creation operators η_{is}, $i = 1,...,d$, $s = 1,...,n$, and their corresponding annihilation operators $\xi_{is} = (\eta_{is})^\dagger$, as follows[32]:

$$\begin{aligned} D^\dagger_{ij} &= D^\dagger_{ji} = \sum_{s=1}^{n} \eta_{is}\eta_{js}, \quad 1\leqslant i\leqslant j\leqslant d, \\ D_{ij} &= D_{ji} = \sum_{s=1}^{n} \xi_{is}\xi_{js}, \quad 1\leqslant i\leqslant j\leqslant d, \\ E_{ij} &= \frac{1}{2}\sum_{s=1}^{n} (\eta_{is}\xi_{js} + \xi_{js}\eta_{is}) \\ &= \sum_{s=1}^{n} \eta_{is}\xi_{js} + \frac{n}{2}\delta_{ij}, \quad i,j = 1,...,d. \end{aligned} \tag{2.1}$$

They satisfy the Hermiticity properties

$$E_{ij} = (E_{ji})^\dagger, \quad D_{ij} = (D^\dagger_{ij})^\dagger, \tag{2.2}$$

and the commutation relations

$$\begin{aligned} &[E_{ij},E_{kl}] = \delta_{jk}E_{il} - \delta_{il}E_{kj}, \\ &[E_{ij},D^\dagger_{kl}] = \delta_{jk}D^\dagger_{il} + \delta_{jl}D^\dagger_{ik}, \\ &[E_{ij},D_{kl}] = -\delta_{ik}D_{jl} - \delta_{il}D_{jk}, \\ &[D^\dagger_{ij},D^\dagger_{kl}] = [D_{ij},D_{kl}] = 0, \\ &[D_{ij},D^\dagger_{kl}] = \delta_{ik}E_{lj} + \delta_{il}E_{kj} + \delta_{jk}E_{li} + \delta_{jl}E_{ki}, \end{aligned} \tag{2.3}$$

from which we note that the operators E_{ij} generate the maximal compact subgroup U(d) of Sp(2d,R). The set of generators (2.1) can be divided into three subsets of raising, weight, and lowering type, as follows:

$$D^\dagger_{ij}, E_{ij} \quad (i<j); \quad E_{ii}; \quad D_{ij}, E_{ij} \quad (i>j), \tag{2.4}$$

where the subsets are separated by semicolons.

In the above equations, n is an arbitrary positive integer. If we now restrict ourselves to n values greater than or equal to 2d, we can realize all the Sp(2d,R) positive discrete series irreps $\langle \lambda \rangle \equiv \langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$ in the space of boson states built from the dn boson creation operators η_{is}. The lowest weight state $|(\lambda)_{\min}\rangle$ of such an irrep satisfies the following equations[33]:

$$D_{ij}|(\lambda)_{\min}\rangle = 0, \quad i\leqslant j, \tag{2.5a}$$

$$E_{ij}|(\lambda)_{\min}\rangle = 0, \quad i>j, \tag{2.5b}$$

$$E_{ii}|(\lambda)_{\min}\rangle = (\lambda_{d+1-i} + n/2)|(\lambda)_{\min}\rangle, \quad i = 1,...,d. \tag{2.5c}$$

From Eqs. (2.5b) and (2.5c), we note that it is the lowest weight state of an irrep $[\lambda] \equiv [\lambda_1 + n/2,...,\lambda_d + n/2]$ of U(d), so it can be characterized by the corresponding Gel'fand pattern $(\lambda)_{\min}$ (see Ref. 34).

The whole representation space $\mathcal{F}_{\langle\lambda\rangle}$ of the irrep $\langle\lambda\rangle$ can be generated from its lowest weight state $|(\lambda)_{\min}\rangle$ by applying polynomials in the $D^\dagger_{ij}$, E_{ij}, and D_{ij} generators, written in normal form as $P(D^\dagger_{ij})P'(E_{ij})P''(D_{ij})$. From Eq. (2.5a), $P''(D_{ij})$ applied to $|(\lambda)_{\min}\rangle$ gives rise to the same state multiplied by an irrelevant constant. The action of all the polynomials $P'(E_{ij})$ upon $|(\lambda)_{\min}\rangle$ generates the representation space of the irrep $[\lambda]$, whose basis states $|(\lambda)\rangle$ can be characterized by the Gel'fand patterns (λ). A discrete nonorthonormal basis of $\mathcal{F}_{\langle\lambda\rangle}$ is therefore given by[27]

$$|\mathbf{N};(\lambda)\rangle = F_{\mathbf{N}}(\mathbf{D}^\dagger)|(\lambda)\rangle, \tag{2.6}$$

where $\mathbf{D}^\dagger$ denotes the $d\times d$ matrix $\|D^\dagger_{ij}\|$, $F_{\mathbf{N}}(\mathbf{D}^\dagger)$ is defined by

$$F_{\mathbf{N}}(\mathbf{D}^\dagger) = \prod_{i<j} (N_{ij}!)^{-1/2}[(1+\delta_{ij})^{-1/2}D^\dagger_{ij}]^{N_{ij}}, \tag{2.7}$$

the quantum numbers N_{ij}, $1\leqslant i\leqslant j\leqslant d$, run over all non-negative integers, and (λ) over all Gel'fand patterns of $[\lambda]$.

In physical applications, it is convenient to use bases classified according to the chain of groups

$$\begin{array}{ccccc} \mathrm{Sp}(2d,R) & \supset & \mathrm{U}(d) & \supset & \mathrm{SO}(d). \\ \langle\lambda\rangle & & [h] & & (k) \end{array} \tag{2.8}$$

Underneath each one of them, we have indicated the label

characterizing its irrep. Such bases can be obtained by coupling the polynomials in the $D^{\dagger}_{ij}$ generators specified by a definite U(d) irrep $[l]$ to the basis states of the U(d) irrep $[\lambda]$, both classified according to the chain (2.8),

$$|([l][\lambda])\alpha[h]\beta(k)q\rangle = [P_{[l]}(\mathbf{D}^{\dagger})\times|(\ )\rangle]^{\alpha[h]}_{\beta(k)q}. \tag{2.9}$$

Here α distinguishes between repeated irreps $[h]$ in the reduction of the product representation $[l]\times[\lambda]$, β does the same for the repeated irreps (k) in the reduction of $[h]$, and q characterizes the row of the SO(d) irrep (k). If the states (2.9) are orthogonal with respect to $[h]$,(k),q, they are not with respect to $[l]$,α, β.

Having reviewed the Sp(2d,R) positive discrete series irreps and some discrete basis of their representation space $\mathscr{F}_{(\lambda)}$, we shall proceed in the next section to define the corresponding unitary-operator CS, which in the subsequent sections will be shown to form a continuous basis of $\mathscr{F}_{(\lambda)}$.

III. PARAMETRIZATION OF THE Sp(2d,R) COHERENT STATES

Following Klauder,[4] Perelomov,[5] and Gilmore,[6] the system of unitary-operator CS corresponding to a unitary irrep ρ of a given Lie group G in a Hilbert space $\mathscr{F}_{\rho}$, and to a fixed vector $|\psi_0\rangle$ of this space, is the set of states $g|\psi_0\rangle$, where g runs over all the group G. If H denotes the stability subgroup of $|\psi_0\rangle$, i.e., the set of all group elements h leaving $|\psi_0\rangle$ unchanged up to a phase factor, then every group element $g\in G$ has a unique decomposition into a product of two group elements, one of which is in H and the other in the coset space G/H,

$$g = kh, \quad h\in H, \quad k\in G/H. \tag{3.1}$$

Since phase factors are irrelevant, the CS can be parametrized by the points k of the coset space G/H,

$$|k\rangle = k\,|\psi_0\rangle. \tag{3.2}$$

In practice, the decomposition (3.1) is carried out by going to the complex extension of the real Lie group G (see Refs. 35 and 36).

In the present case, the Lie group G is the real symplectic group Sp(2d,R), its unitary irrep is specified by $\rho = \langle\lambda\rangle$, and the corresponding representation space is the space $\mathscr{F}_{(\lambda)}$, spanned by the states (2.6) or (2.9). For the reference state $|\psi_0\rangle$, we choose the irrep lowest weight state $|(\lambda)_{\min}\rangle$, characterized by Eq. (2.5). The stability subgroup H of this state depends upon the respective values of $\lambda_1,\lambda_2,\ldots,\lambda_d$, satisfying the conditions $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_d\geqslant 0$.

The case where $\lambda_1=\lambda_2=\cdots=\lambda_d=\lambda$ has been extensively studied in the literature.[21,25,37] The irrep $[\lambda]$ is then one dimensional, and its single base $|(\lambda)_{\min}\rangle$. The stability group H is the maximal compact subgroup U(d) of Sp(2d,R), hence the corresponding CS exist in one-to-one correspondence with the points of the coset space Sp(2d,R)/U(d). From the extremal property (2.5a) of $|(\lambda)_{\min}\rangle$, it follows that the CS can be written as

$$|\mathbf{u}\rangle = \exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{u}^{*}\mathbf{D}^{\dagger})|(\lambda)_{\min}\rangle = \exp\left[\sum_{i<j}(1+\delta_{ij})^{-1}u^{*}_{ij}D^{\dagger}_{ij}\right]|(\lambda)_{\min}\rangle, \tag{3.3}$$

where $\exp(\frac{1}{2}\mathrm{tr}\,\mathbf{u}^{*}\mathbf{D}^{\dagger})$ is an element of the complex extension Sp(2d,C) of Sp(2d,R), and $\mathbf{u}=\|u_{ij}\|$ a complex symmetric $d\times d$ matrix, subject to the condition that $\mathbf{I}-\mathbf{u}^{*}\mathbf{u}$ be a positive-definite (Hermitian) matrix. Here * denotes complex conjugation.

It remains to consider those cases where $\lambda_1,\ldots,\lambda_d$ are not all equal. The stability group H of $|(\lambda)_{\min}\rangle$ is then a proper subgroup of U(d). For such cases, in Ref. 27 Deenen and Quesne introduced the so-called Perelomov PCS, defined by

$$|\mathbf{u};(\lambda)\rangle = \exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{u}^{*}\mathbf{D}^{\dagger})|(\lambda)\rangle, \tag{3.4}$$

where $\mathbf{u}=\|u_{ij}\|$ again parametrizes the coset space Sp(2d,R)/U(d), and $|(\lambda)\rangle$ are the discrete basis states of the U(d) irrep $[\lambda]$, introduced in Eq. (2.6). The success encountered in determining the reproducing kernel and the measure for such states suggests a parametrization of the coset space Sp(2d,R)/H, based upon the factorization

$$\mathrm{Sp}(2d,R)/H = [\mathrm{Sp}(2d,R)/\mathrm{U}(d)][\mathrm{U}(d)/H]. \tag{3.5}$$

The corresponding CS, which are similar to those considered by Rowe,[15] can be written as

$$|\mathbf{u},\mathbf{z}\rangle = \exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{u}^{*}\mathbf{D}^{\dagger})|\mathbf{z}\rangle, \tag{3.6}$$

where $\mathbf{u}$ is the same matrix as above, and $\mathbf{z}$ denotes a set of parameters specifying the points of the coset space U(d)/H. The states $|\mathbf{z}\rangle$ are U(d) [or equivalently SU(d)] CS corresponding to the unitary irrep $[\lambda]$ and the reference state $|(\lambda)_{\min}\rangle$. In the remainder of this section, we shall review such states for the cases of U(2) and U(3), and outline their generalization to U(d) for $d>3$.

The U(2) [or SU(2)] CS, entering the definition of Sp(4,R) CS in the case where $\lambda_1>\lambda_2$, have been known for a long time, and are referred to in the literature as either spin[38] or atomic[39] CS. The stability group H of $|(\lambda)_{\min}\rangle$ is the U(1)$\times$U(1) subgroup of U(2), where the two U(1) groups are generated by E_{11} and E_{22}, respectively. The U(2) CS are parametrized by a single complex variable $z_{21}\equiv z$ varying in the whole complex plane, as follows:

$$|z\rangle = \exp(z^{*}E_{12})|(\lambda)_{\min}\rangle. \tag{3.7}$$

In terms of the U(2) generators, those of SU(2) can be written as

$$J_{+}=E_{12}, \quad J_{-}=E_{21}, \quad J_0=\tfrac{1}{2}(E_{11}-E_{22}). \tag{3.8}$$

The state $|(\lambda)_{\min}\rangle$ is characterized by an angular momentum $j=(\lambda_1-\lambda_2)/2$, and a projection $m=-j$. Hence Eq. (3.7) reduces to the usual definition of SU(2) CS,

$$|z\rangle = \exp(z^{*}J_{+})|j\,{-j}\rangle. \tag{3.9}$$

For the U(3) [or SU(3)] CS, entering the definition of Sp(6,R) CS when $\lambda_1,\lambda_2,\lambda_3$ are not all equal, we have to distinguish between three cases [henceforth referred to as (a), (b), and (c)] according as $\lambda_1>\lambda_2>\lambda_3$, $\lambda_1>\lambda_2=\lambda_3$, or $\lambda_1=\lambda_2>\lambda_3$ (see Ref. 36). From the known action of the U(3) generators on a basis state of a U(3) irrep,[40] it is easy to see that the corresponding stability group H is, respectively, (a) U(1)$\times$U(1)$\times$U(1), where the three U(1) groups are generated by E_{ii}, $i=1,2,3$; (b) U(2)$\times$U(1), where U(2) is generated by E_{ij}, $i,j=1,2$, and U(1) by E_{33}; and (c) U(1)$\times$U(2), where U(1) is generated by E_{11}, and U(2) by E_{ij}, $i,j=2,3$.

For case (a), it is advantageous to use the canonical coset decomposition of the unitary group,[41] to factorize the coset

space $U(3)/U(1)\times U(1)\times U(1)$ (see Ref. 36) into either of the two following products of coset spaces:

$$U(3)/U(1)\times U(1)\times U(1) = [U(3)/U(2)\times U(1)]\,[U(2)\times U(1)/U(1)\times U(1)\times U(1)], \tag{3.10}$$

$$U(3)/U(1)\times U(1)\times U(1) = [U(3)/U(1)\times U(2)]\,[U(1)\times U(2)/U(1)\times U(1)\times U(1)], \tag{3.11}$$

where $U(2)\times U(1)$ and $U(1)\times U(2)$ are generated, respectively, by E_{ij}, $i,j=1,2$, E_{33}, and E_{11}, E_{ij}, $i,j=2,3$. The corresponding parametrizations of the U(3) CS will be denoted by **x** and **y**, the symbol **z** being used whenever a specific choice is not demanded.

In the **x** parametrization, the U(3) CS are defined as follows:

$$|\mathbf{x}\rangle = \exp(x_2^* E_{13} + x_1^* E_{23})\exp(x_3^* E_{12})|(\lambda)_{\min}\rangle, \tag{3.12}$$

in terms of three complex variables $x_{32}\equiv x_1$, $x_{31}\equiv x_2$, and $x_{21}\equiv x_3$, varying in the whole complex plane, the first two parametrizing the coset space $U(3)/U(2)\times U(1)$, and the last one $U(2)\times U(1)/U(1)\times U(1)\times U(1)$. In the **y** parametrization, they are written as

$$|\mathbf{y}\rangle = \exp(y_3^* E_{12} + y_2^* E_{13})\exp(y_1^* E_{23})|(\lambda)_{\min}\rangle, \tag{3.13}$$

where the complex parameters $y_{32}\equiv y_1$, and $y_{31}\equiv y_2$, $y_{21}\equiv y_3$, again varying in the whole complex plane, parametrize the coset spaces $U(1)\times U(2)/U(1)\times U(1)\times U(1)$ and $U(3)/U(1)\times U(2)$, respectively.

The relation between the **x** and **y** parametrizations can be obtained easily by making use of Baker–Campbell–Hausdorff (BCH) formulas in matrix form.[42,43] By realizing the $Gl(3,C)$ generators E_{ij}, $i,j=1,2,3$, by 3×3 matrices with $+1$ at the intersection of row i and column j and 0 elsewhere, the $Gl(3,C)$ group element on the right-hand side of Eq. (3.12) is converted into the 3×3 matrix

$$\begin{pmatrix}1&0&x_2^*\\0&1&x_1^*\\0&0&1\end{pmatrix}\begin{pmatrix}1&x_3^*&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&x_3^*&x_2^*\\0&1&x_1^*\\0&0&1\end{pmatrix}. \tag{3.14}$$

In the same way, the $Gl(3,C)$ group element on the right-hand side of Eq. (3.13) becomes the following matrix:

$$\begin{pmatrix}1&y_3^*&y_2^*\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&y_1^*\\0&0&1\end{pmatrix}=\begin{pmatrix}1&y_3^*&y_2^*+y_1^*y_3^*\\0&1&y_1^*\\0&0&1\end{pmatrix}. \tag{3.15}$$

Identifying the CS $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ now amounts to equating the right-hand sides of Eqs. (3.14) and (3.15). The sought-for relations therefore read

$$x_1 = y_1, \quad x_2 = y_2 + y_1 y_3, \quad x_3 = y_3, \tag{3.16a}$$

or

$$y_1 = x_1, \quad y_2 = x_2 - x_1 x_3, \quad y_3 = x_3. \tag{3.16b}$$

Cases (b) and (c) can be treated by specializing either the **x** or **y** parametrization of case (a). For case (b), by setting $x_3=0$ in Eq. (3.12), we obtain the following CS:

$$|\mathbf{x}\rangle = \exp(x_2^* E_{13} + x_1^* E_{23})|(\lambda)_{\min}\rangle, \tag{3.17}$$

while for case (c), by setting $y_1=0$ in Eq. (3.13), we get

$$|\mathbf{y}\rangle = \exp(y_3^* E_{12} + y_2^* E_{13})|(\lambda)_{\min}\rangle. \tag{3.18}$$

The discussion of the $U(d)$ [or $SU(d)$] CS becomes quite tedious when $d>3$ because there are many possibilities for some of the λ_i's being equal. We shall therefore restrict ourselves here to the generic case for which $\lambda_1>\lambda_2>\cdots>\lambda_d$. The generalizations of the **x** and **y** parametrizations, introduced above for U(3), are connected, respectively, with the factorizations

$$U(d)/U(1)\times U(1)\times\cdots\times U(1) = [U(d)/U(d-1)\times U(1)][U(d-1)\times U(1)/U(d-2)\times U(1)\times U(1)] \times\cdots\times[U(2)\times U(1)\times\cdots\times U(1)/U(1)\times U(1)\times\cdots\times U(1)], \tag{3.19}$$

and

$$U(d)/U(1)\times U(1)\times\cdots\times U(1) = [U(d)/U(1)\times U(d-1)][U(1)\times U(d-1)/U(1)\times U(1)\times U(d-2)] \times\cdots\times[U(1)\times\cdots\times U(1)\times U(2)/U(1)\times\cdots\times U(1)\times U(1)]. \tag{3.20}$$

The corresponding CS can be written as

$$|\mathbf{z}\rangle = [V(\mathbf{z})]^{\dagger}|(\lambda)_{\min}\rangle, \tag{3.21}$$

where

$$V(\mathbf{x}) = \exp(x_{21}E_{21})\cdots\exp\left(\sum_{i=1}^{d-2} x_{d-1,i}E_{d-1,i}\right) \times\exp\left(\sum_{i=1}^{d-1} x_{di}E_{di}\right), \tag{3.22a}$$

$$V(\mathbf{y}) = \exp(y_{d,d-1}E_{d,d-1})\cdots\left(\exp\sum_{i=3}^{d} y_{i2}E_{i2}\right) \times\exp\left(\sum_{i=2}^{d} y_{i1}E_{i1}\right), \tag{3.22b}$$

and both parametrizations are related by

$$x_{ji} = y_{ji} + \sum_{i<k<j} y_{jk}\,y_{ki} + \sum_{i<k<l<j} y_{jl}\,y_{lk}\,y_{ki} + \cdots + y_{j,j-1}\,y_{j-1,j-2}\cdots y_{i+1,i}, \tag{3.23a}$$

or

$$y_{ji} = x_{ji} - \sum_{i<k<j} x_{jk}x_{ki} + \sum_{i<k<l<j} x_{jl}x_{lk}x_{ki} - \cdots + (-1)^{j-i-1}x_{j,j-1}x_{j-1,j-2}\cdots x_{i+1,i}. \tag{3.23b}$$

Note that in the U(2) case both parametrizations coincide.

As is well known,[4–6] the CS do not form an orthonormal

set, and their overlap acts as a reproducing kernel in a Hilbert space of analytic functions. In the next section, we shall proceed to calculate the overlap of two Sp(2d,R) CS, postponing the statement of the reproducing kernel property until Sec. VII.

IV. OVERLAP OF TWO Sp(2d,R) COHERENT STATES

Let

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \langle \mathbf{u}',\mathbf{z}'|\mathbf{u},\mathbf{z}\rangle \tag{4.1}$$

denote the overlap of two Sp(2d,R) CS corresponding to the generic case. By introducing Eq. (3.6) into Eq. (4.1), the overlap can be expressed as the matrix element of some operator between two U(d) CS, as follows:

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \langle \mathbf{z}'|\exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{u}'\mathbf{D})\exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{u}^*\mathbf{D}^\dagger)|\mathbf{z}\rangle. \tag{4.2}$$

In Appendix A of Ref. 27, it has been proved that the operator appearing on the right-hand side of Eq. (4.2) can be rewritten in normally ordered form,

$$\exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{u}'\mathbf{D})\exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{u}^*\mathbf{D}^\dagger) = \exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{a}\mathbf{D}^\dagger)\exp(\mathrm{tr}\ \mathbf{b}\mathbf{E})\exp(\tfrac{1}{2}\mathrm{tr}\ \mathbf{c}\mathbf{D}), \tag{4.3}$$

where the matrices $\mathbf{a},\mathbf{b},\mathbf{c}$ are given in terms of $\mathbf{u}'$ and $\mathbf{u}^*$ by the following relations:

$$\mathbf{a} = \mathbf{U}^{-1}\mathbf{u}^*, \quad \exp \mathbf{b} = \tilde{\mathbf{U}}^{-1},$$
$$\mathbf{c} = \mathbf{u}'\mathbf{U}^{-1}, \quad \mathbf{U} = \mathbf{I} - \mathbf{u}^*\mathbf{u}'. \tag{4.4}$$

Here $\tilde{\mathbf{U}}$ stands for the transpose of $\mathbf{U}$. From the Sp(2d,R) commutation relations and from the definitions of the U(d) CS and of the lowest weight state $|(\lambda)_{\min}\rangle$ given, respectively, in Eqs. (2.3), (3.21), and (2.5), it is obvious that the U(d) CS are annihilated by the generators D_{ij}, i.e.,

$$D_{ij}|\mathbf{z}\rangle = 0. \tag{4.5}$$

Hence, by taking Eq. (4.3) into account, Eq. (4.2) reduces to

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \langle \mathbf{z}'|\exp(\mathrm{tr}\ \mathbf{b}\mathbf{E})|\mathbf{z}\rangle. \tag{4.6}$$

By introducing Eq. (3.21) into Eq. (4.6), the latter becomes

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \langle (\lambda)_{\min}|V(\mathbf{z}')\exp(\mathrm{tr}\ \mathbf{b}\mathbf{E})[V(\mathbf{z})]^\dagger|(\lambda)_{\min}\rangle, \tag{4.7}$$

where $V(\mathbf{z})$ is the Gl(d,C) group element defined in Eq. (3.22). Equation (4.7) expresses the Sp(2d,R) CS overlap as the diagonal matrix element of the Gl(d,C) group element $V(\mathbf{z}')\exp(\mathrm{tr}\ \mathbf{b}\mathbf{E})[V(\mathbf{z})]^\dagger$ with respect to the lowest weight state $|(\lambda)_{\min}\rangle$ of the U(d) [or Gl(d,C)] irrep $[\lambda]$. In the realization of the Gl(d,C) generators E_{ij} in terms of $d\times d$ matrices with $+1$ at the intersection of row i and column j and 0 elsewhere, the Gl(d,C) group element $\exp(\mathbf{b}\mathbf{E})$ corresponds to the matrix $\mathbf{U}^{-1}$, while $V(\mathbf{z})$ is realized by some $d\times d$ matrix $\mathbf{Z}$ with entries 0 and 1 above and on the diagonal, respectively. Hence, to the product of Gl(d,C) group elements $V(\mathbf{z}')\exp(\mathrm{tr}\ \mathbf{b}\mathbf{E})[V(\mathbf{z})]^\dagger$, corresponds the matrix product

$$\mathbf{W} = \mathbf{Z}'\mathbf{U}^{-1}\mathbf{Z}^\dagger, \tag{4.8}$$

where $\mathbf{Z}'$ is obtained from $\mathbf{Z}$ by replacing z_{ij} by z'_{ij}. The CS overlap is therefore nothing else but the element of the Gl(d,C) representation matrix $\mathbf{D}^{[\lambda]}(\mathbf{W})$, whose row and column indices are both equal to $(\lambda)_{\min}$,

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = D^{[\lambda]}_{(\lambda)_{\min}(\lambda)_{\min}}(\mathbf{W}). \tag{4.9}$$

A theorem first stated by Louck[44] for U(d), then extended by Brunet and Seligman[45] to Gl(d,C), assures that the elements of the representation matrix $D^{[\lambda]}_{(\lambda')(\lambda)}(\mathbf{W})$, specified by two Gel'fand patterns (λ) and (λ'), coincide with the (appropriately normalized) double Gel'fand states $P^{[\lambda]}_{(\lambda')(\lambda)}(\mathbf{W})$, characterized by the same Gel'fand patterns and depending upon the d^2 complex variables W_{ij}, $i,j = 1,...,d$. Consequently, Eq. (4.9) leads to the following result:

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = (W_{dd})^{\lambda_1-\lambda_2}(W_{d-1d,d-1d})^{\lambda_2-\lambda_3}\times\cdots\times(W_{2\cdots d,2\cdots d})^{\lambda_{d-1}-\lambda_d}(W_{1\cdots d,1\cdots d})^{\lambda_d+n/2}. \tag{4.10}$$

Here $W_{d-i+1\cdots d,d-i+1\cdots d}$, $1\leqslant i\leqslant d$, is the minor of order i of $\det \mathbf{W}$ defined by

$$W_{d-i+1\cdots d,d-i+1\cdots d} = \sum_p (-1)^p W_{d-i+1,p(d-i+1)}\times W_{d-i+2,p(d-i+2)}\cdots W_{d,p(d)}, \tag{4.11}$$

where the summation is carried out over the $i!$ permutations of the indices $d-i+1,...,d$. From Eq. (4.8) and the definitions of $\mathbf{Z}$ and $\mathbf{Z}'$, it follows that

$$W_{1\cdots d,1\cdots d} = \det \mathbf{W} = (\det \mathbf{U})^{-1}. \tag{4.12}$$

Hence, by setting

$$T_i(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = (\det \mathbf{U})W_{d-i+1\cdots d,d-i+1\cdots d}, \quad i = 1,...,d-1, \tag{4.13}$$

Eq. (4.10) can be put into the following equivalent form:

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = (\det \mathbf{U})^{-\lambda_1-n/2}\times\prod_{i=1}^{d-1}[T_i(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)]^{\lambda_i-\lambda_{i+1}}. \tag{4.14}$$

To obtain a detailed expression for the CS overlap, it remains to calculate explicitly the T_i's. Since the transition from the $\mathbf{x}$ to the $\mathbf{y}$ parametrization can be easily performed by applying Eq. (3.23a), we need to consider only the former, wherein the matrix $\mathbf{Z}$ takes the following form:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ x_{21} & 1 & 0 & \cdots & 0 & 0 \\ x_{31} & x_{32} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ x_{d1} & x_{d2} & x_{d3} & \cdots & x_{d,d-1} & 1 \end{pmatrix}. \tag{4.15}$$

For Sp(4,R), Eq. (4.14) only involves a single T_i, given by

$$T_1(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = U_{11} - U_{12}z' - U_{21}z^* + U_{22}z'z^*. \tag{4.16}$$

For Sp(6,R), there appear two T_i's, respectively equal to

$$T_1(\mathbf{u}',\mathbf{x}';\mathbf{u}^*,\mathbf{x}^*) = U_{12,12} - U_{12,13}x'_1 + U_{12,23}x'_2 - U_{13,12}x^*_1 + U_{23,12}x^*_2 + U_{13,13}x'_1x^*_1 - U_{23,13}x'_1x^*_2 - U_{13,23}x'_2x^*_1 + U_{23,23}x'_2x^*_2, \tag{4.17a}$$

or

$$T_1(\mathbf{u}',\mathbf{y}';\mathbf{u}^*,\mathbf{y}^*)$$
$$= U_{12,12} - U_{12,13}\,y_1' + U_{12,23}(y_2' + y_1'\,y_3')$$
$$- U_{13,12}\,y_1^* + U_{23,12}(y_2^* + y_1^*\,y_3^*) + U_{13,13}\,y_1'\,y_1^*$$
$$- U_{23,13}\,y_1'(y_2^* + y_1^*\,y_3^*) - U_{13,23}(y_2' + y_1'\,y_3')\,y_1^*$$
$$+ U_{23,23}(y_2' + y_1'\,y_3')(y_2^* + y_1^*\,y_3^*), \tag{4.17b}$$

and

$$T_2(\mathbf{u}',\mathbf{x}';\mathbf{u}^*,\mathbf{x}^*)$$
$$= U_{11} - U_{13}(x_2' - x_1'x_3') - U_{12}x_3' - U_{31}(x_2^* - x_1^*x_3^*)$$
$$- U_{21}x_3^* + U_{33}(x_2' - x_1'x_3')(x_2^* - x_1^*x_3^*)$$
$$+ U_{23}(x_2' - x_1'x_3')x_3^*$$
$$+ U_{32}x_3'(x_2^* - x_1^*x_3^*) + U_{22}x_3'x_3^*, \tag{4.18a}$$

or

$$T_2(\mathbf{u}',\mathbf{y}';\mathbf{u}^*,\mathbf{y}^*)$$
$$= U_{11} - U_{13}\,y_2' - U_{12}\,y_3' - U_{31}\,y_2^* - U_{21}\,y_3^*$$
$$+ U_{33}\,y_2'\,y_2^* + U_{23}\,y_2'\,y_3^* + U_{32}\,y_3'\,y_2^* + U_{22}\,y_3'\,y_3^*, \tag{4.18b}$$

where $U_{ij,kl}$ denotes the minor of order 2 of $\det \mathbf{U}$ corresponding to rows i,j, and columns k,l. Since Eq. (4.14) is valid for the generic case, corresponding to case (a) for Sp(6,R), it follows from the discussion carried out in the previous section that the counterpart of Eq. (4.14) for cases (b) and (c) irreps of Sp(6,R) reads, respectively,

$$\hat{K}(\mathbf{u}',\mathbf{x}';\mathbf{u}^*,\mathbf{x}^*) = (\det \mathbf{U})^{-\lambda_1 - n/2}[T_1(\mathbf{u}',\mathbf{x}';\mathbf{u}^*,\mathbf{x}^*)]^{\lambda_1 - \lambda_2}, \tag{4.19}$$

and

$$\hat{K}(\mathbf{u}',\mathbf{y}';\mathbf{u}^*,\mathbf{y}^*) = (\det \mathbf{U})^{-\lambda_1 - n/2}[T_2(\mathbf{u}',\mathbf{y}';\mathbf{u}^*,\mathbf{y}^*)]^{\lambda_1 - \lambda_3}. \tag{4.20}$$

Here $T_1(\mathbf{u}',\mathbf{x}';\mathbf{u}^*,\mathbf{x}^*)$ and $T_2(\mathbf{u}',\mathbf{y}';\mathbf{u}^*,\mathbf{y}^*)$ are deduced from Eqs. (4.17a) and (4.18b) by setting there x_3' and x_3^*, or y_1' and y_1^*, equal to zero. This completes the derivation of the CS overlap for Sp(4,R) and Sp(6,R).

As a final point, let us note that if we set either $\mathbf{z} = \mathbf{z}' = \mathbf{0}$ and $\lambda_1 = \cdots = \lambda_d = \lambda$ or $\mathbf{u} = \mathbf{u}' = \mathbf{0}$ in Eq. (4.14), we obtain, respectively, the Sp(2d,R) CS overlap for the case where all the λ_i's are equal to λ,

$$\hat{K}(\mathbf{u}';\mathbf{u}^*) = \hat{K}(\mathbf{u}',\mathbf{0};\mathbf{u}^*,\mathbf{0}) = (\det \mathbf{U})^{-\lambda - n/2}, \tag{4.21}$$

or the U(d) CS overlap for the irrep $[\lambda]$,

$$\hat{K}(\mathbf{z}',\mathbf{z}^*) = \hat{K}(\mathbf{0},\mathbf{z}';\mathbf{0},\mathbf{z}^*) = \prod_{i=1}^{d-1}[T_i(\mathbf{z}';\mathbf{z}^*)]^{\lambda_i - \lambda_{i+1}}, \tag{4.22}$$

where the functions

$$T_i(\mathbf{z}';\mathbf{z}^*) = T_i(\mathbf{0},\mathbf{z}';\mathbf{0},\mathbf{z}^*) \tag{4.23}$$

are given by

$$T_1(z';z^*) = 1 + z'z^* \tag{4.24}$$

for Sp(4,R), and by

$$T_1(\mathbf{x}';\mathbf{x}^*) = 1 + x_1'x_1^* + x_2'x_2^*, \tag{4.25a}$$

or

$$T_1(\mathbf{y}';\mathbf{y}^*) = 1 + y_1'\,y_1^* + (y_2' + y_1'\,y_3')(y_2^* + y_1^*\,y_3^*), \tag{4.25b}$$

and

$$T_2(\mathbf{x}';\mathbf{x}^*) = 1 + (x_2' - x_1'x_3')(x_2^* - x_1^*x_3^*) + x_3'x_3^*, \tag{4.26a}$$

or

$$T_2(\mathbf{y}';\mathbf{y}^*) = 1 + y_2'\,y_2^* + y_3'\,y_3^*, \tag{4.26b}$$

for Sp(6,R).

If we now go back to the Sp(2d,R) CS (3.6) for the case where the λ_i's are not all equal, we note that although they are defined in terms of an operator product whose factors only depend upon either $\mathbf{u}^*$ or $\mathbf{z}^*$, their overlap (3.14) is not the product of a function of $\mathbf{u}'$ and $\mathbf{u}^*$ by a function of $\mathbf{z}'$ and $\mathbf{z}^*$, since the T_i's depend on both the $\mathbf{u}',\mathbf{u}^*$ and $\mathbf{z}',\mathbf{z}^*$ parameters. However, provided we replace $\mathbf{z}$ by some new parameters $\bar{\mathbf{z}}$, functions of both $\mathbf{u}$ and $\mathbf{z}$, the CS overlap

$$\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \hat{L}(\mathbf{u}',\bar{\mathbf{z}}';\mathbf{u}^*,\bar{\mathbf{z}}^*) \tag{4.27}$$

can be factorized as follows:

$$\hat{L}(\mathbf{u}',\bar{\mathbf{z}}';\mathbf{u}^*,\bar{\mathbf{z}}^*) = \left[D^{[\lambda]}_{(\lambda)_{\min}(\lambda)_{\min}}(\mathbf{U})\right]^{-1}\hat{K}(\bar{\mathbf{z}}';\bar{\mathbf{z}}^*), \tag{4.28}$$

where $D^{[\lambda]}_{(\lambda)_{\min}(\lambda)_{\min}}(\mathbf{U})$ and $\hat{K}(\bar{\mathbf{z}}';\bar{\mathbf{z}}^*)$ are obtained, respectively, by replacing $\mathbf{W}$ by $\mathbf{U}$, and $\mathbf{z}',\mathbf{z}^*$ by $\bar{\mathbf{z}}',\bar{\mathbf{z}}^*$ in Eqs. (4.10) and (4.22). The new parameters $\bar{\mathbf{z}}$ are defined by

$$\bar{z} = (\det \mathbf{U})^{-1/2}(U_{22}z - U_{21}) \tag{4.29}$$

for Sp(4,R), and by

$$\bar{x}_1 = (U_{23,23})^{-1/2}(U_{33}x_1 - U_{32}),$$
$$\bar{x}_2 = (U_{33})^{1/2}(U_{23,23}\det \mathbf{U})^{-1/2}$$
$$\times(U_{23,23}x_2 - U_{23,13}x_1 + U_{23,12}), \tag{4.30a}$$
$$\bar{x}_3 = (U_{33}\det \mathbf{U})^{-1/2}(U_{23,23}x_3 - U_{23,13}),$$

or

$$\bar{y}_1 = (U_{23,23})^{-1/2}(U_{33}\,y_1 - U_{32}),$$
$$\bar{y}_2 = (U_{23,23})^{1/2}(U_{33}\det \mathbf{U})^{-1/2}(U_{33}\,y_2 + U_{32}\,y_3 - U_{31}), \tag{4.30b}$$
$$\bar{y}_3 = (U_{33}\det \mathbf{U})^{-1/2}(U_{23,23}\,y_3 - U_{23,13}),$$

for Sp(6,R). In terms of them, the functions $T_i(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)$ can be rewritten as

$$T_1(\mathbf{u}';z';\mathbf{u}^*,z^*) = (\det \mathbf{U}/U_{22})\,T_1(\bar{z}';\bar{z}^*), \tag{4.31}$$

and

$$T_1(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = (\det \mathbf{U}/U_{33})\,T_1(\bar{\mathbf{z}}';\bar{\mathbf{z}}^*),$$
$$T_2(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = (\det \mathbf{U}/U_{23,23})\,T_2(\bar{\mathbf{z}}';\bar{\mathbf{z}}^*), \tag{4.32}$$

respectively.

V. ACTION OF THE Sp(2d,R) GENERATORS UPON THE COHERENT STATES

When they act upon the CS, the Sp(2d,R) generators are equivalent to some first-order partial differential operators, whose explicit form will be found in the present section. For such purpose, we shall first deal with the generic case in detail, then consider the transition to the remaining ones.

Let X denote any operator acting in $\mathcal{F}_{(\lambda)}$. When it is applied to the bra $\langle \mathbf{u},\mathbf{z}|$, it is equivalent to some partial differential operator $\hat{\mathscr{X}}$ with respect to $\mathbf{u}$ and $\mathbf{z}$,

$$\langle \mathbf{u},\mathbf{z}|X = \hat{\mathscr{X}}\langle \mathbf{u},\mathbf{z}|. \tag{5.1}$$

The Hermitian conjugate of Eq. (5.1) can be written as

$$X^\dagger|\mathbf{u},\mathbf{z}\rangle = \hat{\mathscr{X}}^*|\mathbf{u},\mathbf{z}\rangle, \tag{5.2}$$

where $\hat{\mathscr{X}}^*$ is a partial differential operator with respect to $\mathbf{u}^*$ and $\mathbf{z}^*$. In the next section, we shall prove that the CS satisfy a unity resolution relation, and therefore form a continuous basis of $\mathscr{T}_{\langle\lambda\rangle}$. Hence, in the corresponding representation, $\hat{\mathscr{X}}$ will be the representation of X.

In particular, for the Sp($2d,R$) generators, it directly results from the CS definition (3.6) that

$$\hat{\mathscr{D}} = \mathbf{\Delta}_u, \tag{5.3}$$

where $\mathbf{\Delta}_u$ is a $d\times d$ matrix whose elements $\Delta_{u_{ij}}$ are defined by

$$\Delta_{u_{ij}} = (1+\delta_{ij})\frac{\partial}{\partial u_{ij}}. \tag{5.4}$$

The explicit forms of $\hat{\mathscr{E}}$ and $\hat{\mathscr{D}}^\dagger$ can also be easily found by writing

$$\langle \mathbf{u},\mathbf{z}|X = \langle \mathbf{z}|\exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{uD})X\exp(-\tfrac{1}{2}\mathrm{tr}\,\mathbf{uD})\exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{uD}), \tag{5.5}$$

and using the BCH formula

$$\exp(Y)X\exp(-Y) = X + \sum_{m=1}^{\infty}\frac{1}{m!}[Y,[Y,\ldots,[Y,X]\cdots]]_m, \tag{5.6}$$

for $Y = \frac{1}{2}\,\mathrm{tr}\,\mathbf{uD}$. In matrix notation, the result reads

$$\hat{\mathscr{E}} = \mathbf{u}\mathbf{\Delta}_u + \mathring{\mathscr{E}}, \tag{5.7a}$$

$$\hat{\mathscr{D}}^\dagger = \mathbf{u}\tilde{\mathring{\mathscr{E}}} + \mathring{\mathscr{E}}\mathbf{u} + [\mathbf{u}\mathbf{\Delta}_u - (d+1)\mathbf{I}]\mathbf{u}, \tag{5.7b}$$

where $\mathring{\mathscr{E}}_{ij}$ denotes the partial differential operator with respect to $\mathbf{z}$, representing E_{ij} in the U(d) CS representation corresponding to the irrep $[\lambda]$, i.e.,

$$\langle \mathbf{z}|E_{ij} = \mathring{\mathscr{E}}_{ij}\langle \mathbf{z}|, \tag{5.8}$$

and $\tilde{\mathring{\mathscr{E}}}$ is the transpose of $\mathring{\mathscr{E}}$, i.e., $\tilde{\mathring{\mathscr{E}}}_{ij} = \mathring{\mathscr{E}}_{ji}$. Note that Eqs. (5.3) and (5.7) have the same structure as the corresponding equations in the PCS representation, Eqs. (6.11) of Ref. 27.

The explicit form of $\mathring{\mathscr{E}}_{ij}$ can be found in the same way. By writing the left-hand side of Eq. (5.8) as

$$\langle \mathbf{z}|E_{ij} = \langle(\lambda)_{\min}|V(\mathbf{z})E_{ij}[V(\mathbf{z})]^{-1}V(\mathbf{z}), \tag{5.9}$$

where $V(\mathbf{z})$ is defined in Eq. (3.22), and by repeatedly using the BCH formula (5.6), we get

$$\mathring{\mathscr{E}}_{11} = z\partial + \lambda_2 + n/2, \quad \mathring{\mathscr{E}}_{22} = -z\partial + \lambda_1 + n/2,$$
$$\mathring{\mathscr{E}}_{12} = z(\lambda_1 - \lambda_2 - z\partial), \quad \mathring{\mathscr{E}}_{21} = \partial, \tag{5.10}$$

for the U(2) generators, and

$$\begin{aligned}
\mathring{\mathscr{E}}_{11} &= x_2\partial_2 + x_3\partial_3 + \lambda_3 + n/2 = y_2\delta_2 + y_3\delta_3 + \lambda_3 + n/2,\\
\mathring{\mathscr{E}}_{22} &= x_1\partial_1 - x_3\partial_3 + \lambda_2 + n/2 = y_1\delta_1 - y_3\delta_3 + \lambda_2 + n/2,\\
\mathring{\mathscr{E}}_{33} &= -x_1\partial_1 - x_2\partial_2 + \lambda_1 + n/2\\
&= -y_1\delta_1 - y_2\delta_2 + \lambda_1 + n/2,\\
\mathring{\mathscr{E}}_{12} &= x_3(\lambda_2 - \lambda_3 - x_3\partial_3) + x_2\partial_1\\
&= y_3(\lambda_2 - \lambda_3 + y_1\delta_1 - y_2\delta_2 - y_3\delta_3) + y_2\delta_1,
\end{aligned} \tag{5.11}$$

$$\begin{aligned}
\mathring{\mathscr{E}}_{13} &= x_2(\lambda_1 - \lambda_3 - x_1\partial_1 - x_2\partial_2 - x_3\partial_3)\\
&\quad - x_1x_3(\lambda_2 - \lambda_3 - x_3\partial_3)\\
&= y_2(\lambda_1 - \lambda_3 - y_1\delta_1 - y_2\delta_2 - y_3\delta_3)\\
&\quad + y_1y_3(\lambda_1 - \lambda_2 - y_1\delta_1),\\
\mathring{\mathscr{E}}_{23} &= x_1(\lambda_1 - \lambda_2 - x_1\partial_1 - x_2\partial_2 + x_3\partial_3) - x_2\partial_3\\
&= y_1(\lambda_1 - \lambda_2 - y_1\delta_1) - y_2\delta_3,\\
\mathring{\mathscr{E}}_{21} &= x_1\partial_2 + \partial_3 = \delta_3, \quad \mathring{\mathscr{E}}_{31} = \partial_2 = \delta_2\\
\mathring{\mathscr{E}}_{32} &= \partial_1 = -y_3\delta_2 + \delta_1,
\end{aligned}$$

for the U(3) ones. In Eqs. (5.10) and (5.11), ∂, ∂_i, and δ_i are abbreviated forms of the differentiation symbols $\partial/\partial z$, $\partial/\partial x_i$, and $\partial/\partial y_i$, respectively. The explicit form of the U(3) generators in the $\mathbf{y}$ parametrization can be easily obtained from that in the $\mathbf{x}$ parametrization through the change of variables (3.16). Finally, the expressions of the U(3) generators valid for nongeneric irreps are inferred from Eq. (5.11) by deleting all the terms containing x_3 or ∂_3 for case (b) irreps, and y_1 or δ_1 for case (c) irreps.

It can be easily checked that $\hat{\mathscr{D}}_{ij}, \hat{\mathscr{E}}_{ij}, \hat{\mathscr{D}}^\dagger_{ij}$, as given by Eqs. (5.3) and (5.7), satisfy the Sp($2d,R$) commutation relations (2.3) as it should be. It will be shown in Sec. VII that they also fulfill the required Hermiticity conditions of the Sp($2d,R$) generators, given in Eq. (2.2), with respect to the measure to be determined in the next section.

To conclude the present section, let us point out that the differential equations satisfied by the overlap $\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)$ can be deduced straightforwardly from the Hermiticity properties (2.2). By taking Eqs. (5.1) and (5.2) into account, Eq. (2.2) indeed leads to the following equations:

$$\begin{aligned}
[\hat{\mathscr{E}}'_{ij} - \hat{\mathscr{E}}^*_{ji}]\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = 0,\\
[\hat{\mathscr{D}}'^\dagger_{ij} - \hat{\mathscr{D}}^*_{ij}]\hat{K}(\mathbf{u}',\mathbf{z}',\mathbf{u}^*,\mathbf{z}^*) = 0,
\end{aligned} \tag{5.12}$$

where $\hat{\mathscr{E}}'_{ij}$ and $\hat{\mathscr{D}}'^\dagger_{ij}$ depend upon $\mathbf{u}',\mathbf{z}'$, and the corresponding differentiation operators. When combined with Eqs. (5.3) and (5.7), Eq. (5.12) becomes

$$\begin{aligned}
&[(\mathbf{u}'\mathbf{\Delta}_{u'})_{ij} - (\mathbf{u}^*\mathbf{\Delta}_{u^*})_{ji} + \mathring{\mathscr{E}}'_{ij} - \mathring{\mathscr{E}}^*_{ji}]\\
&\quad\times\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = 0,\\
&\{\mathbf{u}'\tilde{\mathring{\mathscr{E}}}' + \mathring{\mathscr{E}}'\mathbf{u}' + [\mathbf{u}'\mathbf{\Delta}_{u'} - (d+1)\mathbf{I}]\mathbf{u}' - \mathbf{\Delta}_{u^*}\}_{ij}\\
&\quad\times\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = 0,
\end{aligned} \tag{5.13}$$

where it only remains to introduce the explicit expressions (5.10) or (5.11) of $\mathring{\mathscr{E}}_{ij}$.

VI. UNITY RESOLUTION RELATION

From the general theory of unitary-operator CS (see Ref. 5) it follows that the Sp($2d,R$) CS, defined in Eq. (3.6), must satisfy a unity resolution relation

$$\int d\hat{o}(\mathbf{u},\mathbf{z})|\mathbf{u},\mathbf{z}\rangle\langle \mathbf{u},\mathbf{z}| = I_{\langle\lambda\rangle}, \tag{6.1}$$

for some appropriately chosen measure $d\hat{o}(\mathbf{u},\mathbf{z})$. In Eq. (6.1), $I_{\langle\lambda\rangle}$ is the unit operator in $\mathscr{T}_{\langle\lambda\rangle}$, and the parameters $\mathbf{u}$ vary in the domain determined by the condition $\mathbf{I} - \mathbf{u}^*\mathbf{u} > 0$, while each of the parameters $\mathbf{z}$ varies over the whole complex plane. The measure

$$d\hat{\sigma}(\mathbf{u},\mathbf{z}) = \hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*)\,d\mathbf{u}\,d\mathbf{u}^*\,d\mathbf{z}\,d\mathbf{z}^* \tag{6.2}$$

is expressed in terms of a weight function $\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*)$, which is a real-valued, positive function of the indicated variables, vanishing on the boundary of the integration domain. This weight function could be derived from the invariant measure on the coset space $\mathrm{Sp}(2d,R)/H$ (see Ref. 5). The present section purpose is to propose an alternative method for its determination, and to apply it in full details to the $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$ cases.

Let us denote by $\hat{O}$ the operator on the left-hand side of Eq. (6.1). For $\hat{O}$ to be the unit operator in $\mathscr{F}_{\langle\lambda\rangle}$, it is sufficient that it commutes with all the $\mathrm{Sp}(2d,R)$ generators,

$$[E_{ij},\hat{O}] = [D^{\dagger}_{ij},\hat{O}] = [D_{ij},\hat{O}] = 0, \tag{6.3}$$

and that the measure satisfies the normalization condition

$$\int d\hat{\sigma}(\mathbf{u},\mathbf{z}) = 1. \tag{6.4}$$

From Eq. (6.3) and Schur's lemma within the representation space $\mathscr{F}_{\langle\lambda\rangle}$ of the irrep $\langle\lambda\rangle$, we indeed infer that $\hat{O}$ is a multiple of the unit operator in $\mathscr{F}_{\langle\lambda\rangle}$, while from Eq. (6.4) and the relation

$$\langle\mathbf{u},\mathbf{z}|(\lambda)_{\min}\rangle = \langle(\lambda)_{\min}|(\lambda)_{\min}\rangle = 1, \tag{6.5}$$

we conclude that the multiplicative constant is equal to 1. It is an easy matter to satisfy Eq. (6.4) since it merely fixes the normalization coefficient of the weight function. In contrast, Eq. (6.3) determines the dependence of the weight function upon the variables $\mathbf{u},\mathbf{u}^*,\mathbf{z}$, and $\mathbf{z}^*$, and is more difficult to solve.

From the fact that $\hat{O}$ is a Hermitian operator, and from the symmetry and Hermiticity properties of the $\mathrm{Sp}(2d,R)$ generators, given, respectively, in Eqs. (2.1) and (2.2), it follows that only part of the conditions expressed in Eq. (6.3) are independent, namely

$$[E_{ij},\hat{O}] = [D^{\dagger}_{ij},\hat{O}] = 0, \quad i\leqslant j. \tag{6.6}$$

By taking Eqs. (5.1), (5.2), (5.3), and (5.7) into account, Eq. (6.6) can be rewritten as

$$\int d\hat{\sigma}(\mathbf{u},\mathbf{z})[\hat{\mathscr{E}}^*_{ji} - \hat{\mathscr{E}}_{ij}]\,|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}|$$

$$= \int d\hat{\sigma}(\mathbf{u},\mathbf{z})[(\mathbf{u}^*\Delta_{u^*})_{ji} + \mathring{\mathscr{E}}^*_{ji} - (\mathbf{u}\Delta_u)_{ij} - \mathring{\mathscr{E}}_{ij}]$$

$$\times|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}| = 0, \quad i\leqslant j, \tag{6.7a}$$

and

$$\int d\hat{\sigma}(\mathbf{u},\mathbf{z})[\hat{\mathscr{D}}^*_{ij} - \hat{\mathscr{D}}^{\dagger}_{ij}]\,|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}|$$

$$= \int d\hat{\sigma}(\mathbf{u},\mathbf{z})\{\Delta_{u^*} - (\mathbf{u}\tilde{\mathscr{E}} + \mathring{\mathscr{E}}\mathbf{u}) - [\mathbf{u}\Delta_u$$

$$-(d+1)\mathbf{I}]\mathbf{u}\}_{ij}|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}| = 0, \quad i\leqslant j, \tag{6.7b}$$

where we used the fact that $|\mathbf{u},\mathbf{z}\rangle$ and $\langle\mathbf{u},\mathbf{z}|$ only depend upon $\mathbf{u}^*,\mathbf{z}^*$ or $\mathbf{u},\mathbf{z}$, respectively, and the latter are independent variables.

In Eq. (6.7), the first-order partial differential operators acting upon $|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}|$ can be transferred to the weight function by integrating by parts and noting that the weight function vanishing on the domain boundary ensures that of the limit terms. Hence Eqs. (6.7a) and (6.7b) become

$$\int d\mathbf{u}\,d\mathbf{u}^*\,d\mathbf{z}\,d\mathbf{z}^*\{[(\mathbf{u}\Delta_u)_{ij} - (\mathbf{u}^*\Delta_{u^*})_{ji}$$

$$+A_{ij} + a_{ij}]\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*)\}\,|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}| = 0, \quad i\leqslant j, \tag{6.8a}$$

and

$$\int d\mathbf{u}\,d\mathbf{u}^*\,d\mathbf{z}\,d\mathbf{z}^*\left\{\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk} - \Delta_{u^*_{ij}}\right.\right.$$

$$\left.\left.+B_{ij} + b_{ij}\right]\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*)\right\}|\mathbf{u},\mathbf{z}\rangle\,\langle\mathbf{u},\mathbf{z}| = 0, \quad i\leqslant j, \tag{6.8b}$$

where A_{ij},B_{ij} and a_{ij},b_{ij} denote, respectively, the first-order partial differential operators and the functions coming from the terms containing $\mathring{\mathscr{E}}$ or $\mathring{\mathscr{E}}^*$. In Appendix A, their explicit expressions are given for the cases of $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$.

Equations (6.8a) and (6.8b) will be satisfied provided the weight function is a solution of the following system of first-order partial differential equations:

$$[(\mathbf{u}\Delta_u)_{ij} - (\mathbf{u}^*\Delta_{u^*})_{ji} + A_{ij} + a_{ij}]$$

$$\times\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*) = 0, \quad i\leqslant j, \tag{6.9a}$$

$$\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk} - \Delta_{u^*_{ij}} + B_{ij} + b_{ij}\right]$$

$$\times\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*) = 0, \quad i\leqslant j. \tag{6.9b}$$

We have therefore reduced the determination of the weight function to that of a particular solution of Eqs. (6.9a) and (6.9b).

In Appendix B, it is proved that in the cases of $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$, such a solution can be obtained in the following form:

$$\hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z},\mathbf{z}^*) = \hat{A}\,(\det\mathbf{U})^{\alpha}\prod_{i=1}^{d-1}[T_i(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*)]^{\beta_i}. \tag{6.10}$$

Here $\hat{A}$ is the normalization coefficient determined by Eq. (6.4), $\mathbf{U}$ and T_i are defined, respectively, by Eqs. (4.4) and (4.16)–(4.18), where we set $\mathbf{u}' = \mathbf{u}$ and $\mathbf{z}' = \mathbf{z}$, and the exponents α and β_i, $i = 1,\ldots,d-1$, are given by

$$\alpha = \lambda_1 + n/2 - 2, \quad \beta_1 = -(\lambda_1 - \lambda_2 + 2), \tag{6.11}$$

in the $\mathrm{Sp}(4,R)$ case, and by

$$\begin{aligned}
&\alpha = \lambda_1 + n/2 - 2, \quad \beta_1 = -(\lambda_1 - \lambda_2 + 2),\\
&\beta_2 = -(\lambda_2 - \lambda_3 + 2), \quad \text{for case (a) irreps},\\
&\alpha = \lambda_1 + n/2 - 2, \quad \beta_1 = -(\lambda_1 - \lambda_2 + 3),\\
&\beta_2 = 0, \quad \text{for case (b) irreps},\\
&\alpha = \lambda_1 + n/2 - 3, \quad \beta_1 = 0,\\
&\beta_2 = -(\lambda_1 - \lambda_3 + 3), \quad \text{for case (c) irreps},
\end{aligned} \tag{6.12}$$

in the $\mathrm{Sp}(6,R)$ one. Note that for cases (a), (b), and (c) irreps, the variables $\mathbf{z}$ are, respectively, x_1,x_2,x_3 (or y_1,y_2,y_3), x_1,x_2, and y_2,y_3. For those $\mathrm{Sp}(2d,R)$ irreps for which all the λ_i's are equal to λ, the weight function is also given by Eq. (6.10) with[21,25,37]

$$\alpha = \lambda + n/2 - d - 1, \quad \beta_i = 0, \quad i = 1,\ldots,d-1. \tag{6.13}$$

In this case, no variables $\mathbf{z},\mathbf{z}^*$ are present in Eqs. (6.1) and (6.2).

To calculate the normalization coefficient $\hat{A}$ of the weight function, it is convenient to replace $\mathbf{z}$ by the variables $\bar{\mathbf{z}}$, defined in Eqs. (4.29) and (4.30). The transformed weight function $\hat{g}(\mathbf{u},\mathbf{u}^*,\bar{\mathbf{z}},\bar{\mathbf{z}}^*)$ can be written in terms of the old one $\hat{f}$, and the Jacobian J of the transformation, depending only on $\mathbf{u}$ and $\mathbf{u}^*$, as follows:

$$\hat{g}(\mathbf{u},\mathbf{u}^*,\bar{\mathbf{z}},\bar{\mathbf{z}}^*) = \hat{f}(\mathbf{u},\mathbf{u}^*,\mathbf{z}(\mathbf{u},\mathbf{u}^*,\bar{\mathbf{z}}),\mathbf{z}^*(\mathbf{u},\mathbf{u}^*,\bar{\mathbf{z}}^*))|J(\mathbf{u},\mathbf{u}^*)|. \tag{6.14}$$

In all the cases listed in Eqs. (6.11)–(6.13), $\hat{g}$ is found to factorize into two functions $\hat{g}'$ and $\hat{g}''$, only depending upon $\mathbf{u},\mathbf{u}^*$ or $\bar{\mathbf{z}},\bar{\mathbf{z}}^*$,

$$\hat{g}(\mathbf{u},\mathbf{u}^*,\bar{\mathbf{z}},\bar{\mathbf{z}}^*) = \hat{g}'(\mathbf{u},\mathbf{u}^*)\hat{g}''(\bar{\mathbf{z}},\bar{\mathbf{z}}^*). \tag{6.15}$$

The functions $\hat{g}'(\mathbf{u},\mathbf{u}^*)$ and $\hat{g}''(\bar{\mathbf{z}},\bar{\mathbf{z}}^*)$ are given by the following expressions:

$$\hat{g}'(\mathbf{u},\mathbf{u}^*) = \hat{A}'(\det \mathbf{U})^{-d-1} D^{[\lambda]}_{(\lambda)_{\min}(\lambda)_{\min}}(\mathbf{U}), \tag{6.16}$$

and

$$\hat{g}''(\bar{\mathbf{z}},\bar{\mathbf{z}}^*) = \hat{A}'' \prod_{i=1}^{d-1} [T_i(\bar{\mathbf{z}},\bar{\mathbf{z}}^*)]^{\beta_i}, \tag{6.17}$$

respectively. Here $\hat{A}'$ and $\hat{A}''$ are two constants satisfying the condition

$$\hat{A} = \hat{A}'\hat{A}'', \tag{6.18}$$

$D^{[\lambda]}_{(\lambda)_{\min}(\lambda)_{\min}}(\mathbf{U})$ is given by the right-hand side of Eq. (4.10) with $\mathbf{U}$ substituted for $\mathbf{W}$, the functions $T_i(\bar{\mathbf{z}},\bar{\mathbf{z}}^*)$ are obtained from Eqs. (4.24)–(4.26) by replacing $\mathbf{z}$ and $\mathbf{z}^*$ by $\bar{\mathbf{z}}$ and $\bar{\mathbf{z}}^*$, and the exponents β_i are the same as in Eq. (6.10).

A straightforward analysis, similar to that carried out in Appendix B, shows that provided the constant $\hat{A}''$ is chosen in such a way that

$$\int d\mathbf{z}\, d\mathbf{z}^*\, \hat{g}''(\mathbf{z},\mathbf{z}^*) = 1, \tag{6.19}$$

the function $\hat{g}''(\bar{\mathbf{z}},\bar{\mathbf{z}}^*)$ is the weight function of U(d) CS, written in terms of $\bar{\mathbf{z}},\bar{\mathbf{z}}^*$ instead of $\mathbf{z},\mathbf{z}^*$. This means that $\hat{g}''(\mathbf{z},\mathbf{z}^*)$ satisfies the unity resolution relation

$$\int d\mathbf{z}\, d\mathbf{z}^*\, \hat{g}''(\mathbf{z},\mathbf{z}^*)|\mathbf{z}\rangle\langle\mathbf{z}| = I_{[\lambda]} \tag{6.20}$$

in the representation space of the U(d) irrep $[\lambda]$. The constant $\hat{A}''$ satisfying Eq. (6.19) is easily found to be given by

$$\hat{A}'' = \pi^{-1}(\lambda_1 - \lambda_2 + 1) \quad \text{if } \lambda_1 > \lambda_2, \tag{6.21}$$

and

$$\begin{aligned}\hat{A}'' &= \pi^{-3}(\lambda_1-\lambda_2+1)(\lambda_1-\lambda_3+2)(\lambda_2-\lambda_3+1)\\ &\quad \text{if } \lambda_1>\lambda_2>\lambda_3,\\ &= \pi^{-2}(\lambda_1-\lambda_2+1)(\lambda_1-\lambda_2+2) \quad \text{if } \lambda_1>\lambda_2=\lambda_3,\\ &= \pi^{-2}(\lambda_1-\lambda_3+1)(\lambda_1-\lambda_3+2) \quad \text{if } \lambda_1=\lambda_2>\lambda_3,\end{aligned} \tag{6.22}$$

for U(2) and U(3) irreps, respectively.

It now remains to determine the normalization coefficient $\hat{A}'$ of $\hat{g}'(\mathbf{u},\mathbf{u}^*)$. From Eqs. (6.4), (6.14), (6.15), (6.18), and (6.19), it has to satisfy the following condition:

$$\int d\mathbf{u}\, d\mathbf{u}^*\, \hat{g}'(\mathbf{u},\mathbf{u}^*) = 1. \tag{6.23}$$

For $d \leqslant 3$, a straightforward calculation, using the methods of Ref. 46 leads to the following result:

$$\hat{A}' = 2^{-d}\pi^{-d(d+1)/2} \prod_{i<j=1}^{d} (\lambda_i + \lambda_j + n - i - j). \tag{6.24}$$

This completes the determination of the Sp($2d,R$) CS measure for $d \leqslant 3$.

VII. HILBERT SPACE OF ANALYTIC FUNCTIONS

The purpose of the present section is both to summarize the results achieved in the previous sections, and to reformulate them in the language of Hilbert spaces of analytic functions.

The Sp($2d,R$) CS $|\mathbf{u},\mathbf{z}\rangle$, labeled by a complex symmetric matrix $\mathbf{u}$ fulfilling the condition $\mathbf{I} - \mathbf{u}^*\mathbf{u} > 0$, and by a set of complex parameters $\mathbf{z}$, chosen as explained in Sec. III and varying in the whole complex plane, do form a nonorthogonal family of states, satisfying a unity resolution relation in the representation space $\mathscr{F}_{(\lambda)}$ of the irrep $\langle\lambda\rangle$. Hence this set is complete (as a matter of fact, it is overcomplete as proved hereafter), and can be used as a continuous basis in $\mathscr{F}_{(\lambda)}$. Any state $|\psi\rangle$ of $\mathscr{F}_{(\lambda)}$ can be expanded in terms of CS as follows:

$$|\psi\rangle = \int d\hat{\sigma}(\mathbf{u},\mathbf{z})\hat{\psi}(\mathbf{u},\mathbf{z})|\mathbf{u},\mathbf{z}\rangle, \tag{7.1}$$

where

$$\hat{\psi}(\mathbf{u},\mathbf{z}) = \langle\mathbf{u},\mathbf{z}|\psi\rangle \tag{7.2}$$

is an analytic function in the variables $\mathbf{u}$, and a polynomial in the variables $\mathbf{z}$.

In particular, the functional images

$$\hat{\phi}_{\mathbf{N}(\lambda)}(\mathbf{u},\mathbf{z}) = \langle\mathbf{u},\mathbf{z}|\mathbf{N};(\lambda)\rangle \tag{7.3}$$

of the discrete basis states $|\mathbf{N};(\lambda)\rangle$, defined in Eq. (2.6), are polynomials in $\mathbf{u}$ and $\mathbf{z}$, whose explicit form can be found as follows. We first note that from Eq. (6.5), the functional image of the irrep lowest weight state is

$$\hat{\phi}_{\mathbf{0}(\lambda)_{\min}}(\mathbf{u},\mathbf{z}) = 1. \tag{7.4}$$

We next observe that any basis state $|(\lambda)\rangle$ of the U(d) irrep $[\lambda]$ can be built from its lowest weight state by applying an appropriate polynomial in the U(d) generators,

$$|(\lambda)\rangle = P_{(\lambda)}(\mathbf{E})|(\lambda)_{\min}\rangle. \tag{7.5}$$

Hence, its functional image can be obtained by acting with the differential operator $P_{(\lambda)}(\mathring{\mathscr{E}})$ on the function $\hat{\phi}_{\mathbf{0}(\lambda)_{\min}}(\mathbf{u},\mathbf{z})$,

$$\hat{\phi}_{\mathbf{0}(\lambda)}(\mathbf{u},\mathbf{z}) = P_{(\lambda)}(\mathring{\mathscr{E}}). \tag{7.6}$$

The result is a polynomial in $\mathbf{z}$,

$$\hat{\phi}_{\mathbf{0}(\lambda)}(\mathbf{u},\mathbf{z}) = \hat{\phi}_{(\lambda)}(\mathbf{z}) = \langle\mathbf{z}|(\lambda)\rangle, \tag{7.7}$$

that is nothing else than the U(d) CS representation of $|(\lambda)\rangle$. Explicit expressions of $\hat{\phi}_{(\lambda)}(\mathbf{z})$ will be given in paper II of the present series. Finally, the functional image of an arbitrary state $|\mathbf{N};(\lambda)\rangle$ can be obtained by acting with the differential operator $F_{\mathbf{N}}(\hat{\mathscr{D}}^\dagger)$ on the polynomial $\hat{\phi}_{(\lambda)}(\mathbf{z})$,

$$\hat{\phi}_{\mathbf{N}(\lambda)}(\mathbf{u},\mathbf{z}) = F_{\mathbf{N}}(\hat{\mathscr{D}}^\dagger)\hat{\phi}_{(\lambda)}(\mathbf{z}). \tag{7.8}$$

From Eq. (5.7b), it is obvious that the resulting function is a

polynomial in $\mathbf{u}$ and $\mathbf{z}$, of degree $\Sigma_{i<j}N_{ij}$ in $\mathbf{u}$. Its explicit form is however, quite complicated, hence of no practical use. This will contrast with the corresponding result for the annihilation-operator CS to be derived in paper II.

The space $\hat{\mathscr{H}}_{(\lambda)}$ spanned by the analytic functions $\hat{\psi}(\mathbf{u},\mathbf{z})$ corresponding to the states $|\psi\rangle$ of $\mathscr{F}_{(\lambda)}$, can be endowed with the scalar product

$$\langle\hat{\chi},\hat{\psi}\rangle = \int d\hat{\sigma}(\mathbf{u},\mathbf{z})[\hat{\chi}(\mathbf{u},\mathbf{z})]^*\hat{\psi}(\mathbf{u},\mathbf{z}), \tag{7.9}$$

making it into a Hilbert space of analytic functions. From Eqs. (7.1) and (7.2), it is obvious that the scalar product is preserved in the mapping $|\psi\rangle\mapsto\hat{\psi}(\mathbf{u},\mathbf{z})$, i.e.,

$$\langle\chi|\psi\rangle = \langle\hat{\chi},\hat{\psi}\rangle, \tag{7.10}$$

and that the CS overlap $\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)$, defined in Eq. (4.1), acts as a reproducing kernel in $\hat{\mathscr{H}}_{(\lambda)}$, i.e.,

$$\hat{\psi}(\mathbf{u},\mathbf{z}) = \int d\hat{\sigma}(\mathbf{u},\mathbf{z})\hat{K}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*)\hat{\psi}(\mathbf{u}',\mathbf{z}'). \tag{7.11}$$

Moreover, any CS can be expanded into CS as follows:

$$|\mathbf{u},\mathbf{z}\rangle = \int d\hat{\sigma}(\mathbf{u}',\mathbf{z}')\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)|\mathbf{u}',\mathbf{z}'\rangle; \tag{7.12}$$

therefore the CS are not independent and form an overcomplete set.

Let X be any operator acting in $\mathscr{F}_{(\lambda)}$, and $\hat{\mathscr{X}}$ the corresponding partial differential operator in $\hat{\mathscr{H}}_{(\lambda)}$. Then X maps any state $|\psi\rangle$ in $\mathscr{F}_{(\lambda)}$ onto a state $|\chi\rangle$ in the same space,

$$|\chi\rangle = X|\psi\rangle, \tag{7.13}$$

and the corresponding functions in $\hat{\mathscr{H}}_{(\lambda)}$ satisfy the following relation:

$$\hat{\chi}(\mathbf{u},\mathbf{z}) = \hat{\mathscr{X}}\hat{\psi}(\mathbf{u},\mathbf{z}). \tag{7.14}$$

The function $\hat{\chi}(\mathbf{u},\mathbf{z})$ can also be obtained from $\hat{\psi}(\mathbf{u},\mathbf{z})$ by the integral transform

$$\hat{\chi}(\mathbf{u},\mathbf{z}) = \int d\hat{\sigma}(\mathbf{u}',\mathbf{z}')\hat{X}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*)\hat{\psi}(\mathbf{u}',\mathbf{z}'), \tag{7.15}$$

with the integral kernel

$$\hat{X}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*) = \langle\mathbf{u},\mathbf{z}|X|\mathbf{u}',\mathbf{z}'\rangle. \tag{7.16}$$

Equation (7.15) directly results from Eq. (7.13) when use is made of the unity resolution relation (6.1). The relation between the integral kernel $\hat{X}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*)$ and the differential operator $\hat{\mathscr{X}}$ reads

$$\hat{X}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*) = \hat{\mathscr{X}}\hat{K}(\mathbf{u},\mathbf{z};\mathbf{u}'^*,\mathbf{z}'^*). \tag{7.17}$$

Since the mapping from $\mathscr{F}_{(\lambda)}$ into $\hat{\mathscr{H}}_{(\lambda)}$ preserves the scalar product, the Hermiticity properties of the operators X in $\mathscr{F}_{(\lambda)}$ are transferred to their images $\hat{\mathscr{X}}$ in $\hat{\mathscr{H}}_{(\lambda)}$. In particular, for the Sp($2d,R$) generators the following relations are satisfied:

$$\langle\hat{\chi},\hat{\mathscr{E}}_{ij}\hat{\psi}\rangle = \langle\hat{\mathscr{E}}_{ji}\hat{\chi},\hat{\psi}\rangle, \tag{7.18a}$$

$$\langle\hat{\chi},\hat{\mathscr{D}}^{\dagger}_{ij}\hat{\psi}\rangle = \langle\hat{\mathscr{D}}_{ij}\hat{\chi},\hat{\psi}\rangle. \tag{7.18b}$$

It can indeed be checked that Eqs. (7.18a) and (7.18b) are, respectively, equivalent to Eqs. (6.7a) and (6.7b), which were used to determine the measure $d\hat{\sigma}(\mathbf{u},\mathbf{z})$.

From Eqs. (4.14), (5.3), (5.7), and (7.17), the integral kernels corresponding to the Sp($2d,R$) generators can be easily calculated by noting that

$$\Delta_{u'_{ij}}\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) = \hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)\Delta_{u'_{ij}}\ln\hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*), \tag{7.19}$$

and

$$\ln\det\mathbf{U} = \operatorname{tr}\ln\mathbf{U} = -\sum_{m=1}^{\infty} m^{-1}\operatorname{tr}(\mathbf{u}^*\mathbf{u}')^m. \tag{7.20}$$

The results read

$$\begin{aligned}\hat{D}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) &= \hat{D}^{\dagger}_{ij}(\mathbf{u}^*,\mathbf{z}^*;\mathbf{u}',\mathbf{z}') \\ &= \hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)\Big\{(2\lambda_1+n)(\mathbf{U}^{-1}\mathbf{u}^*)_{ij} \\ &\quad + \sum_{k=1}^{d-1}(\lambda_k-\lambda_{k+1})[T_k(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)]^{-1} \\ &\quad \times\Delta_{u'_{ij}}T_k(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)\Big\},\end{aligned} \tag{7.21a}$$

and

$$\begin{aligned}\hat{E}_{ij}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*) &= \hat{K}(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)\Big\{(2\lambda_1+n)(\mathbf{U}^{-1}_{ji}-\tfrac{1}{2}\delta_{ij})+f'_{ij} \\ &\quad + \sum_{k=1}^{d-1}(\lambda_k-\lambda_{k+1})[T_k(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)]^{-1} \\ &\quad \times[(\mathbf{u}'\Delta_{u'})_{ij}+F'_{ij}]T_k(\mathbf{u}',\mathbf{z}';\mathbf{u}^*,\mathbf{z}^*)\Big\},\end{aligned} \tag{7.21b}$$

where F_{ij} and f_{ij} are, respectively, the operational and functional terms of $\hat{\mathscr{E}}_{ij}$, and the primed symbols denote operators or functions depending upon $\mathbf{u}'$ and $\mathbf{z}'$ instead of $\mathbf{u}$ and $\mathbf{z}$. In Eq. (7.21), the first term between braces coincides with the result obtained by Kramer[21] in the case where all the λ_i's are equal, while the summation over k can be easily evaluated by introducing the explicit expressions of the T_k's, and by taking into account that

$$\Delta_{u'_{ij}}U_{kl} = -\delta_{il}u^*_{jk} - \delta_{jl}u^*_{ik}. \tag{7.22}$$

VIII. SOME APPLICATIONS

In this concluding section, we shall outline two possible applications of the formalism developed in the present paper. The first one is concerned with classical approximations to quantum dynamics, the second one with boson realizations of the symplectic algebra.

The path-integral formulation of Klauder[4,47,48] and the related method based upon the time-dependent variational principle[36] make use of continuous and overcomplete sets of Hilbert space vectors. Among those, the families of generalized CS play an important role. The complex parameters v_α, characterizing such overcomplete sets of quantum states, serve as classical coordinates in a generalized phase space, where the motion is described classically by Hamilton-like equations. The latter are written in terms of a generalized Poisson bracket

$$\{f(\mathbf{v},\mathbf{v}^*),g(\mathbf{v},\mathbf{v}^*)\} = \sum_{\alpha\beta}(\hat{\gamma}^{-1})_{\alpha\beta}\left(\frac{\partial f}{\partial v^*_\alpha}\frac{\partial g}{\partial v_\beta}-\frac{\partial f}{\partial v_\beta}\frac{\partial g}{\partial v^*_\alpha}\right). \tag{8.1}$$

In Eq. (8.1), $\hat{\gamma}^{-1}$ is the inverse of the matrix $\hat{\gamma} = \|\hat{\gamma}_{\alpha\beta}\|$, whose elements

$$\hat{\gamma}_{\alpha\beta} = \frac{\partial^2}{\partial v_\alpha\,\partial v^*_\beta}\ln\hat{K}(\mathbf{v},\mathbf{v}^*) \tag{8.2}$$

are defined in terms of the norm $\hat{K}(\mathbf{v},\mathbf{v}^*)$ of the quantum states.

If we choose the Sp(2d,R) CS$|\mathbf{u},\mathbf{z}\rangle$ for the continuous and overcomplete set of states, then the classical coordinates v_α are the variables u_{ij}, $i\leqslant j$, and $\mathbf{z}$. It is straightforward to determine the matrix $\hat{\gamma}$ from the known expression of the overlap, Eq. (4.14), then to invert it to get an explicit expression for the generalized Poisson bracket. In the case where $\lambda_1=\cdots=\lambda_d=\lambda$, the result reads[21]

$$\{f(\mathbf{u},\mathbf{u}^*),g(\mathbf{u},\mathbf{u}^*)\}=[4(2\lambda+n)]^{-1}\sum_{ijkl}(U_{ik}U_{jl}+U_{il}U_{jk})\times\left[(\Delta_{u_{ij}^*}f)(\Delta_{u_{kl}}g)-(\Delta_{u_{kl}}f)(\Delta_{u_{ij}^*}g)\right], \tag{8.3}$$

while in the remaining cases, it assumes much more complicated forms, whose discussion we leave to a forthcoming publication. The functions

$$\hat{\mathfrak{D}}_{ij}^{\dagger}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*),\quad \hat{\mathfrak{D}}_{ij}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*),$$

and

$$\hat{\mathfrak{E}}_{ij}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*),$$

defined by

$$\hat{\mathfrak{D}}_{ij}^{\dagger}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*)=[\hat{K}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*)]^{-1}\hat{D}_{ij}^{\dagger}(\mathbf{u},\mathbf{z};\mathbf{u}^*,\mathbf{z}^*) \tag{8.4}$$

and similar relations for the remaining ones, yield a realization of the Sp(2d,R) Lie algebra in terms of the generalized Poisson bracket. Their detailed expressions can be deduced from Eq. (7.21). The application of this formalism to the Sp(6,R) nuclear collective model provides a classical approximation of the theory. Numerical applications along these lines are in progress.[22]

Knowing the classical solution, one then has to quantize periodic motions.[16,48] In the calculation of the quantum corrections, boson degrees of freedom appear naturally since the bosons are just the quantum version of the classical parameters labeling the quantum states of the overcomplete set.[48] Because of this, and because they are quite useful for studying the relations between microscopic and phenomenological models, the boson realizations of the Sp(2d,R) algebra have been the object of many studies in the last few years.[9,15,18,20,21,23,24,26–28] By boson realizations we mean both non-Hermitian Dyson[49] and Hermitian Holstein–Primakoff[50] realizations.

In Refs. 27 and 28, the relations between various PCS representations and boson realizations, as well as those between the latter have been extensively studied for generic irreps $\langle\lambda\rangle$ of Sp(2d,R). They are directly relevant to establishing relations between boson realizations and (fully) CS representations, such as that derived in the present paper. From Eqs. (3.4) and (3.6), it is indeed obvious that the unitary-operator CS, introduced in the present paper, can be expanded in terms of the so-called Perelomov PCS of Refs. 27 and 28, as follows:

$$|\mathbf{u},\mathbf{z}\rangle=\sum_{(\lambda)}[\hat{\phi}_{(\lambda)}(\mathbf{z})]^*|\mathbf{u};(\lambda)\rangle, \tag{8.5}$$

the coefficients of the expansion being well-known functions of $\mathbf{z}^*$, defined in Eq. (7.7).

According to Refs. 27 and 28, boson realizations of the Sp(2d,R) algebra can be obtained by mapping the Sp(2d,R) representation space $\mathscr{F}_{(\lambda)}$ onto the direct product space $\mathscr{B}\otimes\mathscr{S}_{[\lambda]}$, where $\mathscr{B}$ is the infinite-dimensional space spanned by all the boson states

$$|\mathbf{N}]=\prod_{i\leqslant j}(N_{ij}!)^{-1/2}[(1+\delta_{ij})^{-1/2}\bar{a}_{ij}^{\dagger}]^{N_{ij}}|0], \tag{8.6}$$

built from $d(d+1)/2$ independent boson creation operators $\bar{a}_{ij}^{\dagger}=\bar{a}_{ji}^{\dagger}$, $i,j=1,\ldots,d$, acting upon the vacuum state $|0]$, and $\mathscr{S}_{[\lambda]}$ is the representation space of a U(d) irrep $[\lambda]$, whose bases are denoted by $|(\lambda)]$. If we choose to map the discrete basis states $|\mathbf{N};(\lambda)\rangle$ or $|([l][\lambda])\alpha[h]\beta(k)q\rangle$, defined in Eqs. (2.6) or (2.9), onto the extended boson states

$$|\mathbf{N};(\lambda)]=|\mathbf{N}]\otimes|(\lambda)]=\mathrm{F}_{\mathbf{N}}(\bar{\mathbf{a}}^{\dagger})|\mathbf{0};(\lambda)], \tag{8.7}$$

or

$$|([l][\lambda])\alpha[h]\beta(k)q]=[P_{[l]}(\bar{\mathbf{a}}^{\dagger})\times|\mathbf{0};(\ )]]_{\beta(k)q}^{\alpha[h]}, \tag{8.8}$$

where

$$|\mathbf{0};(\lambda)]=|0]\otimes|(\lambda)], \tag{8.9}$$

and $F_{\mathbf{N}}(\bar{\mathbf{a}}^{\dagger})$ and $P_{[l]}(\bar{\mathbf{a}}^{\dagger})$ are the same polynomials as $F_{\mathbf{N}}(\mathbf{D}^{\dagger})$ and $P_{[l]}(\mathbf{D}^{\dagger})$ but with $D_{ij}^{\dagger}$ replaced by $\bar{a}_{ij}^{\dagger}$, then the PCS $|\mathbf{u};(\lambda)\rangle$ and the CS $|\mathbf{u},\mathbf{z}\rangle$ are mapped, respectively, onto the extended Glauber CS

$$|\mathbf{u};(\lambda)]=|\mathbf{u}]\otimes|(\lambda)], \tag{8.10}$$

and the CS of the direct product group $N(d)\otimes U(d)$,

$$|\mathbf{u},\mathbf{z}]=|\mathbf{u}]\otimes|\mathbf{z}]=\sum_{(\lambda)}[\hat{\phi}_{(\lambda)}(\mathbf{z})]^*|\mathbf{u};(\lambda)]. \tag{8.11}$$

Here

$$|\mathbf{u}]=\exp(\tfrac{1}{2}\mathrm{tr}\,\mathbf{u}^*\bar{\mathbf{a}}^{\dagger})|0] \tag{8.12}$$

is a standard Glauber CS (see Ref. 2), $|\mathbf{z}]$ is a U(d) CS belonging to the space $\mathscr{S}_{[\lambda]}$, and we have used the fact that

$$[\mathbf{z}|(\lambda)]=\langle\mathbf{z}|(\lambda)\rangle=\hat{\phi}_{(\lambda)}(\mathbf{z}). \tag{8.13}$$

In this mapping, both the PCS and the CS representations of the Sp(2d,R) algebra are converted into a Dyson realization of the latter by the replacement of u_{ij} and $\Delta_{u_{ij}}$ by $\bar{a}_{ij}^{\dagger}$ and $\bar{a}_{ij}$, respectively.

In Ref. 28, it was also shown that the transition from the Dyson realization $X_{\mathscr{D}}$ of any Sp(2d,R) generator X to its HP realization X_{HP} can be performed through a similarity transformation as follows:

$$X_{\mathrm{HP}}=T^{-1/2}X_{\mathscr{D}}T^{1/2}, \tag{8.14}$$

where $T^{1/2}$ is the square root of the positive definite, Hermitian operator T, defined by either relation

$$[\mathbf{N}';(\lambda')|T|\mathbf{N};(\lambda)]=\langle\mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle, \tag{8.15}$$

or

$$[([l'][\lambda])\alpha'[h]\beta'(k)q|T|([l][\lambda])\alpha[h]\beta(k)q]=\langle([l'][\lambda])\alpha'[h]\beta'(k)q|([l][\lambda])\alpha[h]\beta(k)q\rangle. \tag{8.16}$$

The operator T is therefore an essential ingredient of boson realizations, whose determination is of considerable interest.

For such purpose, one can start from the equations satisfied by T,[28,51]

$$[T,\bar{\mathbf{a}}^{\dagger}\bar{\mathbf{a}}+\mathring{\mathbf{E}}]=0, \tag{8.17}$$

and

$$T\bar{\mathbf{a}}^{\dagger}T^{-1}=\bar{\mathbf{a}}^{\dagger}\tilde{\mathring{\mathbf{E}}}+\mathring{\mathbf{E}}\bar{\mathbf{a}}^{\dagger}+[\bar{\mathbf{a}}^{\dagger}\bar{\mathbf{a}}-(d+1)\mathbf{I}]\bar{\mathbf{a}}^{\dagger}, \tag{8.18}$$

where $\mathring{\mathbf{E}}$ denotes the matrix of $U(d)$ generators acting in $\mathscr{S}_{[\lambda]}$. By multiplying Eq. (8.18) from the right by T, and by taking matrix elements of the resulting equation between two coupled states (8.8), one gets recursion relations for the matrix elements of T between coupled states.[15,20] In Ref. 20, an explicit algorithm was devised to solve them in the case of Sp(4,R).

In a recent work, Castaños, Kramer, and Moshinsky[52] did propose an alternative to this cumbersome recursive method. For simplicity's sake, their discussion was restricted to Sp(4,R), but it is equally valid for higher-dimensional symplectic groups, as we shall now proceed to show.

We first note that by expanding into powers the exponentials in Eqs. (3.4) and (8.12), we obtain a development of the PCS and of the extended Glauber CS, respectively, on the discrete basis $|\mathbf{N};(\lambda)\rangle$ and $|\mathbf{N};(\lambda)]$,

$$|\mathbf{u};(\lambda)\rangle=\sum_{\mathbf{N}}F_{\mathbf{N}}(\mathbf{u}^{*})|\mathbf{N};(\lambda)\rangle, \tag{8.19}$$

$$|\mathbf{u};(\lambda)]=\sum_{\mathbf{N}}F_{\mathbf{N}}(\mathbf{u}^{*})|\mathbf{N};(\lambda)]. \tag{8.20}$$

The expansion coefficient $F_{\mathbf{N}}(\mathbf{u}^{*})$ is the same in both cases, and it is obtained from Eq. (2.7) by replacing $D^{\dagger}_{ij}$ by u^{*}_{ij}. The definition of the operator T, given in Eq. (8.15), then implies that

$$[\mathbf{u}';(\lambda')|T|\mathbf{u};(\lambda)]=\langle\mathbf{u}';(\lambda')|\mathbf{u};(\lambda)\rangle. \tag{8.21}$$

Finally, by using Eqs. (8.5) and (8.11), Eq. (8.21) can be transformed into the following relation:

$$[\mathbf{u}',\mathbf{z}'|T|\mathbf{u},\mathbf{z}]=\langle\mathbf{u}',\mathbf{z}'|\mathbf{u},\mathbf{z}\rangle. \tag{8.22}$$

Note that for the right-hand side of Eqs. (8.21) and (8.22) to be meaningful, one has to restrict oneself to matrices $\mathbf{u}$ and $\mathbf{u}'$ satisfying the conditions $\mathbf{I}-\mathbf{u}^{*}\mathbf{u}>0$, and $\mathbf{I}-\mathbf{u}'^{*}\mathbf{u}'>0$, although the boson CS appearing on the left-hand side of these equations could of course be defined for any symmetric matrices $\mathbf{u}$ and $\mathbf{u}'$.

Equation (8.22) was obtained for the first time by Castaños, Kramer, and Moshinsky in the Sp(4,R) case.[52] To prove it, they used the property that Glauber CS lead to the Bargmann representation,[53] wherein $\bar{\mathbf{a}}^{\dagger}$ and $\bar{\mathbf{a}}$ are represented by $\mathbf{u}$ and Δ_{u}, respectively. They then derived the partial differential equations satisfied by the kernel $[\mathbf{u}',\mathbf{z}'|T|\mathbf{u},\mathbf{z}]$ as a consequence of Eq. (8.18). By comparing these equations with those satisfied by the Sp($2d$,R) CS overlap, given in Eq. (5.13), they finally concluded that Eq. (8.22) is fulfilled. We would like to emphasize that such a complicated demonstration is unnecessary and would become, moreover, quite cumbersome for higher-dimensional symplectic groups. As proved above, Eq. (8.22) indeed directly follows from the definition of the operator T, as given in Ref. 28.

Equation (8.22) expresses the operator T as a kernel with respect to the CS of the direct product group $N(d)\otimes U(d)$, and shows that this kernel coincides with the overlap of the corresponding Sp($2d$,R) CS. Since a closed, analytic expression of the latter has been given in Eq. (4.14), a significant advance has been made in the determination of T. It now only remains to expand the kernel in terms of discrete basis states to obtain the matrix representation of T without going through the solution of recursion relations. Such a procedure has been applied to Sp(4,R) (see Ref. 54) and hopefully will be extended to Sp(6,R) in the near future.

APPENDIX A: EXPLICIT FORM OF EQ. (6.9)

In this Appendix, we list the explicit expressions of the partial differential operators A_{ij}, B_{ij}, and of the functions a_{ij}, b_{ij}, appearing in Eq. (6.9), satisfied by the weight function $\hat{f}(\mathbf{u},\mathbf{u}^{*},\mathbf{z},\mathbf{z}^{*})$.

In the Sp(4,R) case, they are given by

$$\begin{aligned}&A_{11}=-A_{22}=z\partial-z^{*}\partial^{*},\quad A_{12}=-z^{2}\partial-\partial^{*},\\&B_{11}=2(u_{11}-u_{12}z)z\partial,\quad B_{22}=2(u_{12}-u_{22}z)\partial,\\&B_{12}=(u_{11}-u_{22}z^{2})\partial,\end{aligned} \tag{A1}$$

and

$$\begin{aligned}&a_{11}=a_{22}=0,\quad a_{12}=-(\lambda_{1}-\lambda_{2}+2)z,\\&b_{11}=-(2\lambda_{2}+n-8)u_{11}-2(\lambda_{1}-\lambda_{2}+2)u_{12}z,\\&b_{22}=-(2\lambda_{1}+n-4)u_{22},\\&b_{12}=-(\lambda_{1}+\lambda_{2}+n-6)u_{12}-(\lambda_{1}-\lambda_{2}+2)u_{22}z.\end{aligned} \tag{A2}$$

In the Sp(6,R) case, we have to distinguish between cases (a), (b), and (c) irreps. As an example, we list below the results obtained in the x parametrization for case (a) irreps:

$$\begin{aligned}A_{11}&=x_{2}\partial_{2}+x_{3}\partial_{3}-x_{2}^{*}\partial_{2}^{*}-x_{3}^{*}\partial_{3}^{*},\\A_{22}&=x_{1}\partial_{1}-x_{3}\partial_{3}-x_{1}^{*}\partial_{1}^{*}+x_{3}^{*}\partial_{3}^{*},\\A_{33}&=-x_{1}\partial_{1}-x_{2}\partial_{2}+x_{1}^{*}\partial_{1}^{*}+x_{2}^{*}\partial_{2}^{*},\\A_{12}&=x_{2}\partial_{1}-x_{3}^{2}\partial_{3}-x_{1}^{*}\partial_{2}^{*}-\partial_{3}^{*},\\A_{13}&=-x_{2}(x_{1}\partial_{1}+x_{2}\partial_{2})+(x_{1}x_{3}-x_{2})x_{3}\partial_{3}-\partial_{2}^{*},\\A_{23}&=-x_{1}(x_{1}\partial_{1}+x_{2}\partial_{2})+(x_{1}x_{3}-x_{2})\partial_{3}-\partial_{1}^{*},\\B_{11}&=2u_{11}(x_{2}\partial_{2}+x_{3}\partial_{3})+2u_{12}(x_{2}\partial_{1}-x_{3}^{2}\partial_{3})\\&\quad+2u_{13}[-x_{2}(x_{1}\partial_{1}+x_{2}\partial_{2})+(x_{1}x_{3}-x_{2})x_{3}\partial_{3}],\\B_{22}&=2u_{12}(x_{1}\partial_{2}+\partial_{3})+2u_{22}(x_{1}\partial_{1}-x_{3}\partial_{3})\\&\quad+2u_{23}[-x_{1}(x_{1}\partial_{1}+x_{2}\partial_{2})+(x_{1}x_{3}-x_{2})\partial_{3}],\\B_{33}&=2u_{13}\partial_{2}+2u_{23}\partial_{1}-2u_{33}(x_{1}\partial_{1}+x_{2}\partial_{2}),\\B_{12}&=u_{11}(x_{1}\partial_{2}+\partial_{3})+u_{12}(x_{1}\partial_{1}+x_{2}\partial_{2})\\&\quad+u_{13}[-x_{1}(x_{1}\partial_{1}+x_{2}\partial_{2})+(x_{1}x_{3}-x_{2})\partial_{3}]\\&\quad+u_{22}(x_{2}\partial_{1}-x_{3}^{2}\partial_{3})+u_{23}[-x_{2}(x_{1}\partial_{1}+x_{2}\partial_{2})\\&\quad+(x_{1}x_{3}-x_{2})x_{3}\partial_{3}],\\B_{13}&=u_{11}\partial_{2}+u_{12}\partial_{1}+u_{13}(-x_{1}\partial_{1}+x_{3}\partial_{3})\\&\quad+u_{23}(x_{2}\partial_{1}-x_{3}^{2}\partial_{3})+u_{33}[-x_{2}(x_{1}\partial_{1}+x_{2}\partial_{2})\\&\quad+(x_{1}x_{3}-x_{2})x_{3}\partial_{3}],\\B_{23}&=u_{12}\partial_{2}+u_{13}(x_{1}\partial_{2}+\partial_{3})+u_{22}\partial_{1}\\&\quad-u_{23}(x_{2}\partial_{2}+x_{3}\partial_{3})+u_{33}[-x_{1}(x_{1}\partial_{1}+x_{2}\partial_{2})\\&\quad+(x_{1}x_{3}-x_{2})\partial_{3}],\end{aligned} \tag{A3}$$

and

$$
\begin{aligned}
&a_{11}=a_{22}=a_{33}=0, \quad a_{12}=-(\lambda_2-\lambda_3+2)x_3,\\
&a_{13}=-(\lambda_1-\lambda_3+4)x_2+(\lambda_2-\lambda_3+2)x_1x_3,\\
&a_{23}=-(\lambda_1-\lambda_2+2)x_1,\\
&b_{11}=-(2\lambda_3+n-12)u_{11}-2(\lambda_2-\lambda_3+2)u_{12}x_3\\
&\qquad+2u_{13}[-(\lambda_1-\lambda_3+4)x_2\\
&\qquad+(\lambda_2-\lambda_3+2)x_1x_3],\\
&b_{22}=-(2\lambda_2+n-8)u_{22}-2(\lambda_1-\lambda_2+2)u_{23}x_1,\\
&b_{33}=-(2\lambda_1+n-4)u_{33},\\
&b_{12}=-(\lambda_2+\lambda_3+n-10)u_{12}-(\lambda_1-\lambda_2+2)u_{13}x_1\\
&\qquad-(\lambda_2-\lambda_3+2)u_{22}x_3\\
&\qquad+u_{23}[-(\lambda_1-\lambda_3+4)x_2+(\lambda_2-\lambda_3+2)x_1x_3],\\
&b_{13}=-(\lambda_1+\lambda_3+n-8)u_{13}-(\lambda_2-\lambda_3+2)u_{23}x_3\\
&\qquad+u_{33}[-(\lambda_1-\lambda_3+4)x_2+(\lambda_2-\lambda_3+2)x_1x_3],\\
&b_{23}=-(\lambda_1+\lambda_2+n-6)u_{23}-(\lambda_1-\lambda_2+2)u_{33}x_1.
\end{aligned}
\tag{A4}
$$

For cases (b) and (c) irreps, the explicit expressions of A_{ij} and B_{ij} can be obtained from Eq. (A3), or its counterpart in the y parametrization, by deleting all the terms containing either $x_3, x_3^*, \partial_3, \partial_3^*$ or $y_1, y_1^*, \delta_1, \delta_1^*$. As a result of the integration by parts carried out in Eq. (6.7), such a simple trick cannot be used for a_{ij} and b_{ij}, which therefore have to be recalculated in each case.

APPENDIX B: SOLUTION OF EQ. (6.9)

The purpose of this appendix is to prove that the function $\hat{f}(\mathbf{u},\mathbf{u}^*, \mathbf{z},\mathbf{z}^*)$ defined in Eq. (6.10) is a solution of Eqs. (6.9a) and (6.9b), provided the exponents α and β_i, $i=1,\ldots, d-1$, take the values given in Eqs. (6.11) and (6.12).

By introducing Eq. (6.10) into Eqs. (6.9a) and (6.9b), we obtain the two following equations:

$$
\begin{aligned}
&\alpha\left(\prod_{l=1}^{d-1} T_l\right)[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}](\det \mathbf{U})\\
&\quad+(\det \mathbf{U})\sum_{l=1}^{d-1}\beta_l\left(\prod_{\substack{m=1\\ \neq l}}^{d-1} T_m\right)\\
&\quad\times[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}+A_{ij}]\,T_l\\
&\quad+a_{ij}(\det \mathbf{U})\prod_{l=1}^{d-1} T_l=0, \quad i\leqslant j,
\end{aligned}
\tag{B1a}
$$

and

$$
\begin{aligned}
&\alpha\left(\prod_{l=1}^{d-1} T_l\right)\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}\right]\det \mathbf{U}\\
&\quad+(\det \mathbf{U})\sum_{l=1}^{d-1}\beta_l\left(\prod_{\substack{m=1\\ \neq l}}^{d-1} T_m\right)\\
&\quad\times\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}+B_{ij}\right]T_l\\
&\quad+b_{ij}(\det \mathbf{U})\prod_{l=1}^{d-1} T_l=0, \quad i\leqslant j.
\end{aligned}
\tag{B1b}
$$

In these relations, the action of the differential operators on det U can be easily evaluated as follows. We note that in the case where $\lambda_1=\cdots=\lambda_d=0$, from Eq. (4.21) the CS overlap (where we set $\mathbf{u}'=\mathbf{u}$) reduces to

$$
\hat{K}(\mathbf{u},\mathbf{u}^*)=(\det \mathbf{U})^{-n/2}. \tag{B2}
$$

On the other hand, it fulfills the differential equations (5.12), where $\mathring{\mathscr{E}}=(n/2)\mathbf{I}$. The function det U therefore satisfies the equations

$$
[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}]\det \mathbf{U}=0, \tag{B3a}
$$

and

$$
\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}-2u_{ij}\right]\det \mathbf{U}=0. \tag{B3b}
$$

By taking Eqs. (B3a) and (B3b) into account, Eqs. (B1a) and (B1b) become

$$
\begin{aligned}
&\sum_{l=1}^{d-1}\beta_l\left(\prod_{\substack{m=1\\ \neq l}}^{d-1} T_m\right)[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}+A_{ij}]\,T_l\\
&\quad+a_{ij}\prod_{l=1}^{d-1} T_l=0, \quad i\leqslant j,
\end{aligned}
\tag{B4a}
$$

and

$$
\begin{aligned}
&\sum_{l=1}^{d-1}\beta_l\left(\prod_{\substack{m=1\\ \neq l}}^{d-1} T_m\right)\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}+B_{ij}\right]T_l\\
&\quad+(b_{ij}+2\alpha u_{ij})\prod_{l=1}^{d-1} T_l=0, \quad i\leqslant j.
\end{aligned}
\tag{B4b}
$$

To evaluate the action on T_l of the differential operators appearing in Eqs. (B4a) and (B4b), we could again use some information coming from the differential equations satisfied by the overlap. In this case, it is easier, however, to directly apply the differential operators on T_l. For such a purpose, the following properties are worth noting:

$$
\begin{aligned}
&[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}]\,U_{lm}=-\delta_{il}\,U_{jm}+\delta_{jm}\,U_{li},\\
&[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}]\,U_{lm,pq}\\
&\quad=-\delta_{il}\,U_{jm,pq}-\delta_{im}\,U_{lj,pq}+\delta_{jp}\,U_{lm,iq}+\delta_{jq}\,U_{lm,pi},\\
&\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}\right]U_{lm}=u_{im}\,U_{lj}+u_{jm}\,U_{li},\\
&\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}\right]U_{lm,pq}\\
&\quad=u_{ip}\,U_{lm,jq}+u_{iq}\,U_{lm,pj}+u_{jp}\,U_{lm,iq}+u_{jq}\,U_{lm,pi},
\end{aligned}
\tag{B5}
$$

whose demonstration is straightforward.

In the Sp(4,R) case, taking Eqs. (A1) and (A2) into account, we obtain the following results:

$$
[(\mathbf{u}\Delta_u)_{ij}-(\mathbf{u}^*\Delta_{u^*})_{ji}+A_{ij}]\,T_1=\begin{cases}0, & \text{if } i=j=1,2,\\ -zT_1, & \text{if } i=1, j=2,\end{cases}
\tag{B6}
$$

and

$$
\left[\sum_k u_{ik}(\mathbf{u}\Delta_u)_{jk}-\Delta_{u_{ij}^*}+B_{ij}\right]T_1
=\begin{cases}2(u_{11}-u_{12}z)T_1, & \text{if } i=j=1,\\ 0, & \text{if } i=j=2,\\ (u_{12}-u_{22}z)T_1, & \text{if } i=1, j=2.\end{cases}
\tag{B7}
$$

The left-hand side of Eq. (B4a) therefore vanishes for

$i=j=1,2$. Hence, we are left with a single equation corresponding to $i=1, j=2$, which reads

$$-(\beta_1+\lambda_1-\lambda_2+2)T_1=0. \tag{B8}$$

It imposes that $\beta_1=-(\lambda_1-\lambda_2+2)$, in accordance with Eq. (6.11). On the other hand, Eq. (B4b) leads to the following equations:

$$\begin{aligned}
&[2\beta_1(u_{11}-u_{12}z)+(2\alpha-2\lambda_2-n+8)u_{11}\\
&\quad-2(\lambda_1-\lambda_2+2)u_{12}z]\,T_1=0,\\
&(2\alpha-2\lambda_1-n+4)u_{22}\,T_1=0,\\
&[\beta_1(u_{12}-u_{22}z)+(2\alpha-\lambda_1-\lambda_2-n+6)u_{12}\\
&\quad-(\lambda_1-\lambda_2+2)u_{22}z]\,T_1=0,
\end{aligned} \tag{B9}$$

corresponding to $i=j=1$, $i=j=2$ and $i=1, j=2$, respectively. With the above value of β_1, all three equations are satisfied if $\alpha=\lambda_1+n/2-2$, again in accordance with Eq. (6.11).

For Sp(6,R), the demonstration proceeds along the same lines and will not be detailed here.

[1]E. Schrödinger, Naturwiss. **14**, 664 (1926).
[2]R. J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[3]R. Gilmore and D. H. Feng, in *Progress in Particle and Nuclear Physics*, Vol. 9, *Collective Bands in Nuclei*, edited by D. Wilkinson (Pergamon, Oxford, 1983), p. 479.
[4]J. R. Klauder, J. Math. Phys. **4**, 1055, 1058 (1963); **5**, 177 (1964).
[5]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); Usp. Fiz. Nauk. **123**, 23 (1977) [Sov. Phys. Usp. **20**, 703 (1977)].
[6]R. Gilmore, Ann. Phys. N.Y. **74**, 391 (1972); Rev. Mex. Fís. **23**, 143 (1974).
[7]A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[8]J. Dobaczewski, Nucl. Phys. A **369**, 213, 237 (1981); **380**, 1 (1982).
[9]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. N.Y. **128**, 314 (1980); **131**, 1 (1981).
[10]G. Rosensteel and D. J. Rowe, Ann. Phys. N.Y. **123**, 36 (1979); **126**, 198, 343 (1980).
[11]O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. N.Y. **102**, 493 (1976); B. Buck, L. C. Biedenharn, and R. Y. Cusson, Nucl. Phys. A **317**, 205 (1979).
[12]G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979) [Sov. J. Nucl. Phys. **30**, 333 (1979)]; V. S. Vasilevskii, Yu. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980) [Sov. J. Nucl. Phys. **32**, 510 (1980)]; G. F. Filippov, V. I. Ovcharenko, and Yu. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (Naukova Dumka, Kiev, 1981) (in Russian).
[13]V. Vanagas, "The microscopic nuclear theory within the framework of the restricted dynamics," in *Lecture Notes in Physics* (Univ. Toronto, Toronto, 1977); V. Vanagas, "The microscopic theory of the collective motion in nuclei," in *Group Theory and Its Applications in Physics—1980, AIP Conference Proceedings*, No. 71, edited by T. H. Seligman (A.I.P., New York, 1981).
[14]J. P. Draayer, K. J. Weeks, and G. Rosensteel, Nucl. Phys. A **413**, 215 (1984); P. Park, J. Carvalho, M. Vassanji, D. J. Rowe, and G. Rosensteel, *ibid.* **414**, 93 (1984); G. Rosensteel, J. P. Draayer, and K. J. Weeks, *ibid.* **419**, 1 (1984); M. G. Vassanji and D. J. Rowe, *ibid.* **426**, 205 (1984).
[15]D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984).
[16]F. Arickx, Nucl. Phys. A **268**, 347 (1976); F. Arickx, J. Broeckhove, and E. Deumens, *ibid.* **318**, 269 (1979); **377**, 121 (1982); J. Broeckhove and P. Van Leuven, Phys. Rev. C **29**, 628 (1984); J. Broeckhove, M. Buysse, and P. Van Leuven, Phys. Lett. B **134**, 379 (1984).
[17]D. R. Peterson and K. T. Hecht, Nucl. Phys. A **344**, 361 (1980).
[18]O. Castaños, A. Frank, E. Chacón, P. Hess, and M. Moshinsky, Phys. Rev. C **25**, 1611 (1982); J. Math. Phys. **23**, 2537 (1982); O. Castaños and A. Frank, J. Math. Phys. **25**, 388 (1984).
[19]M. Moshinsky, J. Math. Phys. **25**, 1555 (1984); E. Chacón, P. Hess, and M. Moshinsky, *ibid.* **25**, 1565 (1984); O. Castaños, E. Chacón, and M. Moshinsky, *ibid.* **25**, 2815 (1984).
[20]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985).
[21]P. Kramer, Ann. Phys. N.Y. **141**, 254, 269 (1982); P. Kramer, Z. Papadopolos, and W. Schweizer, *ibid.* **149**, 44 (1983).
[22]P. Kramer, Z. Papadopolos, and W. Schweizer, Nucl. Phys. A **431**, 75 (1984); **441**, 461 (1985).
[23]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982).
[24]J. Deenen and C. Quesne, J. Phys. A: Math. Gen. **16**, 2095 (1983); in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 180, Istanbul, Turkey, 1982, edited by M. Serdaroğlu and E. Inönü (Springer, Berlin, 1983), p. 444.
[25]J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638 (1984).
[26]J. Deenen and C. Quesne, J. Phys. A: Math. Gen. **17**, L405 (1984).
[27]J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984).
[28]J. Deenen and C. Quesne, J. Math. Phys. **26**, 2705 (1985).
[29]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[30]A. U. Klimyk, J. Math. Phys. **24**, 224 (1983).
[31]P. Kramer and Z. Papadopolos, "Hilbert spaces of analytic functions and representations of the positive discrete series of Sp(6,R)," to be published.
[32]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[33]In Ref. 27, there is a misprint in Eq. (2.4c), which should read as Eq. (2.5c) in the present paper.
[34]I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[35]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974), p. 350.
[36]P. Kramer and M. Saraceno, *Geometry of the Time-Dependent Variational Principle in Quantum Mechanics* (Springer, Berlin, 1981).
[37]Reference 25 actually deals with the case where $\lambda_1=\cdots=\lambda_d=0$, but it is straightforward to extend the results to the case where $\lambda_1=\cdots=\lambda_d=\lambda\neq 0$ by replacing n by $n+2\lambda$ in all the formulas.
[38]J. M. Radcliffe, J. Phys. A: Gen. Phys. **4**, 313 (1971).
[39]F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[40]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1968).
[41]Reference 35, p. 401.
[42]Reference 35, p. 149.
[43]R. Gilmore, J. Math. Phys. **15**, 2090 (1974).
[44]J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[45]M. Brunet and T. H. Seligman, Rep. Math. Phys. **8**, 233 (1975).
[46]L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains, Translations of Mathematical Monographs*, Vol. 6 (Am. Math. Soc., Providence, RI, 1963).
[47]J. R. Klauder, Phys. Rev. D **19**, 2349 (1979).
[48]J. P. Blaizot and H. Orland, Phys. Rev. C **24**, 1740 (1981).
[49]F. J. Dyson, Phys. Rev. **102**, 1217 (1956).
[50]T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[51]Equations (8.17) and (8.18) were also derived in Refs. 15 and 20, where T is denoted by κ^2 and K^2, respectively.
[52]O. Castanõs, P. Kramer, and M. Moshinsky, J. Phys. A: Math. Gen. **18**, L493 (1985).
[53]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[54]O. Castaños, P. Kramer, and M. Moshinsky, "Boson realization of sp(4,R). II. The generating kernel formulation," to be published.

The basis for the symmetric irreducible representations of SO(9)

Z. R. Yu[a] and O. Scholten
National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan 48824-1321

H. Z. Sun[b]
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

(Received 26 March 1985; accepted for publication 3 October 1985)

An explicit basis is constructed for the symmetric irreducible representation (irrep) of $SU(9)\supset SO(9)\supset SO(5)\times SU_1(2)\times SU_2(2)$. It is also indicated how good angular momentum states can be constructed. The techniques used are based on the well-known tensor algebra for the infinitesimal generators of simple Lie groups.

I. INTRODUCTION

In nuclear physics, the interacting boson model (IBM) gives a unified description of collective states in medium and heavy nuclei. The original version of the IBM is limited to the consideration of s- and d-bosons.[1] However, for nuclei with large deformation the s–d model is insufficient and we have to introduce a g-boson degree of freedom to be able to account for the experimentally observed results.

Many authors have already identified the limiting symmetries of the interacting s–d–g boson system and obtain the typical spectra for the various limiting symmetries.[2] However, the explicit wave functions for the s–d–g system have not been given. In this paper, we discuss this problem. For the sake of simplicity, as a first step we will focus our attention on the construction of the basis vectors for the g-boson's system.

For the n g-bosons system, the symmetry group is $U(9)$. It is thus necessary to have nine quantum numbers for labeling the states uniquely. However, the physical subgroup chain $U(9)\supset SO(9)\supset SO(3)$, $SO(3)$ being the rotation group associated with angular momentum, only gives four quantum numbers. In this reduction there are several missing labels, for which it is extremely difficult (impossible) to find a simple physical interpretation. For this reason, we adopt the "natural" labeling in which an irrep of $SO(9)$ is considered to be fully reduced with respect to its subgroup $SO(5)\times SU_1(2)\times SU_2(2)$, and a state is labeled by the weight of the irrep of $SO(5)\times SU_1(2)\times SU_2(2)$ to which it belongs. This basis is called the mathematical basis. If the number of g-bosons are small, we can easily express the physical basis in terms of our mathematical basis.

In Sec. II, we will give a reduction formula for the irreps of $SO(5)\times SU_1(2)\times SU_2(2)$, which occur in a given symmetric irrep of $SO(9)$. In addition, we will also give an expression for the generators and Casimir operators of $SO(9)$. In Secs. III and IV, we will explicitly determine the basis states and give some examples.

II. GENERATORS

The branching rule of $SO(9)\supset SO(5)\times SU_1(2)\times SU_2(2)$ can be derived by means of the Schur function method.[3–5] The result can be written as follows:

$$\begin{array}{lll} SO(9)\supset & SO(5)\times SU_1(2)\times SU_2(2), \\ (v0000) & ((p,0),\Sigma,\Sigma), \end{array}$$

where v is the seniority quantum number, $2\Sigma = v-p-2\omega$, and

$$p = 0,1,2,\ldots,v, \tag{1a}$$

$$\omega = 0,1,2,\ldots,[(v-p)/2]. \tag{1b}$$

The symbol $[(v-p)/2]$ stands for the maximum integral of $(v-p)/2$.

In IBM, the groups $SU(9)$, $SO(9)$, and $SO(3)$ are generated by the following set of operators:

$$\begin{aligned} &SU(9):\ (g^\dagger\tilde{g})^{(k)}_q, && k=1,2,\ldots,8, \\ &&& q=0,\pm 1,\ldots,\pm k; \\ &SO(9):\ (g^\dagger\tilde{g})^{(k)}_q, && k=1,3,5,7, \\ &&& q=0,\pm 1,\ldots,\pm k; \\ &SO(3):\ (g^\dagger\tilde{g})^{(k)}_q, && k=1, \\ &&& q=0,\pm 1; \end{aligned} \tag{2}$$

where $g^\dagger_\mu(g_\mu)$, $\mu=0,\pm 1,\ldots,\pm 4$, is the creation (annihilation) operator of the g-boson, and $\tilde{g}_\mu = (-)^\mu g_{-\mu}$ is a tensor operator. A general state vector can be written as $|\phi\rangle = |nv\Omega\rangle$, where Ω is the totality of the quantum numbers needed for full classification besides n and v. In the following we will determine Ω by means of the group chain

$$SO(9)\supset SO(5)\times SU_1(2)\times SU_2(2), \tag{3a}$$

with

$$SO(5)\supset SU_a(2)\times SU_b(2). \tag{3b}$$

For convenience of construction of the state vector $|nv\Omega\rangle$, we introduce a set of uncoupling generators for the group chain (3), as follows:

$$\chi_{\mu\nu} = g^\dagger_\mu\tilde{g}_\nu - g^\dagger_\nu\tilde{g}_\mu, \quad \mu,\nu = 0,\pm 1,\ldots,\pm 4. \tag{4}$$

Obviously,

$$\chi_{\mu\mu}=0, \chi_{\mu\nu}=-\chi_{\nu\mu}, (\chi_{\mu\nu})^\dagger = (-)^{\mu+\nu}\chi_{-\nu-\mu},$$

and in addition they satisfy the following commutation relation:

$$\begin{aligned}[\chi_{\mu\nu},\chi_{\rho\sigma}] = {} & (-)^\nu\delta_{\rho-\nu}\chi_{\mu\sigma} + (-)^\sigma\delta_{\mu-\sigma}\chi_{\nu\rho} \\ & + (-)^\nu\delta_{\sigma-\nu}\chi_{\rho\mu} + (-)^\rho\delta_{\mu-\rho}\chi_{\sigma\nu}. \end{aligned} \tag{5}$$

The relation between $\chi_{\mu\nu}$ and the generators in Eq. (2) is

[a] On leave from Nanjing University of China, Nanjing, China.
[b] On leave from Beijing University of China, Beijing, China.

 0022-2488/86/020442-05$02.50 © 1986 American Institute of Physics

$$(g^{\dagger}\tilde{g})^{k}=\frac{1}{2}\sum_{\mu\nu}C^{kq}_{4\mu 4\nu}\chi_{\mu\nu},$$

$$k=1,3,5,7,\quad q=0,\pm 1,\ldots,\pm k.$$

Using $\chi_{\mu\nu}$ we can construct the generators of the subgroups as follows:

SO(5):

$$\mathrm{SU}_a(2):\lambda_0=\tfrac{1}{2}(-\chi_{1-1}+\chi_{2-2}),$$

$$\lambda_{\pm}=\pm(1/\sqrt{2})\chi_{\pm 1\pm 2}, \tag{6a}$$

$$\mathrm{SU}_b(2):\nu_0=\tfrac{1}{2}(\chi_{1-1}+\chi_{2-2}),$$

$$\nu_{\pm}=\mp(1/\sqrt{2})\chi_{\mp 1\pm 2}, \tag{6b}$$

$$U^{(1/2)(1/2)}_{\alpha\beta}\quad\begin{array}{c|cc}\alpha\backslash\beta & \frac{1}{2} & -\frac{1}{2}\\ \hline \frac{1}{2} & (1/\sqrt{2})\chi_{20} & (1/\sqrt{2})\chi_{10}\\ -\frac{1}{2} & (1/\sqrt{2})\chi_{0-1} & (1/\sqrt{2})\chi_{0-2}\end{array}, \tag{6c}$$

$$\mathrm{SU}_1(2):\quad \sigma_0=\tfrac{1}{2}(-\chi_{3-3}+\chi_{4-4}),$$

$$\sigma_{\pm}=\pm(1/\sqrt{2})\chi_{\pm 3\pm 4}, \tag{6d}$$

$$\mathrm{SU}_2(2):\quad \tau_0=\tfrac{1}{2}(\chi_{3-3}+\chi_{4-4}),$$

$$\tau_{\pm}=\mp(1/\sqrt{2})\chi_{\mp 3\pm 4}. \tag{6e}$$

The remaining generators of SO(9) can be put in the form of a tensor operator, as given in Table I.

Using Eq. (5), we can construct the commutation relations for these operators. For example,

$$[\lambda_0,\lambda_{\pm}]=\pm\lambda_{\pm},\quad [\lambda_+,\lambda_-]=-\lambda_0,$$

$$[\lambda_0,U^{(1/2)(1/2)}_{\alpha\beta}]=\alpha U^{(1/2)(1/2)}_{\alpha\beta},$$

$$[\lambda_{\pm},U^{(1/2)(1/2)}_{\alpha\beta}]=\mp\sqrt{\tfrac{1}{2}(\tfrac{1}{2}\mp\alpha)(\tfrac{1}{2}\pm\alpha+1)}\;U^{(1/2)(1/2)}_{\alpha\pm 1\beta},$$

$$[U^{(1/2)(1/2)}_{\pm(1/2)(1/2)},U^{(1/2)(1/2)}_{\pm(1/2)-(1/2)}]=\tfrac{1}{2}\lambda_{\pm},$$

$$[U^{(1/2)(1/2)}_{(1/2)(1/2)},U^{(1/2)(1/2)}_{-(1/2)\pm(1/2)}]=\tfrac{1}{2}\nu_{\pm},$$

$$[U^{(1/2)(1/2)}_{\pm(1/2)(1/2)},U^{(1/2)(1/2)}_{\mp(1/2)-(1/2)}]=\tfrac{1}{2}(\lambda_0\pm\nu_0),\ \text{etc.} \tag{7}$$

The angular momentum operator now can be written as

$$\hat{L}_0=\sqrt{60}(g^{\dagger}\tilde{g})^{(1)}_0=\nu_0+3\lambda_0+\tau_0+7\sigma_0,$$

$$\hat{L}_{\pm}=\sqrt{60}(g^{\dagger}\tilde{g})^{(1)}_{\pm 1}$$

$$=\sqrt{2}(3\nu_{\pm}+2\tau_{\pm})-\sqrt{20}U^{(1/2)(1/2)}_{\pm(1/2)\mp(1/2)}$$

$$-\sqrt{14}T^{(10)(1/2)}_{\mp(1/2)\mp(1/2),\pm(1/2)\mp(1/2)}. \tag{8}$$

TABLE I. Tensor operator $T^{(10)(1/2)}_{\Lambda\alpha\beta,\gamma\delta}$.

α	β \ $\gamma\delta$	$\frac{1}{2}\frac{1}{2}$	$\frac{1}{2}-\frac{1}{2}$	$-\frac{1}{2}\frac{1}{2}$	$-\frac{1}{2}-\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$-(1/\sqrt{2})\chi_{42}$	$-(1/\sqrt{2})\chi_{32}$	$(1/\sqrt{2})\chi_{2-3}$	$-(1/\sqrt{2})\chi_{2-4}$
$\frac{1}{2}$	$-\frac{1}{2}$	$-(1/\sqrt{2})\chi_{41}$	$-(1/\sqrt{2})\chi_{31}$	$-(1/\sqrt{2})\chi_{1-3}$	$-(1/\sqrt{2})\chi_{1-4}$
$-\frac{1}{2}$	$\frac{1}{2}$	$(1/\sqrt{2})\chi_{4-1}$	$(1/\sqrt{2})\chi_{3-1}$	$(1/\sqrt{2})\chi_{-1-3}$	$(1/\sqrt{2})\chi_{-1-4}$
$-\frac{1}{2}$	$-\frac{1}{2}$	$(1/\sqrt{2})\chi_{4-2}$	$(1/\sqrt{2})\chi_{3-2}$	$(1/\sqrt{2})\chi_{-2-3}$	$(1/\sqrt{2})\chi_{-2-4}$
0	0	$(1/\sqrt{2})\chi_{40}$	$(1/\sqrt{2})\chi_{30}$	$(1/\sqrt{2})\chi_{0-3}$	$(1/\sqrt{2})\chi_{0-4}$

Similarly the Casimir operator of SO(9), as appears in the usual[2] IBM reduction scheme, can be written as

$$\hat{C}_9=\sum_{k,q}(-)^q(g^{\dagger}\tilde{g})^{(k)}_q(g^{\dagger}\tilde{g})^{(k)}_{-q}$$

$$=\hat{C}_5+\sigma^2+\tau^2+2\sqrt{5}(T^{(10)(1/2)}\times T^{(10)(1/2)})^{(0)}, \tag{9}$$

where

$$\hat{C}_5=\lambda^2+\nu^2+2(U^{(1/2)(1/2)}\times U^{(1/2)(1/2)})^{(0)}. \tag{10}$$

If we define the operators $\hat{S}_+$ and $\hat{S}_-$ as

$$\hat{S}_+=\sum_{\mu}(-1)^{\mu}g^{+}_{\mu}\,g^{+}_{-\mu},$$

$$\hat{S}_-=\sum_{\mu}(-1)^{\mu}g_{-\mu}g_{\mu}, \tag{11}$$

we can rewrite Eq. (9) as

$$\hat{C}_9=\tfrac{1}{2}\{\hat{N}(\hat{N}+7)-\hat{S}_+\hat{S}_-\}.$$

Obviously, the eigenvalue of $\hat{C}_9$ is $(v/2)(v+7)$ when acting on the state $|nv\Omega\rangle$. Furthermore, we can prove that $\lambda^2=\nu^2$, and $\sigma^2=\tau^2$. Our group reduction scheme is now completely defined and we can proceed with the explicit construction of the basis vectors.

III. BASIS VECTORS

In correspondence with the group chain (3), the quantum number Ω may be taken as the eigenvalues of the operators $\hat{C}_5,\lambda^2,\lambda_0,\nu_0,\sigma^2,\sigma_0$, and τ_0. Hence, a general state vector $|nv\Omega\rangle$ can be rewritten more explicitly as $|nv\Omega\rangle=|nv;p,\Lambda\alpha\beta;\Sigma\gamma\delta\rangle$, and the eigenvalues for above-stated operators are as follows:

$$\begin{array}{lccccccc}\text{operators} & \hat{C}_5 & \lambda^2 & \lambda_0 & \nu_0 & \sigma^2 & \sigma_0 & \tau_0\\ \text{eigenvalues} & (p/2)(p+3) & \Lambda(\Lambda+1) & \alpha & \beta & \Sigma(\Sigma+1) & \gamma & \delta\end{array} \tag{12}$$

In order to give the concrete structure of $|nv;p,\Lambda\alpha\beta;\Sigma\gamma\delta\rangle$, we first observe that the state $|nv;p,\Lambda\alpha\beta;\Sigma\gamma\delta\rangle$ may be written as

$$|nv;p,\Lambda\alpha\beta;\Sigma\gamma\delta\rangle=N_1(\hat{S}_+)^{\rho}|vv;p,\Lambda\alpha\beta;\Sigma\gamma\delta\rangle, \tag{13a}$$

where

$$N_1=\sqrt{\frac{(2v+7)!!}{2^{\rho}\rho!(2\rho+2v+7)!!}},\quad \rho=(n-v)/2,$$

$$v=n,n-2,n-4,\ldots,1\text{ or }0, \tag{13b}$$

and the operator S_+ is defined in Eq. (11). Hence, the problem is reduced to the construction of the state $|vv;p,\Lambda\alpha\beta;\Sigma\gamma\epsilon\rangle$. This can be achieved by the following procedure.

First, we can start with an unique maximum weight state

$$|\Phi_1\rangle = \left|vv;0,000;\frac{v}{2}\,\frac{v}{2}\,\frac{v}{2}\right\rangle = \sqrt{\frac{1}{v!}}(g_4^\dagger)^v|0\rangle. \tag{14}$$

For the state $|\Phi_1\rangle$, the eigenvalues of the operators $\hat{C}_5$, λ^2,λ_0, ν_0. σ^2, σ_0, and τ_0 are 0,0,0,0,$(v/2)(v/2+1)$, $(v/2)$, and $v/2$, respectively.

It can easily be seen that an eigenstate of the operators C_5 and σ^2 with an eigenvalue $(p/2)(p+3)$ and $\Sigma(\Sigma+1)$, can be written as

$$|\Phi_2\rangle = \left|vv;p,\frac{p}{2}\,\frac{p}{2}\,\frac{p}{2},\Sigma\Sigma\Sigma\right\rangle = N_2(g_4^\dagger)^{2\Sigma}(\hat{Z})^\omega(g_2^\dagger)^p|0\rangle, \tag{15a}$$

with

$$N_2 = \left\{\sum_{x=0}^{\omega}\frac{\omega!\omega!2^{3\omega-3x}5^{2x}(2\omega-2x+2p+3)!!(x+2\Sigma+1)!}{(\omega-x)!x!(2p+3)!!(2\Sigma+1)}\right\}^{-1/2}, \tag{15b}$$

and where p and ω should satisfy Eqs. [1(a) and 1(b)], and the operator $\hat{Z}$ is defined as

$$\hat{Z} = \sqrt{\tfrac{1}{90}}\{2(g_0^+\,g_0^+ - 2\beta_0) + 5(g_3^+\,g_{-3}^+ + g_4^+\,g_{-4}^+)\}, \tag{16a}$$

$$\beta_0^\dagger = (g_1^\dagger\,g_{-1}^\dagger - g_2^\dagger\,g_{-2}^\dagger). \tag{16b}$$

It can be proven that both of the operators $g_0^+\,g_0^+ - 2\beta_0$ and $(g_3^+\,g_{-3}^+ - g_4^+\,g_{-4}^+)$ are invariants of both SO(5) and $SU_1(2)\times SU_2(2)$. Hence the state $|\Phi_2\rangle$ is a simultaneous eigenstate of operators C_5, λ^2, λ_0, ν_0, σ^2, σ_0, and τ_0 with eigenvalues $(p/2)(p+3)$,$(p/2)(p/2+1)$, $p/2$, $p/2$,$\Sigma(\Sigma+1)$,Σ, and Σ, respectively.

Third, we need a state that is an eigenstate of the operator λ^2 (and ν^2) with the eigenvalue $\Lambda(\Lambda+1)$. The procedure followed is described in Refs. 6 and 7, with the result

$$|\Phi_3\rangle = N_3 = (g_4^\dagger)^{2\Sigma}(\hat{Z})^\omega F(p,\Lambda)(g_2^\dagger)^{2\Lambda}|0\rangle, \tag{17a}$$

with

$$N_3 = \left\{\sum_{x=0}\sum_{\substack{y,z,y',z'\\(y+z=y'+z')}}\frac{(N_{p\Lambda}2^{\omega-x}5^x\omega!)(x+\Sigma+1)!4^{y'-z}}{(2\Lambda+1+z)!z!(p-2\Lambda-2z)!(2\Lambda+1+z')!z'!(p-2\Lambda-2z')!}\right.$$
$$\left.\times\frac{(p-2\Lambda-2z+2\omega-2x-2y)!(y+z)!(y+z+\Lambda+1)!}{x!(\omega-x-y)!y!(\omega-x-y')!y'!(\Lambda+1)(\Sigma+1)}\right\}^{-1/2}, \tag{17b}$$

where

$$\Lambda = p/2,(p-1)/2,\ldots,0,$$

$$F(p,\Lambda) = N_{p\Lambda}\sum_{z=0}\frac{1}{2^z(2\Lambda+1+z)!z!(p-2\Lambda-2z)!}(g_0^\dagger)^{p-2\Lambda-2z}(\hat{\beta}_0)^z,$$

$$N_{p\Lambda} = \left\{\sum_{z=0}\left[\frac{1}{2^z(2\Lambda+1+z)!}\right]^2\frac{(z+1)}{((p-2\Lambda)-2z)!}\right\}^{1/2}.$$

Finally, by acting with $\lambda_-^{(\Lambda-\alpha)}$,$\nu_-^{(\Lambda-\beta)}$,$\sigma_-^{(\Sigma-\gamma)}$, and $\tau_-^{(\Sigma-\delta)}$ on the state $|\Phi_3\rangle$, we can obtain a general state vector, which can be written as

$$|vv;p,\Lambda\alpha\,\beta,\Sigma\gamma\delta\rangle = N_4(-)^{\Lambda+\Sigma-\alpha-\gamma}(1/\sqrt{2})^{2\Lambda+2\Sigma-\alpha-\beta-\gamma-\delta}(\hat{Z})^\omega F(p,\Lambda)G(\Lambda\alpha\,\beta;\Sigma\gamma\epsilon)|0\rangle, \tag{18a}$$

where

$$N_4 = (2\Lambda)!(2\Sigma)!\left\{\frac{(\Lambda-\alpha)!(\Lambda-\beta)!(\Sigma-\gamma)!(\Sigma-\delta)!}{2^{2\Lambda+2\Sigma-\alpha-\beta-\gamma-\delta}(\Lambda+\alpha)!(\Lambda+\beta)!(\Sigma+\gamma)!(\Sigma+\delta)!}\right\}^{-1/2}N_3 \tag{18b}$$

and

$$G(\Lambda\alpha\,\beta;\Sigma\gamma\delta) = \sum_{x=0}\sum_{y=0}\Theta(\Lambda\alpha\,\beta x;\Sigma\gamma\delta y)\times\Omega(\Lambda\alpha\,\beta x;\Sigma\gamma\delta y),$$

$$\Theta(\Lambda\alpha\,\beta x;\Sigma\gamma\delta y) = \frac{(2\Lambda)!(\Lambda-\alpha-1)!(\Lambda-\beta)!}{(x-1+\delta_{x0})!(\Lambda-\alpha-x-\delta_{x0})!(\Lambda+\beta-x)!(\alpha-\beta+x)!}$$
$$\times\frac{(2\Sigma)!(\Sigma-\gamma-1)!(\Sigma-\delta)!}{(y-1+\delta_{y0})!(\Sigma-\gamma-y-\delta_{y0})!(\Sigma+\delta-y)!(\gamma-\delta+y)!},$$

$$\Omega(\Lambda\alpha\,\beta x;\Sigma\gamma\delta y) = (g_2^\dagger)^{\Lambda+\beta-x}(g_{-2}^\dagger)^{\Lambda-\alpha-x}(g_1^\dagger)^{\alpha-\beta+x}(g_{-1}^\dagger)^x(g_4^\dagger)^{\Sigma+\delta-y}(g_{-4}^\dagger)^{\Sigma-\gamma-y}(g_3^\dagger)^{\gamma-\delta+y}(g_{-3}^\dagger)^y.$$

We have now obtained a closed, analytic expression for a general basis state, labeled according to (3), in terms of the g-boson operators. In Table II the relation between the basis (13a) and the natural basis is given explicitly for a configuration with $n<2$.

TABLE II. The natural basis of the g-boson system.

n	v	p	Λ	Σ	Natural basis	L
1	1	1	$\frac{1}{2}$	0	$g_2^\dagger\|0\rangle, g_1^\dagger\|0\rangle, g_{-1}^\dagger\|0\rangle, g_{-2}^\dagger\|0\rangle$	4
	1	1	0	0	$g_0^\dagger\|0\rangle$	
	1	1	0	$\frac{1}{2}$	$g_4^\dagger\|0\rangle, g_3^\dagger\|0\rangle, g_3^\dagger\|0\rangle, g_{-4}^\dagger\|0\rangle$	
2	2	2	1	0	$(1/\sqrt{2})g_2^\dagger g_2^\dagger\|0\rangle, g_2^\dagger g_1^\dagger\|0\rangle, (1/\sqrt{2})g_1^\dagger g_1^\dagger\|0\rangle, g_2^\dagger g_{-1}^\dagger\|0\rangle, (1/\sqrt{2})(g_1^\dagger g_{-1}^\dagger + g_2^\dagger g_{-2}^\dagger)\|0\rangle$ $g_{-2}^\dagger g_1^\dagger\|0\rangle, (1/\sqrt{2})g_{-1}^\dagger g_{-1}^\dagger\|0\rangle, g_{-1}^\dagger g_{-2}^\dagger\|0\rangle, (1/\sqrt{2})g_{-2}^\dagger g_{-2}^\dagger\|0\rangle$	8,6,4,2,0
	2	2	$\frac{1}{2}$	0	$g_2^\dagger g_0^\dagger\|0\rangle, g_1^\dagger g_0^\dagger\|0\rangle, g_0^\dagger g_{-1}^\dagger\|0\rangle, g_0^\dagger g_{-2}^\dagger\|0\rangle$	
	2	2	0	0	$(1/\sqrt{10})(2g_0^\dagger g_0^\dagger + g_1^\dagger g_{-1}^\dagger - g_2^\dagger g_{-2}^\dagger)\|0\rangle$	
	2	1	$\frac{1}{2}$	$\frac{1}{2}$	$g_2^\dagger g_4^\dagger\|0\rangle, g_2^\dagger g_3^\dagger\|0\rangle, g_2^\dagger g_{-3}^\dagger\|0\rangle, g_2^\dagger g_{-4}^\dagger\|0\rangle, g_1^\dagger g_4^\dagger\|0\rangle, g_1^\dagger g_3^\dagger\|0\rangle,$ $g_1^\dagger g_{-3}^\dagger\|0\rangle, g_1^\dagger g_{-4}^\dagger\|0\rangle, g_{-1}^\dagger g_4^\dagger\|0\rangle, g_{-1}^\dagger g_3^\dagger\|0\rangle, g_{-1}^\dagger g_{-3}^\dagger\|0\rangle, g_{-1}^\dagger g_{-4}^\dagger\|0\rangle$ $g_{-2}^\dagger g_{-4}^\dagger\|0\rangle, g_{-2}^\dagger g_3^\dagger\|0\rangle, g_{-2}^\dagger g_{-3}^\dagger\|0\rangle, g_{-2}^\dagger g_{-4}^\dagger\|0\rangle$	
	2	1	0	$\frac{1}{2}$	$g_0^\dagger g_4^\dagger\|0\rangle, g_1^\dagger g_3^\dagger\|0\rangle, g_0^\dagger g_{-3}^\dagger\|0\rangle, g_0^\dagger g_{-4}^\dagger\|0\rangle$	
	2	0	0	1	$(1/\sqrt{2})g_4^\dagger g_4^\dagger\|0\rangle, g_4^\dagger g_3^\dagger\|0\rangle, (1/\sqrt{2})g_3^\dagger g_3^\dagger\|0\rangle, g_4^\dagger g_{-3}^\dagger\|0\rangle, (1/\sqrt{2})(g_4^\dagger g_{-4}^\dagger + g_3^\dagger g_{-3}^\dagger)\|0\rangle,$ $g_3^\dagger g_{-4}^\dagger\|0\rangle, (1/\sqrt{2})g_{-3}^\dagger g_{-3}^\dagger\|0\rangle, g_{-3}^\dagger g_{-4}^\dagger\|0\rangle, (1/\sqrt{2})g_{-4}^\dagger g_{-4}^\dagger\|0\rangle$	
	2	0	0	0	$\sqrt{\frac{1}{30}}(2g_0^\dagger g_0^\dagger - 4g_1^\dagger g_{-1}^\dagger + 4g_2^\dagger g_{-2}^\dagger + 5g_3^\dagger g_{-3}^\dagger - 5g_4^\dagger g_{-4}^\dagger)$	
	0	0	0	0	$\sqrt{\frac{1}{18}}(g_0^\dagger g_0^\dagger - 2g_1^\dagger g_{-1}^\dagger + 2g_2^\dagger g_{-2}^\dagger - 2g_3^\dagger g_{-3}^\dagger + 2g_4^\dagger g_{-4}^\dagger)$	

IV. ANGULAR MOMENTUM PROJECTION

The states constructed using the reduction scheme proposed in the present paper do not have good angular momentum, and as such are not physical states. In practical calculations this problem can be solved by requiring the states to be eigenstates of the angular momentum L^2, and the states can be constructed using numerical methods. We will outline here an alternative analytic procedure that can be followed. As an example we discuss only the case $n = 2$. The formulas for general n will be the subject of a subsequent paper.

In order to form states with good angular momentum we will follow the well-known procedure of taking a maximum weight state (14), which has a maximum value of $M = 4v$ and thus a unique value of L, and operating on it with the L_- operator given in Eq. (8). It should be noted that the states (18) are eigenstates of L_0 with eigenvalue $M = \beta + 3\alpha + \delta + 7\gamma$. The physical states thus have to be a linear combination of states with the same value of M, n, and v.

For the case $n = 2$ the action of the operators entering in L_- can be given as follows:

$$\nu_-|n,v,p;\Lambda,\alpha,\beta;\Sigma,\gamma,\delta\rangle = (1/\sqrt{2})\sqrt{(\Lambda+\beta)(\Lambda-\beta+1)}\,|n,v,p,\Lambda,\alpha,\beta-1,\Sigma,\gamma,\delta\rangle,$$

$$\tau_-|n,v,p;\Lambda,\alpha,\beta;\Sigma,\gamma,\delta\rangle = (1/\sqrt{2})\sqrt{(\Sigma+\delta)(\Sigma-\delta+1)}\,|n,v,p;\Lambda,\alpha,\beta;\Sigma,\gamma,\delta-1\rangle,$$

$$U_{-+}|n,v,p;\Lambda,\alpha,\beta;\Sigma,\gamma,\delta\rangle = \frac{1}{2}\sqrt{\frac{(p-2\Lambda)(p+2\Lambda+3)(\Lambda-\alpha+1)(\Lambda+\beta+1)}{(2\Lambda+1)(2\Lambda+2)}}\,|n,v,p;\Lambda+\tfrac{1}{2},\alpha-\tfrac{1}{2},\beta+\tfrac{1}{2},\Sigma,\gamma,\delta\rangle$$
$$-\frac{1}{2}\sqrt{\frac{(p-2\Lambda+1)(p+2\Lambda+2)(\Lambda+\alpha)(\Lambda-\beta)}{2\Lambda(2\Lambda+1)}}\,|n,v,p;\Lambda-\tfrac{1}{2},\alpha-\tfrac{1}{2},\beta+\tfrac{1}{2};\Sigma,\gamma,\delta\rangle,$$

$$T^{(10)(1/2)}_{+(1/2)+(1/2),-(1/2)+(1/2)}|n=2,v,p,\Lambda,\alpha,\beta;\Sigma,\gamma,\delta\rangle = (-1/\sqrt{2})\delta_{v'v}\Big\{\delta_{\Sigma',\Sigma-1/2}\delta_{\Lambda',\Lambda+1/2}\delta_{p',p+1}$$
$$\times\sqrt{\frac{(\Sigma+\gamma)(\Sigma-\delta)(\Lambda+\alpha+1)(\Lambda+\beta+1)}{2\Sigma(2\Lambda+1)}} + \delta_{\Sigma',\Sigma+1/2}\delta_{\Lambda',\Lambda+1/2}\delta_{p',p-1}$$
$$\times\sqrt{\frac{(\Sigma-\gamma+1)(\Sigma+\delta+1)(\Lambda-\alpha)(\Lambda-\beta)}{(2\Sigma+1)2\Lambda}} + (1/\sqrt{10})[\delta_{\Lambda,0}\delta_{\Sigma,0}\delta_{p',1}\delta_{\Lambda',1/2}\delta_{\Sigma',1/2}$$
$$\times(\delta_{p,2}-3\delta_{p,0}) + \delta_{\Lambda',0}\delta_{\Sigma',0}\delta_{p,1}\delta_{\Lambda,1/2}\delta_{\Sigma,1/2}(\delta_{p',2}-3\delta_{p',0})]\Big\}$$
$$\times|n=2,v',p';\Lambda',\alpha+\tfrac{1}{2},\ \beta+\tfrac{1}{2};\Sigma',\gamma-\tfrac{1}{2},\delta+\tfrac{1}{2}\rangle.$$

With the use of these formulas the action of L_- on the highest weight state, given by Eq. (14), can be calculated:

$$L_-|g_4^2\rangle = L_-|2,2,0;0,0,0;1,1,1\rangle = 2\sqrt{2}|2,2,0;0,0,0;1,1,0\rangle,$$

where only the τ_- operator gives a nonzero contribution. Since the state on the left-hand side (lhs) corresponds to the state $|L=8,\quad M=8\rangle$, the state on the rhs necessarily corresponds to $|L=8,\quad M=7\rangle$. By acting once again with the L_- operator, we obtain

$$L_-|2,2,0;0,0,0;1,1,0\rangle = 2\sqrt{2}|2,2,0;0,0,0;1,1,-1\rangle + \sqrt{7}|2,2,1;\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle,$$

where now both the τ_- and the $T^{(10)(1/2)}_{+(1/2)+(1/2),-(1/2)+(1/2)}$ operators contribute. The state on the rhs thus corresponds to a state with $|L=8,\quad M=6\rangle$. The state $|L=6,\quad M=6\rangle$ can be obtained by taking the orthogonal linear combination. This procedure can be used to construct all states with good angular momentum.

V. SUMMARY

In this paper we have given a method for constructing the basis states of a system of g-bosons. The advantage of the present reduction scheme over the conventional $U(9)\supset SO(9)\supset SO(3)\supset SO(2)$ is that there are no missing labels, the states can be uniquely labeled by the values of the Casimir operators. The only disadvantage of the present reduction scheme is that angular momentum is not a good quantum number. In the last section we have given an example of an angular momentum projection procedure that can be followed. The procedure outlined in this paper can, of course, be extended without problems to a system of s-, d-, and g-bosons.

ACKNOWLEDGMENT

Support for this research has been provided by the U.S. National Science Foundation under Grant No. PHY 84-07858.

[1] A. Arima and F. Iachello, Ann. Rev. Nucl. Part. Sci. **31**, 75 (1981).

[2] K. H. Sage and B. R. Barrett, Phys. Rev. C **22**, 1765 (1980); R. D. Ratna Raju, Phys. Rev. C **23**, 518 (1981); Wu Hua Chuan, Phys. Lett. B **110**, 1 (1982); L. J. B. Goldfarb, Phys. Lett. B **104**, 103 (1982); P. Isacker, K. Heyde, M. Waroquier, and G. Wenes. Nucl. Phys. **380**, 383 (1982); Y. S. Ling, Chin. Phys. **3**, 86 (1983); H. Z. Sun, M. Moshinsky, A. Frank, and P. Isacker, Kinam **5**, 135 (1983); V. K. B. Kota (unpublished); H. Z. Sun and Z. R. Yu (unpublished).

[3] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

[4] R. C. King, J. Phys. A: Math. Gen. **8**, 429 (1975).

[5] G. R. E. Blak, R. C. King, and B. G. Wybourne, J. Phys. A: Math. Gen. **16**, 1555 (1983).

[6] K. T. Hecht, Nucl. Phys. **63**, 177 (1965).

[7] H. Z. Sun, Phys. Energ. Fortis Phys. Nucl. **4**, 478 (1980) (in Chinese).

Remarks on bifurcation with symmetry, gradient property, and reducible representations

Giampaolo Cicogna
Dipartimento di Fisica, Universitá di Pisa, 56100 Pisa, Italy

Giuseppe Gaeta
Department of Physics, New York University, New York, New York 10003

Pietro Rossi
Department of Physics, Cornell University, Ithaca, New York 14853

(Received 13 November 1984; accepted for publication 16 September 1985)

The gradient property for bifurcation equations covariant with respect to a group representation that is reducible, but irreducible as a real representation, is examined. In this case, the Schur lemma does not hold in the usual form, and one is faced with problems not present in the irreducible case. It is shown how to handle these problems, and applications to the fundamental real representations of SO(2) and SU(2) are presented that, due to general group theoretical results, represent in a sense the more general situation.

I. INTRODUCTION

In this paper we want to present some remarks about the gradient property of bifurcation equations in the presence of a symmetry[1] when the symmetry is described by a real reducible representation. Precisely, we are interested in the case, important in the physical applications, in which this representation T is reducible, but is irreducible on real numbers, i.e., it is physically irreducible.

The main difference between this case and the irreducible one is that for reducible representations the Schur lemma does not forbid the existence of nontrivial operators commuting with the whole representation; that is, now we can have

$$T_g K = K T_g, \quad \forall g \in G, \tag{1}$$

for operators $K \neq aI$.

If we denote by $C(T)$ the space of intertwining operators, i.e., the set of real operators commuting with the real irreducible representation T,

$$C(T) = \{K / K T_g = T_g K, \ \forall g \in G\}, \tag{2}$$

we have a theorem (8.2.2 of Ref. 2) which ensures $C(T)$ is isomorphic to $\mathbf{R}$, $\mathbf{C}$, or $\mathbf{H}$, where $\mathbf{R}$ is the field of real numbers, $\mathbf{C}$ that of complex numbers, and $\mathbf{H}$ that of quaternions. From this it follows easily that the dimension D of the operators in $C(T)$ can be 1,2, or 4.

The case $D = 1$, $C(T) \approx \mathbf{R}$ is trivial. We will treat in the following examples of the other two cases, SO(2) for $C(T) \approx \mathbf{C}$ and SU(2) for $C(T) \approx \mathbf{H}$.

Both cases $D = 2$ and $D = 4$ raise problems in connection with the gradient property[1,3] of bifurcation equations: we will discuss these problems, and how to handle them, in the following sections.

We stress that, because of the group theoretical results quoted here and at the end of Sec. III, our treatment covers cases that are generic, in the sense that the most general case is a combination of these three basic cases, $C(T) \approx \mathbf{R}$, $\mathbf{C}$, $\mathbf{H}$.

For what concerns the gradient property, the case of SO(2) was dealt with shortly in a previous paper by two of the present authors,[3] but we present it here anew in more detail for clarity and completeness.

It can also be noted that the property $D = 2$ for the group SO(2) is closely related to the existence of nonzero solutions of the classical Hopf bifurcation problem[1,4] [here SO(2) describes the covariance with respect to time translation $t \to t + \tau$, mod 2π], whereas the case $D = 4$ should correspond to a more complicated and less known "quaternionic bifurcation."[4,5]

II. REDUCIBLE REPRESENTATIONS

Let the basis of the representation T be $\{x_1, \ldots, x_n\}$ and denote by $\mathbf{v}_0$ the covariant of first order corresponding to this array of basis functions, namely $\mathbf{v}_0 = \{v_1^0, v_2^0, \ldots, v_n^0\}$ with $v_j^0 = x_j$.

For any K in $C(T)$,

$$\mathbf{v} = K\mathbf{v}_0$$

is a covariant of first order. Therefore, if $C(T) \neq \{I\}$ (for ease of notation, we do not distinguish between K and aK, as well as for $\mathbf{v}$ and $a\mathbf{v}$, where a is a real number) the covariant of first order is not unique or, in the language of Ref. 3, there is more than one fundamental vector of order 1. This means that one could choose $\mathbf{v}$ as well as $\mathbf{v}_0$ as a basis for this representation, that is, substitute $\{x_1, \ldots, x_n\}$ with $\{x_1', \ldots, x_n'\}$, where $x_i' = K_{ij} x_j$, obtaining exactly the same representation T. Of course, all of the $\mathbf{v}$'s obtained in this way are independent, and these are the only covariants of order 1, that is, any covariant of order 1 can be written as $\mathbf{v} = a_i K_i \mathbf{v}_0$, where $a_i \in \mathbf{R}$, $K_i \in C(T)$.

We have now to deal with gradients: since the gradient is defined with respect to some basis and we now have more equivalent bases, the gradient is not defined uniquely any longer. Therefore, when we speak about a gradient, we have to mean with respect to any one of the bases individuated by first-order covariants. This corresponds to the following fact: as everyone knows, if S is a scalar function, its gradient ∇S is a vector (i.e., is a covariant), due to the vectorial character of the operator ∇. But now we can define different ∇_i's, each of them being a vectorial one, and from a single function S we can obtain different vectors by computing its gradients in the different bases, $\nabla_i S$.

 0022-2488/86/020447-04$02.50 © 1986 American Institute of Physics

If two different bases, $\mathbf{x}$ and $\mathbf{y} = R\mathbf{x}$, are given, the gradients in the two bases are related by $\partial/\partial x_i = (\partial y_k/\partial x_i)(\partial/\partial y_k)$, that is, $\nabla_x = R\nabla_y$, denoting by ∇_0 the "ordinary" gradient, that is the one in the basis $\mathbf{v}_0$, any gradient can be expressed in terms of ∇_0 and the operators belonging to $C(T)$,

$$\nabla = a_i K_i \nabla_0,$$

and the $\nabla_i = K_i \nabla_0$ are independent.

At this point it should be clear that the gradient property,[3] namely the fact that any vector is the gradient of a scalar, means that any covariant is obtained by applying one of the gradient operators to a scalar. We can put this in more precise terms: A representation T_g of a group G acting on the basis $\mathbf{v}_0$ is said to have the gradient property at order n if for any covariant $\mathbf{v}$ of order n there are real numbers a_i and a scalar S of order $n+1$ such that $\mathbf{v} = a_i K_i \nabla_0 S$.

It follows from general theorems in bifurcation theory (see Ref. 1) that if a bifurcation equation relative to a system with G as a symmetry group is covariant with respect to the representation T_g and T_g has the gradient property at order n, the bifurcation equation has the gradient property at order n, too, that is, its terms of order n can be written as a gradient.

The gradient property at each order of T_g is, in general, not enough to ensure that the bifurcation equation $F(\lambda, \mathbf{v}) = 0$ is itself a gradient equation. In fact, the bifurcation equation is, in general, written as a sum of covariants

$$F(\lambda, \mathbf{v}) = \sum_n B_n(\lambda, \mathbf{v}),$$

where each B_n is n-linear in $\mathbf{v}$. If T_g has gradient property, each of the covariants can be written as

$$B_n = \sum_i b_i^{(n)} K_i \nabla_0 S^{(n)},$$

and $F(\lambda, \mathbf{v})$ is given by

$$F(\lambda, \mathbf{v}) = \sum_{n,i} b_i^{(n)} K_i \nabla_0 S^{(n)}.$$

Since there is no reason for the $b_i^{(n)}$ to be independent of n, this, in general, cannot be expressed in the form $F(\lambda, \mathbf{v}) = c_i K_i \nabla_0 S$, where S is some scalar function.

A special case is that of representations, like those we consider in the following, such that all the scalars are of the form $(\mathbf{v}, \mathbf{v})^m$. In this case, if the gradient property is satisfied, one has

$$F(\lambda, \mathbf{v}) = \sum_m b_i^{(m)} K_i \nabla_0 (\mathbf{v}, \mathbf{v})^m = f_i K_i \nabla_0 S_0,$$

with $S_0 = (\mathbf{v}, \mathbf{v})$ and

$$f_i = \sum_m 2m(\mathbf{v}, \mathbf{v})^{m-1} b_i^{(m)},$$

so that the bifurcation equation is itself, in this generalized sense, a gradient equation.

III. GRADIENT PROPERTY

After having clarified the concept of gradient property in the case of reducible representations, we can pass to the problem of proving it. In particular, we want to discuss the applicability of the technique used in Ref. 3.

That was based on the computation of the number of independent covariant and invariant completely symmetric tensors in the basis functions of T at any order; this computation was performed by making use of the orthogonality relations for characters and of the Molien function,[1] which together led to an integral formula for the multiplicity of vectors (completely symmetric covariant tensors) and scalars (completely symmetric invariant tensors) at any order, which were equal to the multiplicities $c_n^{(1)}$ of T and $c_n^{(0)}$ of the identity representation T_0 in the symmetrized n-tensor product $(T^{\otimes n})_s$.

We have to check therefore the applicability of the criterion based on the multiplicities of scalars and vectors, and of the integral formula, in the case of a reducible representation T.

As for confronting the multiplicities, the valid criterion is that the gradient property holds at order n if the number of independent vectors of order n, $\#v(n)$, is equal to the number of independent vectors that can be obtained from scalars of order $n+1$ by the action of a gradient operator. Now, if D is the number of independent operators in $C(T)$, from each scalar we have a set of D independent vectors by applying gradient operators. Therefore, if we have $c_{n+1}^{(0)}$ independent scalars at order $n+1$, we can obtain $D \cdot c_{n+1}^{(0)}$ vectors by application of gradient operators. If all of these are independent, the gradient property at order n is equivalent to

$$\#v(n) = D \cdot c_{n+1}^{(0)}.$$

For the representations we are going to deal with, independence is trivially ensured, since we will have at most one scalar at each order, so we will not be concerned about this.

To obtain the number $\#v(n)$, one has to multiply by D the multiplicity of T in $(T^{\otimes n})_s$, that is, we have

$$\#v(n) = Dc_n^{(1)}$$

(with the same remarks as for gradients about independence). Therefore we have, as in the irreducible case, that to ensure the gradient property at order n it is sufficient to have

$$c_n^{(1)} = c_{n+1}^{(0)}.$$

When computing $c_{n+1}^{(0)}$ and $c_n^{(1)}$ we are concerned with the problem of decomposing the symmetrized n-tensor product $P_n = (T^{\otimes n})_s$. If we denote its character by χ_n, we can decompose it into irreducible representations by the standard formula

$$P_n = \sum_a m_a^{(n)} T^{(a)},$$

$$m_a^{(n)} = \int_G \chi_n(g)\overline{\chi}^{(a)}(g)\,d\mu_g,$$

where $\chi_n(g)$ can be obtained via a generating function [1,3]

$$\sum z^n \chi_n(g) = \det(I - zT(g))^{-1}$$

and $d\mu_g$ is the Haar invariant measure, $\chi^{(a)}(g)$ the character of $T^{(a)}(g)$.

If we want to know $c_n^{(1)}$, that is, the multiplicity of T in P_n, we have to decompose T as

$$T = \sum_{a=1}^{R} A_a T^{(a)},$$

and we have

$$c_n^{(1)} = \min_{a \in (1,R)} [m_a^{(n)}/A_a],$$

where the square brackets mean the integer part.

We remark here that another theorem [(8.2.3) of Ref. 2] ensures that for T irreducible on the reals, we have $T = T_1 \oplus T_2$ with T_1 and T_2 not equivalent, if $D = 2$, $C(T) \approx \mathbf{C}$; and $T = T_1 \oplus T_1$ if $D = 4$, $C(T) \approx \mathbf{H}$.

As anticipated, we want now to give short examples of applications of these remarks, choosing the fundamental real representations of the groups that are the most natural ones in view of the group theoretical theorems cited above, that is, SO(2) for $D = 2$, $C(T) \approx \mathbf{C}$, and SU(2) for $D = 4$, $C(T) \approx \mathbf{H}$.

IV. SO(2)

Now we examine the case $G = \mathrm{SO}(2)$. Its fundamental real representation is given by

$$T(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

which could be reduced to

$$T(\theta) = \begin{pmatrix} e^{i\theta} & \\ & e^{-i\theta} \end{pmatrix} = T^{(1)} \oplus T^{(2)}.$$

The set $C(T)$ is given by

$$C(T) = \{I, K\},$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad K = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

to which correspond the vectors $\binom{x}{y}$ and $\binom{-y}{x}$.

We can now easily compute $c_N^{(0)}$ and $c_N^{(1)}$ by standard residue technique. The Haar measure for SO(2) is simply $d\theta/2\pi$, θ varying from 0 to 2π. Therefore we have

$$I^{(0)} = \sum_{n=0}^{\infty} c_n^{(0)} z^n = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{(1 - ze^{i\theta})(1 - ze^{-i\theta})}$$

$$= \frac{1}{2\pi i} \oint \frac{d\omega}{(1 - z\omega)(\omega - z)} = \frac{1}{1 - z^2} = \sum_{n=0}^{\infty} z^{2n}.$$

As for $c_N^{(1)}$, we notice that

$$\int_0^{2\pi} [(1 - ze^{i\theta})(1 - ze^{-i\theta})]^{-1} e^{i\theta}\, d\theta$$

$$= \int_0^{2\pi} [(1 - ze^{i\theta})(1 - ze^{-i\theta})]^{-1} e^{-i\theta}\, d\theta,$$

so that, with the notation used above, $m_1^{(N)} = m_2^{(N)}$ and $c_N^{(1)} = m_1^{(N)} = \frac{1}{2}(m_1^{(N)} + m_2^{(N)})$. Therefore,

$$I^{(1)} = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta}}{(1 - ze^{i\theta})(1 - ze^{-i\theta})}\, d\theta$$

$$= \frac{1}{2\pi i} \oint \frac{\omega}{(1 - z\omega)(\omega - z)} d\omega = \frac{z}{1 - z^2} = \sum_{n=0}^{\infty} z^{2n+1}.$$

So we have that for the fundamental representation of SO(2) the gradient property is satisfied at any order: we have a scalar at each even order, which is, of course, $(\mathbf{v}_0, \mathbf{v}_0)^m = (K\mathbf{v}_0, K\mathbf{v}_0)^m$, and two vectors at each odd order, $\mathbf{v}_0$ and $K\mathbf{v}_0$, while there are no scalars at odd orders nor vectors at even orders.

This same result holds for all the representations of SO(2), given by

$$T_m(\theta) = \begin{pmatrix} \cos(m\theta) & \sin(m\theta) \\ -\sin(m\theta) & \cos(m\theta) \end{pmatrix},$$

for which the computation differs only for a trivial change in the integration variable, $\theta' = m\theta$.

The bifurcation equation for SO(2) is therefore of the form, as expected from Sec. II,

$$F(\lambda, \mathbf{v}) = F_0(\lambda, \|v\|)\mathbf{v} + F_1(\lambda, \mathbf{v})K\mathbf{v} = 0.$$

Each nth-order component can be written as

$$F_n(\lambda, \mathbf{v}) = S_n(\lambda, \|v\|)\mathbf{u},$$

where

$$\mathbf{u} = \begin{pmatrix} a^{(n)}x_1 - b^{(n)}x_2 \\ a^{(n)}x_2 + b^{(n)}x_1 \end{pmatrix}, \quad \mathbf{v} = (x_1, x_2),$$

or, with the notation of Sec. II,

$$\mathbf{u} = \tfrac{1}{2}(aI + bK)\nabla_0(\mathbf{v}, \mathbf{v}).$$

The nth-order bifurcation equation is therefore a gradient equation in our "generalized" sense: i.e., taking the gradient with respect to the basis $\mathbf{v}'$ obtained from $\mathbf{v} = (x_1, x_2)$ by applying the operator $(aI + bK)^{-1}$, namely $\mathbf{v}_n' = ((aI - bK)/(a^2 + b^2))\mathbf{v}$, and not with respect to the "ordinary" gradient ∇_0.

We remind the reader that, as suggested by the notation, a's and b's depend on the order n, so that also the basis $\mathbf{v}'$ depends on it.

V. SU(2), OR QUATERNIONIC BIFURCATION

We come now to the case $G = \mathrm{SU}(2)$, representative of "quaternionic bifurcation" (see Refs. 4 and 5). We consider here the representation T obtained by decomplexifying the standard (defining) representation

$$\Gamma = \begin{pmatrix} A & B \\ -B^* & A^* \end{pmatrix}, \quad |A|^2 + |B|^2 = 1.$$

If we write $A = \alpha + i\eta$, $B = \beta + i\gamma$, it results in

$$T = \begin{pmatrix} \alpha & -\eta & \beta & -\gamma \\ \eta & \alpha & \gamma & \beta \\ -\beta & -\gamma & \alpha & \eta \\ \gamma & -\beta & -\eta & \alpha \end{pmatrix},$$

$$\alpha^2 + \beta^2 + \gamma^2 + \eta^2 = 1,$$

$$\alpha, \beta, \gamma, \eta \in \mathbf{R}.$$

It is perhaps useful to stress that we are using a parametrization different than the one usual in quantum mechanics, namely $A = \exp(i(\varphi + \psi)/2)\cos\theta/2$; $B = \exp(i(\varphi + \psi)/2)\sin\theta/2$. Our parametrization is the standard one in lattice gauge theory computations, and it is easier to handle in the present case anyway.

The decomplexified representation T can be written in compact form, using the operator K defined in the previous section,

$$K = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

as

$$T=\begin{pmatrix} \alpha I+\eta K & \vdots & \beta I+\gamma K \\ \cdots & \cdots & \cdots \\ -\beta I+\gamma K & \vdots & \alpha I-\eta K \end{pmatrix},$$

as is obvious, since K is a representation of the imaginary unity i over R [2].

Now we have $D=4$; $C(T)$ is spanned by $\{I, K_1, K_2, K_3\}$, where

$$K_1=\begin{pmatrix} 0 & 1 & \vdots & & \\ -1 & 0 & \vdots & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & \vdots & 0 & 1 \\ & 0 & \vdots & -1 & 0 \end{pmatrix},$$

$$K_2=\begin{pmatrix} & & \vdots & 0 & 1 \\ & 0 & \vdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & \vdots & & \\ & -1 & \vdots & 0 & \\ -1 & 0 & \vdots & & \end{pmatrix},$$

$$K_3=\begin{pmatrix} & & \vdots & 1 & 0 \\ & 0 & \vdots & 0 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & \vdots & & \\ & & \vdots & 0 & \\ 0 & 1 & \vdots & & \end{pmatrix}.$$

It is immediate to see $K_i^2=-1$, $K_i K_j=\epsilon_{ijh}K_h$, so that effectively $C(T)\approx\mathbb{H}$.

One has also, in accordance with the general group theoretical results quoted above,

$$T=T_1\oplus T_1,\quad T_1=\Gamma.$$

Now we can consider the gradient property. With our parametrization, the Haar measure is given simply by

$$d\mu=\delta(1-\alpha^2-\beta^2-\gamma^2-\eta^2)d\alpha\,d\beta\,d\gamma\,d\eta.$$

Since, in the integrals we have to compute, only class functions appear, it is more convenient to perform integration over classes using the class measure.

The conjugate classes of SU(2), in the present parametrization, are labeled by the value of α. The class measure $d\nu$ is, after a simple integration in β,γ,η,

$$d\nu=(2/\pi)(1-\alpha^2)^{1/2}d\alpha,\quad \alpha\in[-1,1].$$

The determinant is a class function and is given by

$$\det(1-zT(\alpha))=(1+z^2-2z\alpha)^2$$

(we remind the reader that $T=\Gamma\oplus\Gamma$, which accounts for the square), so that we have for $I^{(0)}$, through the standard position $\alpha=\cos\theta$; $e_{i\theta}=\omega$,

$$\begin{aligned} I^{(0)}&=\frac{2}{\pi}\int_{-1}^{1}\frac{(1-\alpha^2)^{1/2}}{(1+z^2-2z\alpha)^2}d\alpha \\ &=\frac{2}{\pi}\int_0^{\pi}\frac{\sin^2\alpha}{(1+z^2-2z\cos\alpha)^2}d\alpha \\ &=\frac{1}{\pi}\int_0^{2\pi}\frac{\sin^2\alpha}{(1+z^2-2z\cos\alpha)^2}d\alpha \\ &=\frac{-1}{4\pi i}\oint\frac{(\omega-\omega^{-1})^2}{(1-z\omega)^2(1-z\omega^{-1})^2}\frac{d\omega}{\omega} \\ &=-\frac{1}{4\pi i}\oint\frac{(\omega^2-1)^2}{(1-z\omega)^2(\omega-z)^2\omega}d\omega, \end{aligned}$$

and now, evaluating residues, one obtains

$$\begin{aligned} I^{(0)}&=2\pi i\left(\frac{-1}{4\pi i}\right)\left[\frac{1}{z^2}-\frac{(1+z^2)}{z^2(1-z^2)}\right]=\frac{1}{1-z^2} \\ &=\sum_{n=0}^{\infty}z^{2n}=\sum_n c_n^{(0)}z^n, \end{aligned}$$

so that now we have one scalar at each even order, and no scalars at odd orders.

As for $I^{(1)}$, and therefore $c_N^{(1)}$, recalling $T=T_1\oplus T_1$, we have to compute the multiplicity of T_1 in $P_n=(T^{\otimes n})_s$ and divide it by 2, as seen in the preceding discussion:

$$I^{(1)}=\sum_n c_n^{(1)}z^n=\frac{1}{2}\int_G\frac{\overline{\chi}^{(1)}(\alpha)}{\det(1-zT(\alpha))}d\mu_\alpha.$$

Using the same positions as before, one obtains

$$\begin{aligned} I^{(1)}&=\frac{1}{2}\frac{2}{\pi}\int_{-1}^{1}\frac{2\alpha}{(1+z^2-2z\alpha)^2}(1-\alpha^2)^{1/2}d\alpha \\ &=\frac{2}{\pi}\int_0^{\pi}\frac{\cos\alpha\sin^2\alpha}{(1+z^2-2z\cos\alpha)^2}d\alpha \\ &=\frac{1}{\pi}\int_0^{2\pi}\frac{\cos\alpha\sin^2\alpha}{(1+z^2-2z\cos\alpha)^2}d\alpha \\ &=\frac{-1}{8\pi i}\oint\frac{(\omega+\omega^{-1})(\omega-\omega^{-1})^2}{(1-z\omega)^2(1-z\omega^{-1})^2}\frac{d\omega}{\omega} \\ &=\frac{-1}{8\pi i}\oint\frac{(\omega^2-1)^2(\omega^2+1)}{(1-z\omega)^2(\omega-z)^2\omega^2}d\omega. \end{aligned}$$

Evaluation of this with the residue formula gives, with a short computation,

$$\begin{aligned} I_{(1)}&=2\pi i\left(\frac{-1}{8\pi i}\right)\left[\frac{2(1+z^2)}{z^3}-\frac{2(1+z^4)}{(1-z^2)z^3}\right] \\ &=\frac{z}{1-z^2}=\sum_{n=0}^{\infty}z^{2n+1}. \end{aligned}$$

Therefore we have, recalling now $D=4$, four independent vectors at each odd order, and no vectors at even orders. That is, even in this case, the gradient property is satisfied at any order.

As for the possibility of writing the bifurcation equation in gradient form, the same remarks as in the SO(2) case apply, with the role of the operator K played by the three operators K_1, K_2, K_3.

[1] D. H. Sattinger, "Group theoretic methods in bifurcation theory," *Lecture Notes in Mathematics*, Vol. 762 (Springer, Berlin, 1979).

[2] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1976).

[3] G. Gaeta and P. Rossi, "Gradient property of bifurcation equation for systems with rotational symmetry," J. Math. Phys. **25**, 1671 (1984).

[4] M. Golubitsky, "The Benard problem, symmetry and the lattice of isotropy subgroups," in *Bifurcation Theory Mechanics and Physics*, edited by C. P. Borter (Reidel, Dordrecht, 1983).

[5] G. Cicogna, "Quaternionic bifurcation," J. Phys. A **18**, L829 (1985).

A unified treatment of SU(N) inner and outer multiplicities

Jeffrey R. Schmidt
University of Wisconsin—Madison, Physics Department, Madison, Wisconsin 53706

(Received 3 June 1985; accepted for publication 16 September 1985)

A combinatorial approach is developed for calculation of weight multiplicities at or near the center of a weight diagram. The result may be used to determine the inner multiplicity of such a weight, or to decompose the product of several representations into irreducible summands.

I. INTRODUCTION

Inner and outer multiplicities of weights in the groups SU(N) may be calculated by a variety of methods. For direct product decompositions in SU(N), physicists usually exploit the similarity of the symmetric group representation theory to that of SU(N), and use direct multiplication of the Young tableaux.[1] However, this rapidly becomes tedious if we wish to reduce the product of several representations to a sum of irreducibles. The more Lie-theoretic methods involve a double sum over the Weyl group, or else complete knowledge of all weights and multiplicities of the factor representations. Even then, products of three or more representations must be reduced pairwise.

In this paper we shall develop a method of extracting multiplicities, both inner and outer, by combinatorial means. We do not distinguish the product of two representations from the product of an arbitrary number. The basis of our algorithm is the character theory of SU(N). We make use of the operator techniques from the theory of distributions.

II. DEFINITIONS

We have the following expressions for inner and outer multiplicities, respectively,[2,3]

$$\gamma_\lambda(\mu) = \left(\frac{1}{2\pi}\right)^{n-1} \int_0^{2\pi} \chi_\lambda(\phi_1,\ldots,\phi_{n-1}) e^{-i\mu\cdot\phi}\, d\phi_1\cdots d\phi_{n-1}, \tag{2.1}$$

$$\Gamma_{\lambda\otimes\lambda'}(\lambda'') = \frac{1}{n!}\int_0^{2\pi} \Delta\bar{\Delta}\chi_\lambda\chi_{\lambda'}\bar{\chi}_{\lambda''}\, d\phi_1\cdots d\phi_{n-1}, \tag{2.2}$$

$$\Delta = \prod_{i<j}(\epsilon_i - \epsilon_j), \quad \epsilon_j = e^{2\pi i\phi_j}, \tag{2.3}$$

$$\chi_\lambda(\phi_1,\ldots,\phi_{n-1}) = \chi_\lambda(\epsilon_1,\ldots,\epsilon_n). \tag{2.4}$$

The parameters $\epsilon_1,\epsilon_2,\ldots,\epsilon_n$ are the eigenvalues of $U\in$SU(N). Since U is unimodular we have

$$\prod_{i=1}^{n}\epsilon_i = 1. \tag{2.5}$$

Therefore we choose the first $n-1$ eigenvalues to be independent and

$$\epsilon_n = \prod_{i=1}^{n-1}\epsilon_i^{-1} = \prod_{i=1}^{n-1}\bar{\epsilon}_i. \tag{2.6}$$

Since χ_λ and $\Delta\bar{\Delta}$ can be written in terms of polynomials in quantities that are complex exponentials, it is clear that the integrands in (2.1) and (2.2) will be a sum of complex exponentials. Integration will give delta functions. The only nonzero contributions will then come from the constant term in the integrands. In general the characters are complicated functions and integration difficult. Instead of integrating, any operator, which when applied to the integrand extracts the constant part, may be used as a substitute.

Let $U = \mathrm{diag}(\epsilon_1,\epsilon_2,\ldots,\epsilon_n) \in$ SU(n) and consider a representation with Young frame

(2.7)

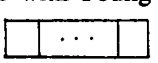
Let X_i, $i = 1,2,\ldots;n$ be the basis in which U has the diagonal form above. Then a basis of U in the representation with the mentioned Young frame is

$$\{X_{i_1}X_{i_2}\cdots X_{i_p} | i_1 < i_2 < \cdots < i_p\}. \tag{2.8}$$

Then, since $U: X_i \to \epsilon_i X_i$ (no sum) the character is

$$\mathrm{Tr}(U) = \sum_{i_1<i_2<\cdots<i_p} \epsilon_{i_1}\epsilon_{i_2}\cdots\epsilon_{i_p}$$
$$= a_p(\epsilon_1,\ldots,\epsilon_n) = \chi_{(1,1,\ldots,1,0,\ldots,0)}(U). \tag{2.9}$$

The representation of U with Young frame



(2.10)

has character

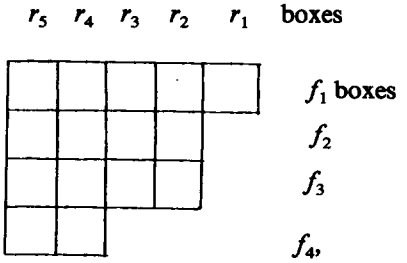
$$\chi_{(p,0,\ldots,0)}(U) = \mathrm{Tr}(U) = \sum_{i_1<i_2<\cdots<i_p} \epsilon_{i_1}\cdots\epsilon_{i_p} = h_p(\epsilon_1,\ldots,\epsilon_n). \tag{2.11}$$

The functions a_p and h_p are symmetric functions[4] with the following generators:

$$\prod_{i=1}^{n}\frac{1}{(1-\epsilon_i X)} = \sum_{k=0}^{\infty} x^k h_k(\epsilon_1,\ldots,\epsilon_n), \tag{2.12}$$

$$\prod_{i=1}^{n}(1-\epsilon_i X) = \sum_{k=0}^{n}(-1)^k X^k a_k(\epsilon_1,\ldots,\epsilon_n).$$

The characters of arbitrary representations may be found in terms of the a and h functions by Weyls' second formula[2,5]

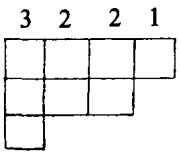


(2.13)

 0022-2488/86/020451-04$02.50 © 1986 American Institute of Physics

$$\chi(f_1,\ldots,f_n)=\det(h_{f_i-i+j}),\tag{2.14}$$
$$\chi(r_1,\ldots,r_m)=\det(a_{r_i-j+i}).$$

Example:



$$\chi=\begin{vmatrix}h_4 & h_2 & 0\\ h_5 & h_3 & h_0\\ h_6 & h_4 & h_1\end{vmatrix}=\begin{vmatrix}a_1 & a_0 & 0 & 0\\ a_3 & a_2 & a_1 & a_0\\ a_4 & a_3 & a_2 & a_1\\ a_6 & a_5 & a_4 & a_3\end{vmatrix}.\tag{2.15}$$

Define the following operators with respect to the symmetric functions:

$$d_1=\frac{d}{da_1}+a_1\frac{d}{da_2}+a_2\frac{d}{da_3}+\cdots,$$
$$\vdots\tag{2.16}$$
$$d_s=\frac{d}{da_s}+a_1\frac{d}{da_{s+1}}+a_2\frac{d}{da_{s+2}}+\cdots,$$
$$D_s=\frac{1}{s!}(d_1^s).\tag{2.17}$$

The operator D_s is called the Hammond operator,[4,6] which has the following actions on a symmetric function $\phi(x_1,x_2,\ldots,x_n)$: (1) transform $\phi(x_1,\ldots,x_n)\to\phi(x_1,\ldots,x_n,x_{n+1})$, (2) differentiate s times with respect to x_{n+1}, and (3) divide by $s!$ and set $x_{n+1}=0$. The functions a_1, a_2,... are independent, so that by the fundamental theorem on symmetric functions, there is a unique function ϕ', so

$$\phi(x_1,\ldots,x_n)=\phi'(a_1,\ldots,a_m),\quad a_i=a_i(x_1,\ldots,x_n).\tag{2.18}$$

For certain special choices of ϕ, one may deduce the following actions:

$$D_sh_p=h_{p-s},\quad D_sh_ph_q=\sum_{i=0}h_{p-i}h_{q-s+i},$$
$$D_1a_p=a_{p-1},\quad D_sa_p=0,\quad \text{for } s>1,\tag{2.19}$$
$$D_sa_{i_1}\cdots a_{i_m}=0,\quad s>m,\quad D_sa_{i_1}\cdots a_{i_s}=a_{i_1-1}\cdots a_{i_s-1}.$$

Furthermore by symmetry

$$D_{i_1}D_{i_2}\cdots D_{i_l}a_{j_1}\cdots a_{j_m}=D_{j_1}\cdots D_{j_m}a_{i_1}\cdots a_{i_l}.\tag{2.20}$$

The Hammond operator $D_{m_1}D_{m_2}\cdots D_{m_n}$ will give the number of times that the monomial $x_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$ occurs in the function ϕ when applied to ϕ. Hence we replace the integrals in (2.1) and (2.3) by Hammond operators acting on the integrals expressed as symmetric functions[3]

$$\gamma_{(f_1,\ldots,f_n)}(m_1,\ldots,m_n)=D_{m_1}D_{m_2}\cdots D_{m_n}\det M,\tag{2.21}$$

where M has form (2.14) or (2.15). We now specialize to the inner and outer multiplicities of central weights, either the zero weight or the fundamental weights. Here SU(n) has $(n-1)$ fundamental weights of the form

$$(\underbrace{1,1,\ldots,1}_{p},0,\ldots,0)\quad\text{with frame}\quad\begin{array}{|c|}\hline\\\hline\vdots\\\hline\\\hline\end{array}\quad p\text{ boxes}$$

and character $a_p(\epsilon_1,\ldots,\epsilon_n)$. The number of occurrences of this representation as an irreducible summand in the tensor product $\lambda\otimes\lambda'$ is

$$D_x^n((\Delta\bar{\Delta}/n!)\chi_\lambda\chi_\lambda\cdot\bar{a}_p),\tag{2.22}$$

where x is chosen to exhaust the expression in parentheses. Note that from (2.16) it follows that

$$\bar{a}_p(\epsilon_1,\ldots,\epsilon_n)=a_p(\bar{\epsilon}_1,\ldots,\bar{\epsilon}_n)=a_{n-p}(\epsilon_1,\ldots,\epsilon_n).\tag{2.23}$$

The Weyl measure $\Delta\bar{\Delta}$ can also be written as a polynomial in symmetric functions by recognizing its occurrence in the theory of equations.[7] First, since

$$\Delta=\prod_{i<j}(\epsilon_i-\epsilon_j),$$
$$\bar{\Delta}=\prod_{i<j}^{n}\Big(\prod_{k\neq i}\epsilon_k-\prod_{k'\neq j}\epsilon_{k'}\Big)$$
$$=a_n^{n-1}(\epsilon_1,\ldots,\epsilon_n)(-1)^{n(n-1)/2}\Delta$$
$$=(-1)^{n(n-1)/2}\Delta.\tag{2.24}$$

Hence the measure becomes

$$\Delta\bar{\Delta}=(-1)^{n(n-1)/2}\Delta^2.\tag{2.25}$$

Since Δ is an alternating homogeneous function, its square is symmetric and homogeneous. The Weyl measure is in fact, up to a sign, the discriminant of the polynomial whose roots are $\epsilon_1,\epsilon_2,\ldots,\epsilon_n$. Let

$$f(x)=\prod_{i=1}^{n}(x-\epsilon_i),\tag{2.26}$$
$$f'(\epsilon_1)=(\epsilon_1-\epsilon_2)\cdots(\epsilon_1-\epsilon_n),$$
$$f'(\epsilon_2)=(\epsilon_2-\epsilon_1)\cdots(\epsilon_2-\epsilon_n),$$
$$\vdots\tag{2.27}$$
$$f'(\epsilon_n)=(\epsilon_n-\epsilon_1)\cdots(\epsilon_n-\epsilon_{n-1}),$$
$$\prod_{i=1}^{n}f'(\epsilon_i)=(-1)^{n(n-1)/2}(\epsilon_1-\epsilon_2)^2\cdots(\epsilon_{n-1}-\epsilon_n)^2$$
$$=\Delta\bar{\Delta}=R(f,f'),\tag{2.28}$$

where $R(f,f')$ is Sylvesters' form of the discriminant[7] of $f(x)$.

Examples:

$$f(x)=(x-\epsilon_1)(x-\epsilon_2)=a_0x^2+a_1x+a_2,$$
$$R(f,f')=\begin{vmatrix}a_0 & a_1 & a_2\\ 2a_0 & a_1 & 0\\ 0 & 2a_0 & a_1\end{vmatrix}=4a_2-a_1^2,\tag{2.29}$$
$$f(x)=(x-\epsilon_1)(x-\epsilon_2)(x-\epsilon_3)$$
$$=a_0x^3+a_1x^2+a_2x+a_3,$$
$$R(f,f')=\begin{vmatrix}a_0 & a_1 & a_2 & a_3 & 0\\ 0 & a_0 & a_1 & a_2 & a_3\\ 3a_0 & 2a_1 & a_2 & 0 & 0\\ 0 & 3a_0 & 2a_1 & a_2 & 0\\ 0 & 0 & 3a_0 & 2a_1 & a_2\end{vmatrix}$$
$$=27a_3^2+4a_2^3-a_1^2a_2^2+4a_1^3a_3-18a_1a_2a_3.\tag{2.30}$$

To summarize

$$\Delta\bar{\Delta} = R(f,f'),$$

for SU(n) with

$$f(X) = \prod_{i=1}^{n}(X-\epsilon_i) = \sum_{i=0}^{n} a_i X^{n-i}.$$

An application is now in order. We solve the following problem: How many spin-0 representations occur as irreducible summands in the tensor product of N spin-$\frac{1}{2}$ particle representations? For SU(2) the previous analysis shows

$$\Delta\bar{\Delta}/2! = 2a_2 - \tfrac{1}{2}a_1^2. \tag{2.31}$$

Hence the number in question is

$$D^2_{N/2+1}(2a_2 - \tfrac{1}{2}a_1^2)a_1^N, \tag{2.32}$$

which is clearly zero if N is odd. Explicit evaluation using the facts that $a_1 = \epsilon_1 + \epsilon_2$, $a_2 = \epsilon_1\epsilon_2$ shows

$$D^2_{N/2+1}(2a_2a_1^N - \tfrac{1}{2}a_1^{N+2}) = \begin{cases} 0, & N \text{ odd}, \\ 2\binom{N}{N/2} - \frac{1}{2}\binom{N+2}{N/2+1}, & N \text{ even}. \end{cases} \tag{2.33}$$

A trivial case that does not require the operator approach is the direct sum decomposition of the tensor product of m defining fundamental representations of SU(n). The coefficient of each irreducible summand is clearly just the number of standard tableaux associated with the Young frame that labels the representation[1]

$$(\otimes\,\square)^m = \sum_{f_1+\cdots+f_n=m,\; f_1\ge f_2\ge\cdots\ge f_n} (f_1,\ldots,f_n)\left\{\frac{m!\prod_{i<j}^{n}(f_i-f_j+j-i)}{\prod_{i=1}^{n}(f_i+n-i)!}\right\}. \tag{2.34}$$

The number of trivial representations contained in the direct sum decomposition of $(\otimes\,\square)^{Nm}$ in SU(N) is

$$\frac{(Nm)!\prod_{k=0}^{m-1}(k)!}{\prod_{k=0}^{m-1}(N+k)!}. \tag{2.35}$$

III. ITERATIVE SOLUTIONS AND MULTIPLICITY FORMULAS

Decompose a_m^2 into monomial symmetric functions of the form

$$2^p1^q \equiv \sum_{i_1\ne i_2\ne\cdots\ne i_{p+q}} x_{i_1}^2\cdots x_{i_p}^2 x_{i_{p+1}}\cdots x_{i_{p+q}}. \tag{3.1}$$

We note first of all that the only monomial symmetric functions that occur in the decomposition of a_m^p are those of the form

$$1^{j_1}2^{j_2}\cdots p^{j_p}, \tag{3.2}$$

with $j_1 + 2j_2 + \cdots + p\cdot j_p = mp$ by (2.19) and (2.20). Then in (3.1) we must have $2m = 2p + q$. Starting with $m = 1$ and building up the series using

$$D^2_m a_2^p a_1^q = D^2_{m-p} a_1^q \tag{3.3}$$

together with the multinomial theorem, the result is

$$D^2_m a_2^p a_1^q = q!/(q/2)!(q/2)!. \tag{3.4}$$

From this it is deduced that coefficients of the form

$$D^3_m a_1^{p_1} a_2^{p_2}$$

are sufficient to generate the entire monomial decomposition of a_m^3. This is due to the fact that

$$D^3_m a_1^{p_1}a_2^{p_2} = D^{p_1}_1 D^{p_2}_2 a_m^3 = D^{p_1}_1 D^{p_2}_2 D^{p_3}_3 a^3_{m+p_3} = D^3_{m+p_3} a_1^{p_1}a_2^{p_2}a_3^{p_3}, \tag{3.5}$$

$$D^3_m a_1^{p_1}a_2^{p_2} = D^2_m \sum_{m_1+m_2=m}\binom{p_1}{m_1}\binom{p_2}{m_2} a_1^{p_1+m_2-m_1}a_2^{p_2-m_2} = D^2_m \sum_{m_1=0}^{m}\binom{p_1}{m_1}\binom{p_2}{m-m_1} a_1^{p_1+m-2m_1}a_2^{p_2-m+m_1} = \sum_{m_1=0}^{m}\binom{p_1}{m_1}\binom{p_2}{m-m_1}\frac{(p_1+m-2m_1)!}{\{(p_1+m)/2-m_1)!\}^2}. \tag{3.6}$$

Introduce the notation

$$D^2_m a_2^p a_1^q = \begin{Bmatrix}(2)^p & (1)^q\\ & (m)^2\end{Bmatrix} = \begin{Bmatrix}(1)^q\\ (m)^2\end{Bmatrix},$$

$$D^3_m a_2^p a_1^q = \begin{Bmatrix}(2)^p & (1)^q\\ & (m)^3\end{Bmatrix}, \tag{3.7}$$

$$D^4_m a_3^r a_2^p a_1^q = \begin{Bmatrix}(3)^r & (2)^p & (1)^q\\ & & (m)^4\end{Bmatrix},$$

and one obtains the recursion formula

$$\begin{Bmatrix}(1)^{p_1} & (2)^{p_2}\\ & (m)^3\end{Bmatrix} = \sum_{m_1=0}\binom{p_1}{m_1}\binom{p_2}{m-m_1}\begin{Bmatrix}(1)^{p_1+m-2m_1}\\ (m)^2\end{Bmatrix}. \tag{3.8}$$

Consider now $D^4_m a_1^{p_1}a_2^{p_2}a_3^{p_3}$, a term that occurs in the problem for SU(4):

$$D^4_m a_1^{p_1}a_2^{p_2}a_3^{p_3} = D^3_m \sum_{m_1+m_2+m_3=m}\binom{p_1}{m_1}\binom{p_2}{m_2}\binom{p_3}{m_3} a_1^{p_1-m_1+m_2}a_2^{p_2-m_2+m_3}a_3^{p_3-m_3} = \sum\binom{p_1}{m_1}\binom{p_2}{m_2}\binom{p_3}{m_3} D^3_{m+m_3-p_3} a_1^{p_1-m_1+m_2}a_2^{p_2-m_2-m_3} = \sum_{m_1+m_2+m_3=m}\binom{p_1}{m_1}\binom{p_2}{m_2}\binom{p_3}{m_3}\begin{Bmatrix}(1)^{p_1-m_1+m_2} & (2)^{p_2-m_2+m_3}\\ & (m+m_3-p_3)^3\end{Bmatrix}. \tag{3.9}$$

Special cases occur when a p_i is zero, an example being

$$D^4_m a_2^{p_2}a_3^{p_3} = \sum_{m_2=0}\binom{p_2}{m_2}\binom{p_3}{m-m_2}\begin{Bmatrix}(1)^{m_2} & (2)^{p_2+m-m_2}\\ & (2m-p_3-m_2)^3\end{Bmatrix}. \tag{3.10}$$

It should be noted that in (3.9) only a few terms in general will contribute, since for $x > y$ the binomial coefficient $\binom{y}{x}$ is zero. It is also clear that the last three formulas in the appendix are easy consequences of (3.9). The process of (3.9) can be iterated to obtain the monomial decomposition coefficients for a symmetric function a_m raised to any power p in terms of the coefficients for $p-1, p-2,\ldots,2,1$. It is worth noting that for SU(3), the decomposition of the product of p quark and p antiquark representations to singlets involves the coefficient

TABLE I. Values of functions T, S, Γ.

p	$T(p)$	$S(p)$	$\Gamma_p(0)$
0	1	6	1
1	3	36	1
2	15	240	2
3	93	1 710	6
4	639	12 726	23
5	4 653	97 608	103
6	35 169	765 288	513
7	272 835	·	·
		·	·
8	2157 759	·	·

$$\sum_{k=0}^{p}\binom{p}{k}^2\binom{2p-2k}{p-k}=\sum_k\sum_q\binom{p}{k}\binom{p}{k}\binom{p-k}{q}\binom{p-k}{q}$$

$$=\sum_n\sum_q\frac{p!p!}{k!k!q!q!(p-k-q)!(p-k-q)!}$$

$$=\sum_{k,q}\left\{\frac{p!}{k!q!(p-k-q)!}\right\}^2, \tag{3.11}$$

which is the sum of the squares of the trinomial coefficients. Also in SU(2), we note that

$$\frac{2p!}{p!p!}=\binom{2p}{p}=\sum_{k=0}^{p}\binom{p}{k}^2. \tag{3.12}$$

Using (2.30), (3.6), and (3.11), the number of irreducible singlets contained as summands in the tensor product of p quark and p antiquark representations of SU(3) is

$$\Gamma_{(\square\otimes\bar{\square})^p}(0)=\tfrac{9}{2}T(p)-\tfrac{1}{6}T(p+2)-3T(p+1)+\tfrac{4}{3}S(p)$$

$$\equiv\Gamma_p(0), \tag{3.13}$$

where

$$T(p)=\sum_{q=0}^{p}\binom{p}{q}^2\binom{2p-2q}{p-q}$$

and (3.14)

$$S(p)=\sum_{q=0}^{p}\binom{p}{q}\binom{p+3}{q+1}\binom{2p+2-2q}{p+1-q}.$$

A few of these numbers are tabulated in Table I.

Arbitrary tensor products may be decomposed into irreducible summands by substituting their character determinants (2.14) or (2.15) in terms of symmetric functions together with the symmetric form of $\Delta\bar{\Delta}$ into (2.22). This method has the advantage over Klimyk's and Steinberg's formulas[1] in that products of two or more representations may be reduced without having to reduce the factors pairwise. The Weyl group of SU(n) is of order $n!$ so that Steinberg's formula involves a sum over $(n!)^2$ terms, each of which involves evaluation of a partition function. Klimyk's formula requires complete knowledge of all weights, multiplicities and their stabilizers in the Weyl group. Our present method requires much less information and calculation.

APPENDIX: COMBINATORIAL FORMULAS IN TERMS OF HAMMOND OPERATORS

Multinomial theorem

$$(\epsilon_1+\cdots+\epsilon_n)^m=\sum_{i_1\cdots i_n}^{i_1+\cdots+i_n=m}\frac{m!}{(i_1)!(i_2)!\cdots(i_n)!}\epsilon_1^{i_1}\epsilon_2^{i_2}\cdots\epsilon_n^{i_n}$$

in symmetric function notation

$$(1)^m=\sum\frac{m!}{(\lambda!)^{l_\lambda}\cdots(\mu!)^{l_\mu}\cdots}(\lambda)^{l_\lambda}\cdots(\lambda)^{l_\mu}\cdots,$$

$$D_{m_1}D_{m_2}\cdots D_{m_n}a_1^m$$

$$=m!/m_1!m_2!\cdots m_n!,\quad m_1+\cdots+m_n=m,$$

$$D_p^q a_1^m=m!/(p!)^q,\quad pq=m,$$

$$D_n^m a_q a_1^p=\binom{m}{q}(p!)/\{(n-1)!\}^q(n!)^{m-q},\quad nm=q+p.$$

[1] A. O. Barut, and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific, Warsaw, 1977).
[2] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, NJ, 1946).
[3] J. R. Schmidt, J. Math. Phys. **26**, 2721 (1985).
[4] P. A. MacMahan, *Combinatory Analysis*, Vol. I (Cambridge U.P., Cambridge, England, 1915).
[5] E. D'Hoker, J. Math. Phys. **25**, 1 (1984).
[6] F. N. David, M. G. Kendall, and D. E. Barton, *Symmetric Functions and Allied Tables* (Cambridge, New York, 1966).
[7] J. V. Uspensky, *Theory of Equations* (McGraw-Hill, New York, 1948).

A simple construction of twist-eating solutions

Pierre van Baal
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794
Bert van Geemen
Department of Mathematics, Budapestlaan 6, P. O. Box 80010, 3508 TA Utrecht, The Netherlands

(Received 7 May 1985; accepted for publication 30 September 1985)

A simple general construction of all solutions to the set of equations $[\Omega_\mu, \Omega_\nu] = \exp(2\pi i n_{\mu\nu}/N)I$, where $\Omega_\mu \in SU(N)$ or $U(N)$ and $\mu, \nu = 1, 2, \ldots, 2g$, is given.

I. REDUCTION TO A CANONICAL FORM

Twisted guage fields on the hypertorus, both in the continuum[1] and on the lattice,[2] posed the interesting mathematical problem of finding matrices Ω_μ in $SU(N)$ or $U(N)$ (called twist-eating solutions), such that

$$[\Omega_\mu, \Omega_\nu] = \Omega_\mu\, \Omega_\nu \Omega_\mu^{-1}\, \Omega_\nu^{-1} = \exp(2\pi i n_{\mu\nu}/N)I. \tag{1}$$

Here n is called the twist tensor; it is skew symmetric with integer entries mod N. The index μ runs from 1 up to $2g$ (the dimension of space-time; odd dimensions need not be considered separately). For details see Refs. 3 and 4, where the full solution of this problem for $g<2$ was found (see also Ref. 5).

By means of a Sl $(2g, \mathbf{Z})$ transformation X, we can always transform n to its standard[3] form n^s:

$$n^s = \begin{pmatrix} \emptyset & & e_1 & & \\ & \ddots & & & \\ & & & & e_g \\ -e_1 & & & \ddots & \\ & & & \emptyset & \\ & & -e_g & & \end{pmatrix}, \tag{2}$$

where $e_1|e_2|\cdots|e_g$ and $n = {}^tXn^sX$. (For integer p and q the symbol $p|q$ means that p divides q.) If $[\tilde{\Omega}_\mu, \tilde{\Omega}_\nu] = \exp(2\pi i n^s_{\mu\nu}/N)I$, then Eq. (1) is solved by

$$\Omega_\nu = \prod_\nu \tilde{\Omega}_\nu^{X_{\nu\mu}}. \tag{3}$$

The standard form n^s is not unique since we can add a multiple of N to each $n_{\mu\nu}$. However, transformation (3) is invertible[4]; the specific choice of n^s is therefore irrelevant. To be precise,

$$\tilde{\Omega}_\mu = Z_\mu \prod_\nu \Omega_\nu^{(X^{-1})_{\mu\nu}},$$

with Z_μ an element of the center of $SU(N)$, depending only on n and X.

Define

$$f_j = \gcd(e_j, N), \quad N_j = N_{g+j} = N/f_j, \quad j = 1, 2, \ldots, g. \tag{4}$$

(Greek indices will always run from 1 up to $2g$ and Latin indices from 1 up to g; gcd = greatest common divisor.) From the commutation relations it follows that

$$[\tilde{\Omega}_j^{N_j}, \tilde{\Omega}_{g+j}] = [\tilde{\Omega}_j, \tilde{\Omega}_{g+j}^{N_j}] = I.$$

Hence, the $\tilde{\Omega}_\mu^{N_\mu} \in SU(N)$ or $U(N)$ commute, so they can be simultaneously diagonalized. Let $A \in SU(N)$ be such that the

$$W_\mu = A\tilde{\Omega}_\mu^{N_\mu} A^{-1} \tag{5}$$

are diagonal matrices. As $[W_\mu, A\tilde{\Omega}_\mu A^{-1}] = I$ for all μ, ν we can choose diagonal matrices Λ_μ such that

$$\Lambda_\mu^{N_\mu} = W_\mu \text{ and } [\Lambda_\mu, A\Omega_\nu A^{-1}]. \tag{6}$$

If we define

$$\Omega'_\mu = \Lambda_\mu^{-1} A\tilde{\Omega}_\mu A^{-1}, \tag{7}$$

then the Ω'_μ satisfy

$$[\Omega'_\mu, \Omega'_\nu] = \exp(2\pi i n^s_{\mu\nu}/N)I, \quad (\Omega'_\mu)^{N_\mu} = I. \tag{8}$$

Next we will further simplify these commutation relations.

Recall that $\gcd(e_j/f_j, N_j) = 1$; hence there exist integers M_j such that

$$M_j(e_j/f_j) \equiv 1 \pmod{N_j}. \tag{9}$$

Define

$$U_j = (\Omega'_j)^{M_j}, \quad U_{g+j} = \Omega'_{g+j}. \tag{10}$$

This transformation can also be inverted: $\Omega'_j = U_j^{(e_j/f_j)}$, where we used that $(\Omega'_j)^{N_j} = I$. As $[U_j, U_{g+j}] = [(\Omega'_j)^{M_j}, \Omega'_{g+j}] = \exp(2\pi i e_j M_j/N)I$ and $e_j M_j/N = M_j(e_j/f_j)/N_j = N_j^{-1} \pmod{\mathbf{Z}}$, we see that the U_μ satisfy the commutation relations (1) with a twist tensor m in standard form:

$$m = \begin{pmatrix} & & f_1 & & \\ & \emptyset & & \ddots & \\ & & & & f_g \\ -f_1 & & & \emptyset & \\ & \ddots & & & \\ & & -f_g & & \end{pmatrix}. \tag{11}$$

(Note that $f_1|f_2|\cdots|f_g$ and moreover each f_j divides N.) In particular,

$$[U_j, U_{g+j}] = \exp(2\pi i N_j^{-1}), \quad (U_\mu)^{N_\mu} = I. \tag{12}$$

Hence to find all solutions to Eq. (1) it suffices to determine all solutions to Eq. (12).

II. THE GENERAL SOLUTION FOR THE CANONICAL FORM

Theorem: There exist matrices $U_\mu \in Gl(N)$ satisfying Eq. (12) if and only if $N_1 N_2 \cdots N_g$ divides N, where $N_i = N/f_i$.

Proof: Note that the subgroup K of $Gl(N)$ generated by

 0022-2488/86/020455-03$02.50 © 1986 American Institute of Physics

U_μ is finite. Moreover the U_j $(1<j<g)$ generate an Abelian subgroup, in particular there is a basis of $\mathbb{C}^N$ consisting of simultaneous eigenvectors for all U_j $(1\leqslant j<g)$. Let v be such a basis vector and assume $U_j v=\exp(2\pi i a_j/N_j)v$. Then $U_j\,(U_{g+j}v)=\exp(2\pi i(a_j+1)/N_j)U_{g+j}v$, hence the vectors Kv span a subspace V of dimension $N_1N_2\cdots N_g$. It is easy to see that K acts irreducibly on V. Proceeding with this method in the K-invariant complementary subspace of V we see that $\mathbb{C}^N$ is a direct sum of k K-invariant subspaces, each of dimension $N_1N_2\cdots N_g$. So $N=kN_1N_2\cdots N_g$.

To prove the converse, let V be a vector space of dimension $N_1N_2\cdots N_g$ with a basis $e(b_1,b_2,\ldots,b_g)$, with $b_j\in\mathbb{Z}/N_j\mathbb{Z}$. Define linear maps $U'_\mu: V\to V$ by

$$\begin{aligned}U'_j e(b_1,b_2,\ldots,b_g)&=\exp(2\pi i b_j/N_j)e(b_1,b_2,\ldots,b_g),\\ U'_{g+j}e(b_1,\ldots,b_j,\ldots,b_g)&=e(b_1,\ldots,b_j+1,\ldots,b_g).\end{aligned}\tag{13}$$

It is easy to check that the U'_μ satisfy Eq. (12). Now assume $N=kN_1N_2\cdots N_g$, then $\mathbb{C}^N\cong V^k$ and define $U_\mu\in\mathrm{Gl}(N)$ by the block diagonal sum of k copies of U'_μ. Then obviously the U_μ also satisfy Eq. (12). ∎

We point out that the finite group K generated by the U_μ is a Heisenberg group. All irreducible representations were constructed in Ref. 6. Solutions of Eq. (12) form representations ρ of K, which, when restricted to the center $C(K)$ $(=\{\lambda I\,|\,\lambda^{N_1}=1\}\simeq Z_{N_1})$ of K, is given by $\rho(c)=c$, $\forall c\in C(K)$. This implies that each irreducible component of ρ has to be the unique so-called Schrödinger representation[6] [Eq. (13)]. Hence, ρ is unique up to a similarity transformation.

More directly, following closely the above proof of the theorem, it is easily seen that for $k=1$, $e(b_1,b_2,\ldots,b_g)$ and $U_{g+j}^{(b_j-a_j)}v$ are to be identified. Similar statements for $k>1$ reproduce the block diagonal form, and two solutions to Eq. (12) have to be equivalent, i.e., $\exists A\in\mathrm{SU}(N)$, $U_\mu^{(2)}=AU_\mu^{(1)}A^{-1}$, $\forall\mu$. We will conclude this note with a few remarks.

The U_μ are unitary matrices. The explicit matrices for Eq. (13) are given by

$$\begin{aligned}U'_j&=1_{N_1}\otimes\cdots\otimes Q_{N_j}\otimes\cdots\otimes 1_{N_g},\\ U'_{g+j}&=1_{N_1}\otimes\cdots\otimes P_{N_j}\otimes\cdots\otimes 1_{N_g},\end{aligned}\tag{14a}$$

with

$$Q_n=\mathrm{diag}(1,e^{2\pi i/n},\ldots,e^{2\pi i(n-1)/n}),$$

$$P_n=\begin{pmatrix}0 & 1 & & 0\\ & \ddots & & \\ & & 0 & \\ \vdots & & & 1\\ & & \ddots & \\ 1 & & \cdots & 0\end{pmatrix}.\tag{14b}$$

This establishes the relation with the previous constructions.[3–5]

A solution Ω_μ to the original Eq. (1) is clearly specified by $A\in\mathrm{SU}(N)$ and Λ_μ, a diagonal unitary matrix [see Eqs. (5) and (6)], together with U'_μ [see Eqs. (13) and (14)]. Equation (6) implies that Λ_μ is a multiple of the identity in each block of U'_μ: $\Lambda_\mu=\mathrm{diag}\,(\lambda_\mu^{(1)}I,\ldots,\lambda_\mu^{(k)}I)$, with I the $N_1N_2\cdots N_g$-dimensional identity matrix. Hence, the pair (A,Λ_μ) forms the group $G=\mathrm{SU}(N)\times\mathrm{U}(1)^k$. On the other hand, the uniqueness of solutions to Eq. (12) guarantees that for each $\{\Lambda_\mu\}$ satisfying $\Lambda_\mu^{N_\mu}$ for all μ, there exists an (in general not unique) $A\in\mathrm{SU}(N)$ such that for all μ,

$$\Lambda_\mu U'_\mu=AU'_\mu A^{-1},\quad \Lambda_\mu^{N_\mu}=I.\tag{15}$$

[This can be explicitly verified for[3] Eq. (14).] Equation (15) specifies a subgroup H of G. The solutions to Eq. (1) are [for $\Omega_\mu\in\mathrm{U}(N)$] in 1–1 corresponding with G/H. These solutions are described by $2gk$ inequivalent continuous parameters [$2g(k-1)$ for $\Omega_\mu\in\mathrm{SU}(N)$]. A case of special interest is $k=1$ for $\Omega_\mu\in\mathrm{SU}(N)$, where the solution space for Eq. (1) modulo equivalence is discrete and isomorphic to $\Pi_{j=1}^g(Z_N/Z_{N_j})^2$, with $N^{2(g-1)}$ elements.

Suppose $N=k\,\Pi_{i=1}^g N_i$, define

$$m_i=-e_i/\gcd(e_i,N)=-(e_i/f_i).\tag{16}$$

Obviously both $n_{\mu\nu}$ and $N\cdot Pf(n/N)$ are multiples of k, since $e_i=-m_i k\,\Pi_{j\neq i}N_j$ and $N\cdot Pf(n/N)=-k\,\Pi_i m_i$. Consequently $N\cdot Pf(n/N)\in\mathbb{Z}$ is a necessary condition for existence of a solution to Eq. (1). Next observe that $\gcd(m_i,N_i)=1$ and $N_g\,|N_{g-1}|\cdots|N_1$. Hence $\gcd(m_i,N_j)=1$, for all $j\geqslant i$, so

$$\gcd(n_{\mu\nu},N\cdot Pf(n/N),N)=k\,\gcd\Big(\prod_{i=2}^g N_i,\prod_{i=2}^g m_i\Big).\tag{17}$$

Given a solution, it is clearly unique up to a similarity transformation and Z_N factors if and only if $k=1$. Hence $\gcd(n_{\mu\nu},N\cdot Pf(n/N),N)=1$ is a sufficient condition for uniqueness. For $g=2$ it is also necessary, as can be seen from Eq. (17) and $\gcd(m_2,N_2)=1$. Furthermore, in the case $g=2$, $N\cdot Pf(n/N)=-e_1e_2/N$. We can write $e_i=m_i f_i$, and $N=f_2c$ with $\gcd(m_i,c)=1$. Hence $N\cdot Pf(n/N)=-m_1m_2f_1/c\in\mathbb{Z}$ implies that f_1 is a multiple of c. So $N/N_1N_2=f_1f_2/N=f_1/c\in\mathbb{Z}$. Consequently for $g=2$, $N\cdot Pf(n/N)\in\mathbb{Z}$ is also sufficient for existence of solutions to Eq. (1).

That the above criteria [i.e., $N\cdot Pf(n/N)$ is sufficient for existence and $\gcd(n_{\mu\nu},N\cdot Pf(n/N),N)=1$ is necessary for uniqueness] cannot be extended beyond $g=2$ can be seen from the following two examples constructed by Coste[7]: (i) $g=3$, $N=2^23^6$, $e_1=e_2=3^4$, and $e_3=2^43^4$ (hence $N_1=N_2=2^23^2$ and $N_3=3^2$), so $N\cdot Pf(n/N)=e_1e_2e_3/N^2=1$ but $N_1N_2N_3=4N$ does not divide N, and no solution exists; and (ii) $g=3$, $N=2^27^3$, $e_1=e_2=2\cdot3\cdot7^2$, and $e_3=2^3\cdot3\cdot7^2$ (hence $N_1=N_2=2\cdot7$ and $N_3=7$), so $\gcd(n_{\mu\nu},N\cdot Pf(n/N),N)=2$, but $N_1N_2N_3=N$ and the solution is unique.

Note added in proof: After completion of this work, we received a preprint by Lebedev and Polikarpov.[8] Their results coincide with those of Ref. 6 and this paper.

ACKNOWLEDGMENTS

We are grateful to Professor T. A. Springer for pointing out the connection with the Heisenberg group and Antoine Coste for the counterexamples to the $g=2$ criteria when extended to $g>2$.

B. v. G. is financially supported by the ZWO project on moduli and P. v. B. is supported in part by National Science Foundation Grant No. PHY-81-09110A-03.

[1]G. 't Hooft, Nucl. Phys. B **153**, 141 (1979); Acta Phys. Austriaca Suppl. **22**, 531 (1980).
[2]J. Groeneveld, J. Jurkiewicz, and C. P. Korthals-Altes, Physica Scripta **23**, 1022 (1981); A. Gonzales-Arroyo and M. Okawa, Phys. Rev. D **27**, 2397 (1983).
[3]P. van Baal, Commun. Math. Phys. **92**, 1 (1983).
[4]P. van Baal, "Twisted boundary conditions: A non-perturbative probe for pure non-Abelian gauge theories," thesis Utrecht, July 1984.
[5]J. Ambjørn and H. Flyvbjerg, Phys. Lett. B **97**, 241 (1980); G. 't Hooft, Commun. Math. Phys. **81**, 267 (1981); P. van Baal, Commun. Math. Phys. **85**, 529 (1982); Y. Brihaye, Phys. Lett. B **122**, 154 (1983); Y. Brihaye, G. Maiella, and P. Rossi, Nucl. Phys. B **222**, 309 (1983).
[6]B. van Geemen and P. van Baal, "The group theory of twist-eating solutions," Stony Brook preprint ITP-SB-84-86, to appear in Proc. K. Ned. Akad. Wet, Ser. B.
[7]A. Coste (private communication).
[8]D. R. Lebedev and M. I. Polikarpov, "Extrema of the twisted Eguchi–Kawai action and the finite Heisenberg group," preprint ITEP-41, 1985.

Differentiation of retarded integrals and the divergence theorem for retarded functions with discontinuities

F. I. Cooperstock and P. H. Lim
Department of Physics, University of Victoria, Victoria, British Columbia, Canada V8W 2Y2

(Received 24 June 1985; accepted for publication 3 October 1985)

Theorems expressing the time derivatives of retarded volume and surface integrals are presented as well as the Gauss divergence theorem for retarded functions with discontinuities. These theorems greatly facilitate the analysis of gravitational radiation from the motion of disjoint matter distributions in general relativity and could find useful application in other branches of physics.

I. INTRODUCTION

In the course of our research on the two-body problem in general relativity,[1-3] which is concerned with the calculation of gravitational radiation emitted when two disjoint gravitationally interacting bodies accelerate in free-fall, certain interesting mathematical problems arose. In order to deduce the field and hence the radiation, it was necessary to consider time derivatives of the retarded volume integrals of functions with discontinuities as well as to employ the Gauss divergence theorem with such retarded functions. We are using the word "retarded" in the usual sense that at the field-point $\mathbf{r}$ at time t, contributions from source-points $\mathbf{r}'$ are to be evaluated at time $t-|\mathbf{r}-\mathbf{r}'|/c$ to allow for the propagation of signals with the speed of light.[4]

After developing methods to handle such integrals, we realized that the mathematical problems are of more general interest in that the wave equation and its causal retarded solutions, as well as the application to disjoint sources with discontinuities, are common to a variety of areas in physics. Accordingly, it was deemed desirable to present the following results, which we have found invaluable in our treatment and understanding of the two-body problem, as a separate study accessible to researchers in all areas of mathematical and theoretical physics. For completeness, we also include a treatment of time derivatives of surface integrals.

Consider a time-independent volume V, bounded by a surface S, in which scalar and vector functions $f(\mathbf{r},t)$ and $\mathbf{F}(\mathbf{r},t)$ are defined. Both f and $\mathbf{F}$ are continuous everywhere except on a closed time-dependent surface $D(t)$ within V. The (time-dependent) volume contained in the interior/exterior of D is denoted $V_{\rm in}(t)/V_{\rm ex}(t)$; in these regions, f is denoted $f_{\rm in}/f_{\rm ex}$. The velocity of an element of D is $\mathbf{u}(t)$. The theorems considered below divide into two classes: (A) theorems for volume integrals defined on a time slice $t=$ const, and (B) theorems for retarded integrals, which we develop in turn.

II. TIME DERIVATIVES OF VOLUME INTEGRALS

(A) We now demonstrate the following:

$$\frac{d}{dt}\int_{V_{\rm in}} f_{\rm in}\,dV=\int_{V_{\rm in}}\left(\frac{\partial f_{\rm in}}{\partial t}\right)dV+\oint_D d\mathbf{s}\cdot\mathbf{u}f_{\rm in}, \tag{1}$$

$$\frac{d}{dt}\int_{V_{\rm ex}} f_{\rm ex}\,dV=\int_{V_{\rm ex}}\left(\frac{\partial f_{\rm ex}}{\partial t}\right)dV-\oint_D d\mathbf{s}\cdot\mathbf{u}f_{\rm ex}. \tag{2}$$

Equation (1) follows from the assertion that $V_{\rm in}$ is t dependent. Suppressing spatial coordinates in the integrands, we have

$$\frac{d}{dt}\int_{V_{\rm in}} f_{\rm in}\,dV = \lim_{\delta t\to 0}\left\{\frac{\int_{V_{\rm in}(t+\delta t)} f_{\rm in}(t+\delta t)dV-\int_{V_{\rm in}(t)} f_{\rm in}(t)dV}{\delta t}\right\}. \tag{3}$$

Clearly, the right side of (3) has contributions from the following two spatial regions: (i) the region $V_{\rm overlap}$, comprising the intersection $V_{\rm in}(t+\delta t)\cap V_{\rm in}(t)$ [the contribution to (3) from this region in the limit is $\int_{V_{\rm in}(t)}(\partial f_{\rm in}/\partial t)dV$], and (ii) the region V^* comprising points lying inside $V_{\rm in}(t+\delta t)$ and $V_{\rm in}(t)$ but outside $V_{\rm overlap}$, i.e.,

$$V^*\equiv(V_{\rm in}(t)\cup V_{\rm in}(t+\delta t))-V_{\rm overlap}. \tag{4}$$

In the limit as $\delta t\to 0$, the contribution to (3) from this region takes the form

$$\oint_D d\mathbf{s}\cdot\mathbf{u}f_{\rm in}.$$

Addition of these contributions to (3) yields Eq. (1).

Equation (2) may be proved in a similar manner; the minus sign on the right side of (2) arising because $d\mathbf{s}$ points from $V_{\rm in}$ to $V_{\rm ex}$.

Finally, addition of (1) and (2) yields the useful relation

$$\frac{d}{dt}\int_V f\,dV=\int_V\left(\frac{\partial f}{\partial t}\right)dV+\oint_D d\mathbf{s}\cdot\mathbf{u}f|, \tag{5}$$

where

$$f|\equiv f_{\rm in}-f_{\rm ex} \tag{6}$$

on D. Note that in (5), $(\partial f/\partial t)$ is defined in V^* but not on D (see Ref. 5).

(B) We now discuss theorems corresponding to (1), (2), and (5) for retarded integrals.

First, it is important to develop a consistent notation for retarded integrals. The archetypical retarded integral corresponding to the integral on the left side of (1) is of the form

$$\phi(\mathbf{r},t)=\int_{[V_{\rm in}]} dV' f_{\rm in}(\mathbf{r}',\,t-R), \tag{7}$$

where

 0022-2488/86/020458-03$02.50 © 1986 American Institute of Physics

$$\mathbf{R} \equiv \mathbf{r} - \mathbf{r}'. \tag{8}$$

In (7), the region $[V_{\text{in}}]$ of integration is the intersection of the interior world tube INT, which at the time t defines $V_{\text{in}}(t)$, with the past light cone $\text{LC}(t)$ from the field point r at time t. More precisely, $[V_{\text{in}}]$ is designated $[V_{\text{in}}]_{\text{LC}(t)}$:

$$[V_{\text{in}}]_{\text{LC}(t)} = \text{INT} \cap \text{LC}(t). \tag{9}$$

In the following calculations, the prime is henceforth dropped from dV' in the integrand of (7) and the retarded integrand is denoted with square brackets $[f_{\text{in}}]$.

We now show that the retarded analogs of (1), (2), and (5) are, respectively,

$$\frac{\partial}{\partial t}\int_{[V_{\text{in}}]}[f_{\text{in}}]\,dV = \int_{[V_{\text{in}}]}\left[\frac{\partial f_{\text{in}}}{\partial t}\right]dV + \oint_{[D]} d\mathbf{s}\cdot\left[\frac{\mathbf{u}f_{\text{in}}}{W}\right], \tag{10}$$

$$\frac{\partial}{\partial t}\int_{[V_{\text{ex}}]}[f_{\text{ex}}]\,dV = \int_{[V_{\text{ex}}]}\left[\frac{\partial f_{\text{ex}}}{\partial t}\right]dV - \oint_{[D]} d\mathbf{s}\cdot\left[\frac{\mathbf{u}f_{\text{ex}}}{W}\right], \tag{11}$$

$$\frac{\partial}{\partial t}\int_{V}[f]dV = \int_{V}\left[\frac{\partial f}{\partial t}\right]dV + \oint d\mathbf{s}\cdot\left[\frac{\mathbf{u}f|}{W}\right], \tag{12}$$

where square brackets in the integrands indicate retardation to times $t - R$ as in (7); square brackets on V_{ex} and D indicate intersection with $\text{LC}(t)$ as defined in (9);

$$W \equiv 1 - \mathbf{u}\cdot\hat{R}, \quad \hat{R} \equiv \mathbf{R}/R; \tag{13}$$

and $f|$ in (12) is defined in (6). Note that partial-t derivatives appear in (10)–(12) because retarded integrals are functions of the field-point coordinates r.

To prove (10), we again start from first principles. The left side of (10) is, from (7),

$$\frac{\partial\phi}{\partial t} = \lim_{\delta t\to 0}\left\{\frac{\phi(\mathbf{r},t+\delta t) - \phi(\mathbf{r},t)}{\delta t}\right\}. \tag{14}$$

As with (3), the right side of (14) [with (7)] has contributions from two spatial regions. In this case, the regions, illustrated in Fig. 1, are[6] the following. (i) The region $[V_{\text{overlap}}]$ is defined by

$$[V_{\text{overlap}}] \equiv [V_{\text{in}}]_{\text{LC}(t+\delta t)} \cap [V_{\text{in}}]_{\text{LC}(t)}.$$

The contribution to (14) from this region in the limit is clearly

$$\int_{[V_{\text{in}}]_{\text{LC}(t)}}\left[\frac{\partial f_{\text{in}}}{\partial t}\right]dV,$$

which is the first integral on the right side of (10).

(ii) The region $[V^*]$ is comprised of points lying inside $[V_{\text{in}}]_{\text{LC}(t+\delta t)}$ and $[V_{\text{in}}]_{\text{LC}(t)}$ but outside $[V_{\text{overlap}}]$. The contribution to (14) from $[V^*]$ in the limit as $\delta t\to 0$ is

$$\oint_{[D]} d\mathbf{s}\cdot\left[\frac{\mathbf{u}f_{\text{in}}}{W}\right],$$

where W, defined in (13), arises because the integral is evaluated on the past light cone from the field point at r.

Addition of the contributions from (i) and (ii) reduces (14) with (7) to the form of (10).

Equation (11) is proven in a similar manner, and addition of (10) and (11) yields (12).

III. GAUSS'S THEOREM

(A) Gauss's theorem for the volume integral $\int_V(\nabla\cdot\mathbf{F})dV$, defined on the slice $t = \text{const}$, is clearly

$$\int_V(\nabla\cdot\mathbf{F})dV = \oint_S d\mathbf{s}\cdot\mathbf{F} + \int_D d\mathbf{s}\cdot\mathbf{F}|, \tag{15}$$

where $\mathbf{F}|$ is the vector discontinuity corresponding to (6):

$$\mathbf{F}| \equiv \mathbf{F}_{\text{in}} - \mathbf{F}_{\text{ex}} \tag{16}$$

on $D(t)$.

(B) A simple calculation yields Gauss's theorem for the retarded integral $\int_V[\nabla\cdot\mathbf{F}]dV$:

$$\int_V[\nabla\cdot\mathbf{F}]dV = \oint_S d\mathbf{s}\cdot[\mathbf{F}] + \oint_D d\mathbf{s}\cdot[\mathbf{F}|] - \int_V\left[\frac{\partial}{\partial t}(\hat{R}\cdot\mathbf{F})\right]dV, \tag{17}$$

where the vertical slash indicates discontinuities as in (16).

IV. TIME DERIVATIVES OF SURFACE INTEGRALS

This section considers time derivaties of D-surface integrals of vector discontinuities of F, defined in (16).

(A) A straightforward series of calculations using (15) and (5) yields

$$\frac{d}{dt}\oint_D d\mathbf{s}\cdot\mathbf{F}| = \oint_D d\mathbf{s}\cdot\frac{\partial\mathbf{F}}{\partial t}\Big| + \oint_D d\mathbf{s}\cdot\mathbf{u}\{(\nabla\cdot\mathbf{F})|\}, \tag{18}$$

where the vertical slash again defines discontinuities as in (16) and (6).

(B) The retarded analog of (18) is found from (17) with (12):

$$\frac{\partial}{\partial t}\oint_{[D]} d\mathbf{s}\cdot[\mathbf{F}|] = \oint_{[D]} d\mathbf{s}\cdot\left[\frac{\partial\mathbf{F}}{\partial t}\Big|\right] + \oint_{[D]} d\mathbf{s}\cdot\left[\frac{\mathbf{u}(\nabla\cdot\mathbf{F})|}{W}\right] + \oint_{[D]} d\mathbf{s}\cdot\left[\frac{\mathbf{u}(\partial/\partial t)(\hat{R}\cdot\mathbf{F})|}{W}\right]. \tag{19}$$

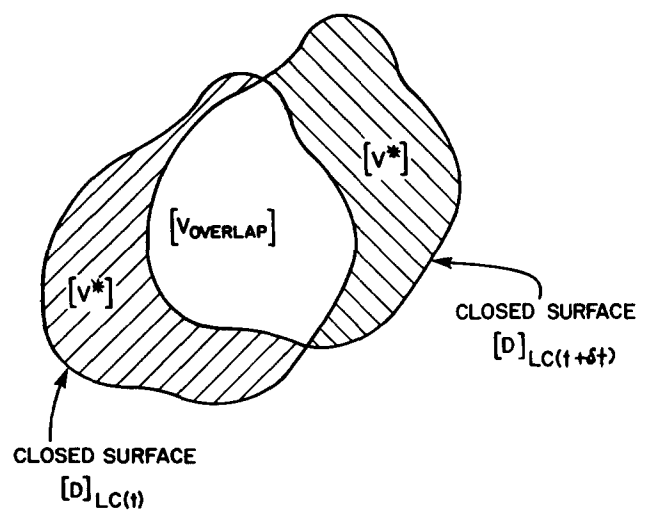


FIG. 1. Region of integration for the retarded integrals.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Council of Canada, Grant No. A5340 and the University of Victoria Faculty Research Grant.

[1]F. I. Cooperstock and P. H. Lim, "The new formula for gravitational radiation energy loss and the axially symmetric two-body problem," Invited Paper presented at the Canadian Conference on General Relativity and Relativistic Astrophysics, Dalhousie University, Halifax, Nova Scotia, Canada, April 25–27, 1985, Proceedings to appear in Can. J. Phys.

[2]F. I. Cooperstock and P. H. Lim, Astrophys. J. (to be published); F. I. Cooperstock and P. H. Lim, Phys. Rev. Lett. **55**, 265 (1985).

[3]See F. I. Cooperstock, Phys. Rev. D **10**, 3171 (1974); Phys. Rev. D **25**, 3126 (1982); F. I. Cooperstock and D. W. Hobill, Phys. Rev. D **20**, 2995 (1979); Gen. Relativ. Gravit. **14**, 361 (1982); and F. I. Cooperstock and P. H. Lim, Phys. Rev. D **15**, 2105 (1977) for background on this problem.

[4]In what follows, we shall choose units in which $c = 1$.

[5]Alternatively, $(\partial f/\partial t)$ could be defined on D as the (infinite) t derivative of a step discontinuity. A simple δ-distribution analysis shows that in such a case, the contribution from D is precisely the surface integral in (5).

[6]Note that these regions are the retarded analogs of (i) and (ii) following (3) above.

Transmission through a system of potential barriers. II. Necessary condition for complete transparency. A maximum transmission problem

Örjan Dammert
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 5 August 1981; accepted for publication 9 September 1985)

A smooth one-dimensional system of N potential barriers of arbitrary shapes (unequal or equal) is considered. A general necessary condition for complete transparency is obtained that can be understood as a constraint on the reflection coefficients pertaining to the single barriers of the system. A maximum transmission problem of a general kind is solved and the solution is used to give a physical interpretation of the necessary condition. The sub- and superbarrier cases are treated in a unified way. The exact final formulas can readily be converted into accurate approximate ones by insertion of available phase-integral expressions (of an arbitrary order) for certain characteristic quantities appearing in the formulas.

I. INTRODUCTION

We study the one-dimensional Schrödinger equation

$$\frac{d^2\psi}{dz^2} + Q^2(z)\psi = 0, \tag{1}$$

where

$$Q^2(z) = (2m/\hbar^2)[E - V(z)], \tag{2}$$

z being a complex variable (the real values of which will be denoted by x) and E being the energy of a particle with mass m moving in the potential field $V(x)$ that, having N humps of various shapes, forms a general system of N potential barriers. We assume that $Q^2(z)$ is an analytical function of z in the complex plane.

In a previous paper[1] (to be referred to as I), the transmission coefficient pertaining to the above-mentioned system of $N(\geqslant 1)$ potential barriers of general shapes (unequal or equal) is expressed exactly, by means of the phase-integral method of N. Fröman and P. O. Fröman, in terms of quantities characterizing the separate barriers and wells of the system. The present article is based on the results in I, and the reader is referred to that paper for the notation and general background and also for illustrative figures and for further references to relevant papers. Some key facts about the phase-integral method are given in Appendix A of I.

In two forthcoming papers,[2,3] also based on the results in I, we shall treat in the first one transmission through a system of N identical potential barriers of a general shape and, in the second paper, bound states of a general potential well containing N humps of arbitrary shapes.

The problem to be treated in the present paper can be regarded as a generalization of a corresponding problem for a system of rectangular barriers. The latter problem, being easier to formulate and visualize, is therefore well suited as an introduction to our real work. So let us start by considering the rectangular system of N barriers shown in Fig. 1, which is defined by

$$V(x) = \begin{cases} V_i, & \text{for } x_0 + s_i < x < x_0 + s_i + a_i, \\ 0, & \text{elsewhere on the } x \text{ axis}, \end{cases} \tag{3}$$

where

$$s_i = \sum_{k=1}^{i-1} (a_k + b_k), \quad s_1 = 0, \quad i = 1,2,...,N.$$

Let the energy E of the incident particle take some fixed value in the discussion below. The reflection coefficients $R_i(E)$ for the separate barriers ($i = 1,2,...,N$) as well as the transmission coefficient $T(E)$ for the whole barrier system then naturally become fixed. We ask the following question. How can we alter the form of the barrier system without changing the reflection coefficients $R_i(E)$ of the separate barriers? As we realize, this can be done by altering, quite arbitrarily and independently, the distances b_i ($i = 1,2,...,N$) between the separate barriers. In addition, we can continuously change the form of each separate barrier $B(i)$, within certain limits, by letting a change of the height V_i be accom-

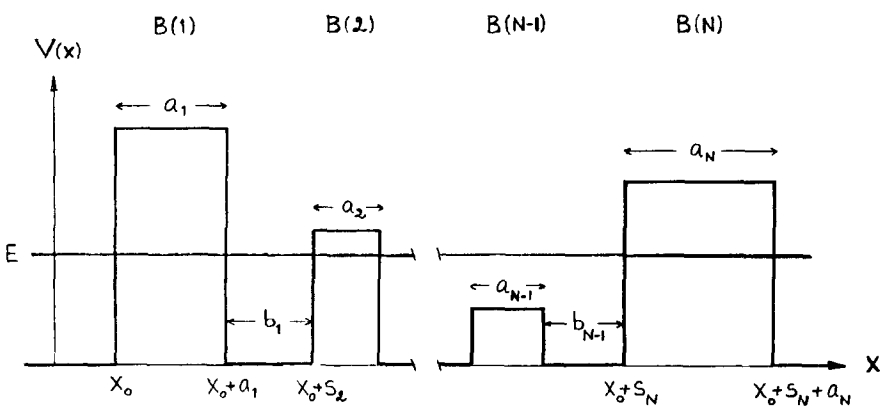


FIG. 1. A system of N rectangular barriers. The width of the barrier $B(i)$ is a_i, its height is V_i, and the distance to the next barrier to the right is b_i; E is the energy of the incident particle; and the quantity s_i is defined by $s_i = \Sigma_{k=1}^{i-1}(a_k + b_k)$ and $s_1 = 0$.

 0022-2488/86/020461-10$02.50 © 1986 American Institute of Physics

panied by the proper change of the width a_i so as to keep $R_i(E)$ constant during the modification. We may regard all the different barrier systems, which can be obtained by altering the parameters b_i, a_i, and V_i ($i = 1,2,...$) in the way described above, as members of a family of barrier systems. The transmission coefficient $T(E)$ is certainly not the same for every member of this family, however, but depends strongly on the parameters b_i, a_i, and V_i. We now ask two questions. What is the maximum value of $T(E)$ that can be attained through alterations in all possible ways of the form of the given rectangular barrier system, limited only by the restriction that the reflection coefficients $R_i(E)$ of the separate barriers shall be unchanged by the modifications and that the rectangular property shall be preserved. The reader is reminded that E is held fixed. What condition must be fulfilled by the original rectangular barrier system [fulfilled by the reflection coefficients $R_i(E)$ of the separate barriers] in order that the maximum value of $T(E)$ shall be equal to unity? An answer to the latter question would give a necessary condition for complete transparency.

We may pose similar questions with respect to a smooth barrier system. The following extreme case will serve as an introductory example. We consider the potential

$$V(x) = \sum_{i=1}^{N} C_i \exp\left[\frac{-(x-a_i)^2}{b_i^2}\right], \tag{4}$$

where a_i, b_i, and C_i are real constants. The distances $|a_i - a_j|$, $i \neq j$, are assumed to be very large. The curve of $V(x)$ then shows N barrier humps at very large distances from each other. We realize that by changing moderately the distances $|a_i - a_j|$, $i \neq j$, and thereby altering the form of the barrier system, we can make the transmission coefficient $T(E)$ for the barrier system change considerably, while at the same time the reflection coefficients pertaining to the component barriers remain essentially constant. With regard to these barrier systems, obtained by varying the parameters a_i ($i = 1,...,N$), we could ask the above-mentioned questions about maximum transmission and possible transparency.

However, our aim is to consider general smooth barrier systems containing barrier humps at arbitrary (not necessarily large) distances from each other. For systems of that kind we want to formulate a natural generalization of the earlier presented problem for rectangular barrier systems.

For this purpose we shall first consider a potential $V(x)$ having just one single hump. Let us take

$$V(x;C,a) = C \exp[-x^2/a^2], \tag{5}$$

where C and a are real constants. The potential curve is pictured in Fig. 2. Again we let the energy E of the incident particle be fixed in the discussion below. The points t_1 and t_2 are the classical turning points where $V(x) = E$, and x_1 and x_2 are points to the left and to the right, respectively, of the turning points. Let us apply the results in I to the potential (5), making some comments without going into detail.

From (44a) and (28) in I, putting $N = 1$, we find for the reflection coefficient $R(E)$ the following exact expression:

$$R(E) = (B_1/A_1)^2. \tag{6}$$

In (6), A_1 and B_1 stand for the limiting values, as $x_1 \to -\infty$ and $x_2 \to +\infty$, of the quantities A_1 and B, which are generally defined by (14a) and (14b) in I as follows:

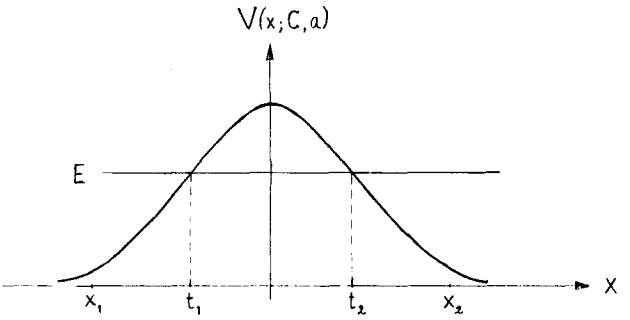


FIG. 2. Subbarrier transmission through the potential barrier $V(x;C,a) = C\exp[-x^2/a^2]$. Here E is the energy of the incident particle. The points t_1 and t_2 are the classical turning points where $V(x) = E$, and x_1 and x_2 are points to the left and to the right, respectively, of the turning points. For further details about the phase-integral method the reader is referred to Figs. 1(a), 1(b), 2(a)–2(d), and 3 in Ref. 1.

$$A_1 = |F_{12}(x_1,x_2)|, \tag{7a}$$

$$B_1 = |F_{22}(x_1,x_2)|. \tag{7b}$$

The elements of the matrix $\mathbf{F}(x_1,x_2)$ are given by the convergent series (3.22a)–(3.22d) in Ref. 4. Formulas (B1) and (B2) in I give useful phase-integral expressions for A_1 and B_1, consisting of an approximate plus a correction part. We obtain from (6) above and (B5) in I the exact formula

$$\frac{1}{R(E)} = \left(\frac{A_1}{B_1}\right)^2 = [1 + \exp(-2K_1)] \times (1 + \exp[-K_1 - \tfrac{1}{2}(|K_1| + K_1)]O(\mu_1)). \tag{8}$$

The quantities μ_1 and $\mu_1 \exp[\frac{1}{2}(|K_1| - K_1)]$ are assumed to be small compared to unity. The symbol $O(\mu_1)$ denotes a quantity at most of the order of magnitude μ_1. From (8) we conclude that $(B_1/A_1)^2$ is practically insensitive to changes in the positions of x_1 and x_2, if these points are situated far enough from the turning points t_1 and t_2. In fact, formula (8) with neglect of the correction term gives the same approximate value of $R(E)$ in the case when x_1 and x_2 are situated at finite distances from the barrier top as in the limiting case when $x_1 \to -\infty$ and $x_2 \to +\infty$. The difference between the two cases is only seen in the different values of the pertaining correction terms.

In the subbarrier case illustrated in Fig. 2, the first-order phase-integral expression for K_1 is given by

$$K_1 = \frac{(2m)^{1/2}}{\hbar} \int_{t_1}^{t_2} (V(x) - E)^{1/2}\, dx. \tag{9}$$

From (8) and (9) we clearly see how $R(E)$ is affected by a variation of the form of $V(x)$. If the value of C in (5) is increased, the potential curve in Fig. 2 is raised, which means that $K_1(E)$ in (9) increases. But by decreasing appropriately the parameter a in (5) at the same time, we can arrange that $K_1(E)$ becomes unaltered by the variation of $V(x)$. However, the changes of C and a also affect the correction term in the exact expression (8) for the reflection coefficient $R(E)$. This correction term is much smaller than the term

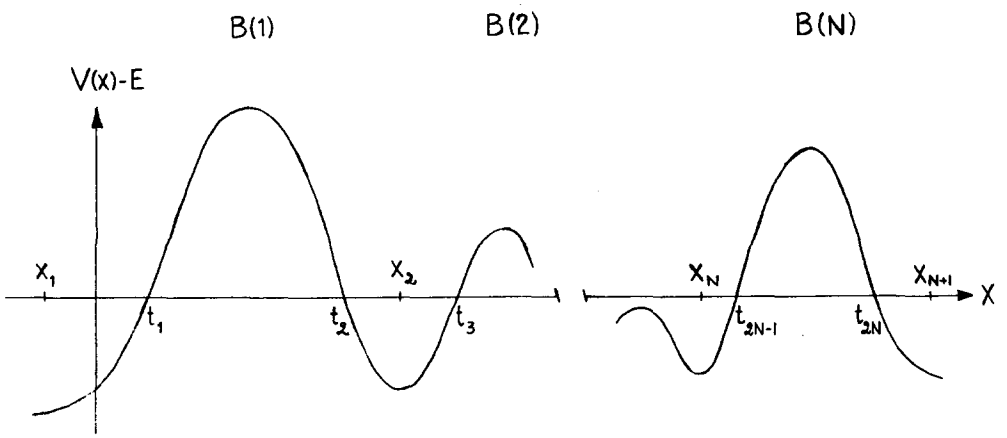


FIG. 3. A general smooth system of N potential barriers. For further details concerning the phase-integral method the reader is referred to the figures in Ref. 1.

$[1+\exp(-2K_1)]$. One realizes that it is possible, by means of a final small adjustment of the parameter a, to change the value of the dominant term $[1+\exp(-2K_1)]$ precisely enough to compensate for the variation of the correction term, thus keeping $R(E)$ unaltered. Making these connected changes of C and a in the potential (5), we obtain a potential $V(x;\acute{C},\acute{a})$ that has the same reflection coefficient $R(E)$ as the original potential (5), for the fixed value E of the energy of the incident particle. The considerations above show the possibility of changing the form of the given potential $V(x)$ in (5) in such a way that the value of $R(E)$, for the fixed E, is preserved during the alteration. Also, in the case when x_1 and x_2 take finite values, similar arguments can be used to show the possibility of altering the form of the potential $V(x)$ in such a way that the quantity $(B_1/A_1)^2$, where x_1 and x_2 take finite values, is conserved during the alterations. The use of higher-order phase-integral expressions, instead of the first-order ones, introduces certain modifications of the quantities K_1 and μ_1 in (8), but the general reasoning above remains valid.

The reflection coefficient $R(E)$ [see (6)] is a measure of that obstacle to an incident particle of energy E that is represented by the barrier. In this sense $R(E)$ is a measure of the size of the barrier relative to an incident particle of energy E. Bearing this in mind, we shall call the quantity

$$(B_n/A_n)^2 \tag{10}$$

the size of the barrier $B(n)$, relative to an incident particle of energy E, also when the barrier $B(n)$ is part of a multibarrier system and its end points x_n and x_{n+1} assume finite values (cf. Fig. 3).

We now proceed to consider arbitrary smooth systems of N potential barriers of the kind shown in Fig. 3. The barrier $B(n)$ is by definition that part of the potential $V(x)$ that is situated between x_n and x_{n+1}. Previous considerations indicate that the sizes of the separate barriers, i.e., $(B_n/A_n)^2$ for $n=1,2,...,N$, are those quantities pertaining to a smooth barrier system that naturally correspond to the reflectivities of the separate barriers belonging to a rectangular barrier system.

Here we are interested in altering the form of the potential $V(x)$, i.e., altering the "distances" between the single barrier humps and at the same time varying the "heights" and the "widths" of the humps, in all possible ways limited only by the restriction that the sizes $(B_n/A_n)^2$ of the separate barriers, relative to some fixed energy E of the incident particle, shall be unchanged by the variations. We shall refrain from trying to give a formal mathematical proof of the existence of such variations of the potential $V(x)$, but shall simply assume the existence and refer to the previous discussion in support of the plausibility of this assumption.

Exploiting the final formulas (43a) and (43b) in I for the transmission coefficient, we shall in the present paper derive an exact expression for that maximum value of the transmission coefficient $T(E)$, pertaining to a given system of N potential barriers of arbitrary shapes (unequal or equal), that can be obtained by varying, at some fixed energy E of the incident particle, the shape of the potential $V(x)$ in all possible ways consistent with the requirement that the sizes of the separate barriers (relative to that particular E) remain unchanged. The multibarrier potential $V(x)$ involved in this maximum problem is thus specified only to the extent that the sizes of the N single barriers (relative to the particular E) have given constant values, but is otherwise freely deformable.

In the treatment we shall repeatedly use the inequality (B7) in Appendix B of I, i.e.,

$$A_n > B_n > 0, \quad n=1,2,...,N, \tag{11}$$

which follows from the assumption that the absolute value of the error term in (B2) of I is less than unity. This is an assumption concerning the smallness of the quantity μ_n, defined by (B6) in I.

From the formula for maximum transmission, we shall then immediately obtain a necessary condition for total transmission through a general system of N potential barriers, $G_N(E)<0$, which is a constraint on the sizes of the single barriers. As further explained in Appendix A, this condition can be understood as requiring the reflection coefficient for the biggest barrier in the system to be smaller than or equal to a certain maximum reflection coefficient pertaining to the rest of the barriers in the system.

In Appendix B we shall give relations between characteristic quantities associated with two single barriers that are mirror images of each other.

The final exact formulas can be converted readily into accurate approximate ones, suitable for evaluation, by insertion of available phase-integral expressions (of an arbitrary order) for certain characteristic quantities appearing in the formulas. These phase-integral expressions together with rigorous error estimates are listed in Appendix B of I.

We shall proceed as follows. We consider all those N barrier systems that, for a particular energy E of the incident particle, are characterized by the same given set of values of $(B_n(E)/A_n(E))^2$ for $n = 1,2,...,N$, but that differ from each other as regards the values of $\beta_n(E)$ for $n = 1,2,...,N-1$. The quantities A_n, B_n, and β_n are defined by Eqs. (18a), (18b), (29a), and (29b) in I. For those systems and this energy E, we shall determine the maximum possible value of the transmission coefficient. This will be achieved technically by calculating the minima of the quantities $|P_{11}(x_1,x_{n+1})|$ and $|P_{12}(x_1,x_{n+1})|$, which appear in formulas (43a) and (43b) in I for the transmission coefficient, considering A_n and B_n $(n = 1,2,...,N)$ to be constants and β_n $(n = 1,2,...,N-1)$ to be independent variables. As a preliminary, we shall in Sec. II introduce some useful quantities and in Sec. III prove a few assertions.

II. DEFINITIONS

For a general system of $N(\geqslant 1)$ potential barriers, we shall below define certain quantities, which will enter into the final formulas (57)–(60) for extreme transmission or reflection, and which will also be used in Secs. III and IV in the derivation of these formulas. Although seemingly uncalled for at the moment, these quantities can hardly be avoided.

We define, for $i = 1,2,...,n$ and $n \leqslant N$,

$$\Gamma_n(i) = \frac{1}{2}\left(\frac{A_i + B_i}{A_i - B_i}\prod_{k=1}^{n}(A_k - B_k) - \frac{A_i - B_i}{A_i + B_i}\prod_{k=1}^{n}(A_k + B_k)\right), \tag{12}$$

where A_k and B_k are defined by (18a) and (18b) in I.

Letting B_m/A_m be a quantity in the sequence B_1/A_1, $B_2/A_2,...,B_n/A_n$ satisfying the conditions

$$B_m/A_m \geqslant B_k/A_k, \quad \text{for } k = 1,2,...,n, \tag{13}$$

we further define

$$G_n = \Gamma_n(m) = \frac{1}{2}\left(\frac{A_m + B_m}{A_m - B_m}\prod_{k=1}^{n}(A_k - B_k) - \frac{A_m - B_m}{A_m + B_m}\prod_{k=1}^{n}(A_k + B_k)\right), \tag{14}$$

$$S_n = \frac{1}{2}\left(\frac{A_m + B_m}{A_m - B_m}\prod_{k=1}^{n}(A_k - B_k) + \frac{A_m - B_m}{A_m + B_m}\prod_{k=1}^{n}(A_k + B_k)\right). \tag{15}$$

We also define

$$P_n^2 = (P_n)^2 = \prod_{k=1}^{n}(A_k^2 - B_k^2). \tag{16}$$

From (16) above and (20) in I, it follows that

$$P_n^2 = \exp\left(-\sum_{\nu=1}^{n} 2K_\nu\right). \tag{17}$$

From (14)–(16) it is easily seen that

$$S_n^2 - G_n^2 = (S_n + G_n)(S_n - G_n) = P_n^2,$$

i.e.,

$$S_n = \sqrt{G_n^2 + P_n^2}. \tag{18}$$

Defining furthermore

$$H_n = \frac{1}{2}\left(\prod_{k=1}^{n}(A_k + B_k) - \prod_{k=1}^{n}(A_k - B_k)\right), \tag{19}$$

$$M_n = \frac{1}{2}\left(\prod_{k=1}^{n}(A_k + B_k) + \prod_{k=1}^{n}(A_k - B_k)\right), \tag{20}$$

we obtain

$$M_n^2 - H_n^2 = (M_n + H_n)(M_n - H_n) = P_n^2. \tag{21}$$

Hence,

$$M_n = \sqrt{H_n^2 + P_n^2}. \tag{22}$$

It should here be remarked, since it is not evident from the expressions above, that all the quantities A_k, B_k, and K_ν, and consequently also $\Gamma_n(i)$, G_n, S_n, P_n, H_n, and M_n, are functions of the energy E of the incident particle.

III. THEOREMS

We shall now prove a few assertions, some of which will be used in the demonstration of the statements (33a)–(33d) in Sec. IV, and some in a forthcoming paper.[3] The principal results in the present section are derived for the case $n \geqslant 3$. The special case $n = 2$ is treated separately while the case $n = 1$ is omitted. For the physical meaning of the quantities $(B_i/A_i)^2$ we refer to the Introduction.

For $n \geqslant 3$ we assert that

$$\Gamma_n(i) \geqslant 0 \Rightarrow \frac{B_i}{A_i} > \frac{B_k}{A_k}, \quad \text{for} \quad k = 1,2,...,n \quad \text{but} \quad k \neq i. \tag{23}$$

Proof: Assuming $\Gamma_n(i) \geqslant 0$, we obtain from (12)

$$\frac{A_i - B_i}{A_i + B_i} \leqslant \prod_{\substack{k=1 \\ k\neq i}}^{n}\frac{A_k - B_k}{A_k + B_k}. \tag{24}$$

Using (11), recalling that $n \geqslant 3$, we get from (24)

$$\frac{A_i - B_i}{A_i + B_i} < \frac{A_k - B_k}{A_k + B_k}, \quad \text{for any} \quad k \neq i.$$

Hence,

$$\frac{B_i}{A_i} > \frac{B_k}{A_k}, \quad k \neq i,$$

which concludes the proof.

For $n = 2$ we have instead

$$\Gamma_2(i) > 0 \Rightarrow \frac{B_i}{A_i} > \frac{B_k}{A_k}, \quad k = 1,2 \quad \text{but} \quad k \neq i, \tag{25a}$$

and

$$\Gamma_2(i) = 0 \Rightarrow \frac{B_1}{A_1} = \frac{B_2}{A_2}. \tag{25b}$$

From the theorems (23) and (25a) and the definition of G_n in (13) and (14), one immediately obtains the following corollaries.

For $n \geqslant 3$, it is true that

$$\Gamma_n(i) \geqslant 0 \Rightarrow \begin{cases} \Gamma_n(i) = G_n, \\ \Gamma_n(k) \neq G_n, \quad \text{if} \quad k \neq i, \end{cases} \tag{26a}$$

and, for $n = 2$,

$$\Gamma_2(i) > 0 \Rightarrow \begin{cases} \Gamma_2(i) = G_2, \\ \Gamma_2(k) \neq G_2, \quad \text{if} \quad k \neq i \end{cases} \tag{26b}$$

We shall also prove the following assertions.

For $n \geqslant 3$, it is true that

$$G_n < 0 \Rightarrow \Gamma_n(k) < 0, \quad \text{for} \quad k = 1,2,...,n, \tag{27}$$

and

$$G_n \geqslant 0 \Rightarrow \begin{cases} G_n = \Gamma_n(m) \geqslant 0, & \text{where } m \text{ satisfies the} \\ & \text{relations (13);} \\ \Gamma_n(k) < 0, & \text{for } k = 1,2,...,n \text{ but } k \neq m. \end{cases} \tag{28}$$

Proof: Negating the right-hand member of (27), we obtain by using (26a) the relation $G_n \geqslant 0$, which is the very negation of the left-hand member of (27). Assertion (27) is thus proved by contradiction.

Turning to the assertion (28), we realize from (13) and (14) that G_n is identical with at least one of the quantities $\Gamma_n(k)$, $k = 1,2,...,n$. It is therefore true that $G_n = \Gamma_n(m)$, where m is a particular one of the integers $1,2,...,n$. As a consequence, $\Gamma_n(k)$, where $k \neq m$, must be < 0, since the contrary assumption, i.e., $\Gamma_n(k) \geqslant 0$, according to (26a) would imply that $\Gamma_n(k) = G_n$ and $\Gamma_n(m) \neq G_n$, which contradicts the statement $G_n = \Gamma_n(m)$ above. Assertion (28) is thereby proved.

If the potential $V(x)$ of a barrier system satisfies the equation

$$V(x) = V(-x) \tag{29}$$

for all real values of x, we shall say that the barrier system is symmetrical. Let us consider a symmetrical system of N barriers. We conveniently choose the points $x_1, x_2, ..., x_{N+1}$ (separating the single barriers) on the real axis such that they, in pairs, become symmetrically situated with respect to the origin, i.e., such that $x_1 = -x_{N+1}$, $x_2 = -x_N$, $x_3 = -x_{N-1}$, etc. The relations (B10a) and (B10b) in Appendix B are then valid for each pair of barriers, $B(i)$ and $B(k)$, that are symmetrically situated with respect to the origin.

For a symmetrical system of N barriers, where the points $x_1, x_2, ..., x_{N+1}$ are chosen as described above, the following assertions are true.

(i) If N is even $\geqslant 4$, then

$$G_N < 0. \tag{30}$$

(ii) If N is odd $\geqslant 3$, then (putting $N = 2r + 1$) we have

$$G_N \geqslant 0 \Leftrightarrow \Gamma_N(r+1) \geqslant 0, \tag{31a}$$

$$G_N \geqslant 0 \Rightarrow G_N = \Gamma_N(r+1). \tag{31b}$$

Proof: (i) We assume that $N = 2r \geqslant 4$. Since N is even, the system consists of pairs of barriers, each pair being symmetrically situated with respect to the origin. To each pair the relations (B10a) and (B10b) apply, from which we realize that the values A_k and B_k characterize both barriers of that pair, to which the barrier $B(k)$ belongs. This is true also of the values A_m and B_m that appear in (14). Hence, from (14) together with (B10a), (B10b), and (11), it follows that

$$\begin{aligned} G_N &= \frac{1}{2}\Bigg((A_m + B_m) \prod_{\substack{k=1 \\ k \neq m}}^{N} (A_k - B_k) \\ &\quad - (A_m - B_m) \prod_{\substack{k=1 \\ k \neq m}}^{N} (A_k + B_k) \Bigg) \\ &= \frac{1}{2}(A_m^2 - B_m^2) \Bigg(\prod_{\substack{k=1 \\ k \neq m}}^{r} (A_k - B_k)^2 \\ &\quad - \prod_{\substack{k=1 \\ k \neq m}}^{r} (A_k + B_k)^2 \Bigg) < 0. \end{aligned}$$

(ii) We assume that $N = 2r + 1 \geqslant 3$. The barrier system then contains a central barrier, $B(r+1)$, around which the rest of the barriers, in pairs, are symmetrically located. Equations (B10a) and (B10b) apply to each such pair. We shall first prove that $\Gamma_N(i) < 0$, if $i \neq r + 1$. From (12), using (B10a), (B10b), and (11), we obtain in this case

$$\begin{aligned} \Gamma_N(i) &= \frac{1}{2}\Bigg((A_i + B_i) \prod_{\substack{k=1 \\ k \neq i}}^{N} (A_k - B_k) \\ &\quad - (A_i - B_i) \prod_{\substack{k=1 \\ k \neq i}}^{N} (A_k + B_k) \Bigg) \\ &= \frac{1}{2}(A_i^2 - B_i^2) \Bigg((A_{r+1} - B_{r+1}) \prod_{\substack{k=1 \\ k \neq i}}^{r} (A_k - B_k)^2 \\ &\quad - (A_{r+1} + B_{r+1}) \prod_{\substack{k=1 \\ k \neq i}}^{r} (A_k + B_k)^2 \Bigg) < 0. \end{aligned}$$

Thus, if $N = 2r + 1 \geqslant 3$, we have

$$\Gamma_N(i) < 0, \quad \text{for } i = 1,2,...,N \quad \text{but} \quad i \neq r+1. \tag{32}$$

From (28) and (32) we conclude that

$$G_N \geqslant 0 \Rightarrow G_N = \Gamma_N(r+1) \geqslant 0,$$

which proves assertion (31b) and half of assertion (31a). From (26a) we see that

$$\Gamma_N(r+1) \geqslant 0 \Rightarrow G_N \geqslant 0,$$

which completes the proof of assertion (31a).

Summing up the content of (30), (31a), and (31b) concerning a symmetrical system of N ($\geqslant 3$) barriers, we see that $G_N < 0$ in all cases except the particular one when N is odd and $\Gamma_N(r+1) \geqslant 0$, in which case $G_N \geqslant 0$.

IV. MAXIMA AND MINIMA OF $|P_{11}(x_1, x_{n+1})|$ AND $|P_{12}(x_1, x_{n+1})|$

In this section we shall prove the following assertion. On the assumption that A_k and B_k have given constant values while the quantities β_k may vary independently of each other ($k = 1,2,...,n$), the maxima and minima of $|P_{11}(x_1, x_{n+1})|$ and $|P_{12}(x_1, x_{n+1})|$ are given by the following scheme, valid for $n \geqslant 1$:

$$\left.\begin{matrix} |P_{12}(x_1, x_{n+1})|_{\max}, & \text{for } n \text{ odd} \\ |P_{11}(x_1, x_{n+1})|_{\max}, & \text{for } n \text{ even} \end{matrix}\right\} = M_n, \tag{33a}$$

$$\left.\begin{matrix}|P_{11}(x_1,x_{n+1})|_{\max}, & \text{for } n \text{ odd} \\ |P_{12}(x_1,x_{n+1})|_{\max}, & \text{for } n \text{ even}\end{matrix}\right\} = H_n, \tag{33b}$$

$$\left.\begin{matrix}|P_{12}(x_1,x_{n+1})|_{\min}, & \text{for } n \text{ odd} \\ |P_{11}(x_1,x_{n+1})|_{\min}, & \text{for } n \text{ even}\end{matrix}\right\} = \begin{cases} P_n, & \text{if } G_n \leqslant 0, \\ S_n, & \text{if } G_n > 0, \end{cases} \tag{33c}$$

$$\left.\begin{matrix}|P_{11}(x_1,x_{n+1})|_{\min}, & \text{for } n \text{ odd} \\ |P_{12}(x_1,x_{n+1})|_{\min}, & \text{for } n \text{ even}\end{matrix}\right\} = \begin{cases} 0 & \text{if } G_n \leqslant 0, \\ G_n, & \text{if } G_n > 0. \end{cases} \tag{33d}$$

Proof: The proof of (33a)–(33d) will be carried out by complete induction.

Part 1: Recalling the definitions (14), (15), (19), and (20) and using that $G_1 = B_1$ and that $B_1 > 0$, according to (11), we find that Eqs. (33a)–(33d) assert, for $n = 1$,

$$|P_{12}(x_1,x_2)|_{\max} = M_1 = A_1,$$
$$|P_{11}(x_1,x_2)|_{\max} = H_1 = B_1,$$
$$|P_{12}(x_1,x_2)|_{\min} = S_1 = A_1,$$
$$|P_{11}(x_1,x_2)|_{\min} = G_1 = B_1.$$

According to Eqs. (28) and (29a) in I, the matrix $\mathbf{P}(x_1,x_2)$ is given by

$$\mathbf{P}(x_1,x_2) = \begin{pmatrix} B_1 & A_1 \\ A_1 & B_1 \end{pmatrix}.$$

Assertions (33a)–(33d) are thus seen to be true for $n = 1$.

Part 2: Let us next show that the assertions (33a)–(33d) must be true for n, if they are true for $n-1$. We shall first consider the case when n is odd. Thus, we assume that (33a)–(33d) are true for $n-1$, where n ($\geqslant 3$) is an odd integer. From (36) in I, and (17) in the present paper, we then obtain

$$|P_{11}(x_1,x_n)|^2 - |P_{12}(x_1,x_n)|^2 = P_{n-1}^2. \tag{34}$$

Hence,

$$|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)| = \frac{P_{n-1}^2}{|P_{11}(x_1,x_n)| + |P_{12}(x_1,x_n)|}. \tag{35}$$

From (35), observing with the aid of (34) that $|P_{11}(x_1,x_n)|$ and $|P_{12}(x_1,x_n)|$ simultaneously assume their maximum values (alternatively, minimum values), we conclude that

$$(|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)|)_{\min} = |P_{11}(x_1,x_n)|_{\max} - |P_{12}(x_1,x_n)|_{\max} \tag{36a}$$

$$(|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)|)_{\max} = |P_{11}(x_1,x_n)|_{\min} - |P_{12}(x_1,x_n)|_{\min}. \tag{36b}$$

From Eqs. (33) and (28) in I, we obtain

$$\mathbf{P}(x_1,x_{n+1}) = \mathbf{P}(x_1,x_n)\mathbf{P}(x_n,x_{n+1}) = \begin{pmatrix} P_{11}(x_1,x_n) & P_{12}(x_1,x_n) \\ P_{21}(x_1,x_n) & P_{22}(x_1,x_n) \end{pmatrix} \times \begin{pmatrix} B_n & A_n \\ A_n \exp\{2i\beta_{n-1}\} & B_n \exp\{2i\beta_{n-1}\} \end{pmatrix}.$$

Hence,

$$P_{11}(x_1,x_{n+1}) = B_n P_{11}(x_1,x_n) + \exp\{2i\beta_{n-1}\} A_n P_{12}(x_1,x_n), \tag{37a}$$

$$P_{12}(x_1,x_{n+1}) = A_n P_{11}(x_1,x_n) + \exp\{2i\beta_{n-1}\} B_n P_{12}(x_1,x_n). \tag{37b}$$

Proof of (33d): Let us determine $|P_{11}(x_1,x_{n+1})|_{\min}$. One realizes from (37a) that

$$|P_{11}(x_1,x_{n+1})|_{\min} > 0 \tag{38}$$

if and only if either

$$(B_n|P_{11}(x_1,x_n)| - A_n|P_{12}(x_1,x_n)|)_{\min} > 0 \tag{39}$$

or

$$(A_n|P_{12}(x_1,x_n)| - B_n|P_{11}(x_1,x_n)|) > 0. \tag{40}$$

With the intention of writing the inequalities (39) and (40) in a different form, we start by rewriting the left-hand member of (39). Utilizing the identity

$$\begin{aligned} &B_n|P_{11}(x_1,x_n)| - A_n|P_{12}(x_1,x_n)| \\ &\quad = \tfrac{1}{2}(A_n + B_n)(|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)|) \\ &\qquad - \tfrac{1}{2}(A_n - B_n)(|P_{11}(x_1,x_n)| + |P_{12}(x_1,x_n)|), \end{aligned} \tag{41}$$

we obtain, with the aid of (36a),

$$\begin{aligned} &(B_n|P_{11}(x_1,x_n)| - A_n|P_{12}(x_1,x_n)|)_{\min} \\ &\quad = \tfrac{1}{2}(A_n + B_n)(|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)|)_{\min} \\ &\qquad - \tfrac{1}{2}(A_n - B_n)(|P_{11}(x_1,x_n)| + |P_{12}(x_1,x_n)|)_{\max} \\ &\quad = \tfrac{1}{2}(A_n + B_n)(|P_{11}(x_1,x_n)|_{\max} - |P_{12}(x_1,x_n)|_{\max}) \\ &\qquad - \tfrac{1}{2}(A_n - B_n)(|P_{11}(x_1,x_n)|_{\max} + |P_{12}(x_1,x_n)|_{\max}). \end{aligned} \tag{42}$$

Using Eqs. (33a) and (33b), which were assumed valid for $n-1$, and the definitions (12), (19), and (20), we obtain from (42)

$$\begin{aligned} &(B_n|P_{11}(x_1,x_n)| - A_n|P_{12}(x_1,x_n)|)_{\min} \\ &\quad = \tfrac{1}{2}(A_n + B_n)(M_{n-1} - H_{n-1}) \\ &\qquad - \tfrac{1}{2}(A_n - B_n)(M_{n-1} + H_{n-1}) = \Gamma_n(n). \end{aligned} \tag{43}$$

By means of (43), the relation (39) can be written

$$\Gamma_n(n) > 0. \tag{44}$$

Consider next the left-hand member of (40). With the aid of (41) and (36b) we get

$$\begin{aligned} &(A_n|P_{12}(x_1,x_n)| - B_n|P_{11}(x_1,x_n)|)_{\min} \\ &\quad = \tfrac{1}{2}(A_n - B_n)(|P_{11}(x_1,x_n)| + |P_{12}(x_1,x_n)|)_{\min} \\ &\qquad - \tfrac{1}{2}(A_n + B_n)(|P_{11}(x_1,x_n)| - |P_{12}(x_1,x_n)|)_{\max} \\ &\quad = \tfrac{1}{2}(A_n - B_n)(|P_{11}(x_1,x_n)|_{\min} + |P_{12}(x_1,x_n)|_{\min}) \\ &\qquad - \tfrac{1}{2}(A_n + B_n)(|P_{11}(x_1,x_n)|_{\min} - |P_{12}(x_1,x_n)|_{\min}). \end{aligned} \tag{45}$$

Using (33c) and (33d), assumed valid for $n-1$, we obtain from (45)

$$(A_n|P_{12}(x_1,x_n)| - B_n|P_{11}(x_1,x_n)|)_{\min} = \begin{cases} \tfrac{1}{2}(A_n - B_n)(S_{n-1} + G_{n-1}) \\ \qquad - \tfrac{1}{2}(A_n + B_n)(S_{n-1} - G_{n-1}), & \text{if } G_{n-1} > 0, \\ -B_n P_{n-1}, & \text{if } G_{n-1} \leqslant 0. \end{cases} \tag{46}$$

Since $-B_n P_{n-1} < 0$, it follows from (46) that the relation

(40) can equivalently be written as

$$\tfrac{1}{2}(A_n - B_n)(S_{n-1} + G_{n-1}) - \tfrac{1}{2}(A_n + B_n)(S_{n-1} - G_{n-1}) > 0. \quad (47)$$

Making use of (23), (25a), and the definitions (12)–(15), we can write the inequality (47) in the simplified form

$$\Gamma_n(m) > 0, \text{ where } m \neq n. \quad (48)$$

Utilizing the equivalence between (39) and (44) and between (40) and (48), we obtain from (38)–(40)

$$|P_{11}(x_1, x_{n+1})|_{\min} > 0 \Leftrightarrow \Gamma_n(m) > 0. \quad (49)$$

Since, according to the definitions (13) and (14),

$$G_n = \Gamma_n(m), \quad (50)$$

Eq. (48) also can be written as

$$|P_{11}(x_1, x_{n+1})|_{\min} > 0 \Leftrightarrow G_n > 0. \quad (51)$$

By making minor changes in the derivation proceeding from (38) to (51), starting instead from the statement

$$|P_{11}(x_1, x_{n+1})|_{\min} > 0 \quad (52a)$$

implies that either

$$|P_{11}(x_1, x_{n+1})|_{\min} = (B_n |P_{11}(x_1, x_n)| - A_n |P_{12}(x_1, x_n)|)_{\min} > 0 \quad (52b)$$

or

$$|P_{11}(x_1, x_{n+1})|_{\min} = (A_n |P_{12}(x_1, x_n)| - B_n |P_{11}(x_1, x_n)|)_{\min} > 0, \quad (52c)$$

one can easily prove that

$$|P_{11}(x_1, x_{n+1})|_{\min} > 0 \Rightarrow |P_{11}(x_1, x_{n+1})|_{\min} = G_n. \quad (53)$$

From (51) and (53) we conclude that

$$|P_{11}(x_1, x_{n+1})|_{\min} = \begin{cases} 0, & \text{if } G_n \leqslant 0, \\ G_n, & \text{if } G_n > 0. \end{cases} \quad (54)$$

The assertion (33d) is thus shown to be true for n.

Proof of (33c): Inserting (54) into the formula

$$|P_{12}(x_1, x_{n+1})|^2 - |P_{11}(x_1, x_{n+1})|^2 = P_n^2, \quad (55)$$

which follows from (36) in I and (17) in the present paper, we find, with the aid of (18),

$$|P_{12}(x_1, x_{n+1})|_{\min} = \begin{cases} \sqrt{0 + P_n^2} = P_n, & \text{if } G_n \leqslant 0, \\ \sqrt{G_n^2 + P_n^2} = S_n, & \text{if } G_n > 0, \end{cases}$$

which means that (33c) is valid for n.

Proof of (33b): Using (37a), the definitions (19) and (20), and also (33a) and (33b), assumed valid for $n - 1$, we get

$$|P_{11}(x_1, x_{n+1})|_{\max} = B_n |P_{11}(x_1, x_n)|_{\max} + A_n |P_{12}(x_1, x_n)|_{\max} = B_n M_{n-1} + A_n H_{n-1} = \tfrac{1}{2}(M_{n-1} + H_{n-1})(A_n + B_n) - \tfrac{1}{2}(M_{n-1} - H_{n-1})(A_n - B_n) = H_n. \quad (56)$$

The assertion (33b) is thus true for n.

Proof of (33a): From (55), (56), and (22) we obtain

$$|P_{12}(x_1, x_{n+1})|_{\max} = \sqrt{H_n^2 + P_n^2} = M_n,$$

which means that Eq. (33a) holds for n.

We discuss the case when n is even. It has now been shown that the assertions (33a)–(33d) are true for n if they are true for $n - 1$, provided that n is an odd integer $\geqslant 3$. The induction from $n - 1$ to n, in the case that n is an even integer $\geqslant 2$, can be carried out in nearly the same way and will therefore not be repeated here. Assertions (33a)–(33d) are thus proved.

V. MAXIMUM AND MINIMUM TRANSMISSION

Let us consider all the possible shapes of a barrier system, consisting of N barriers, that are compatible with the requirement that the quantities B_k/A_k $(k = 1,2,...,N)$ assume certain given values for a particular energy E of the incident particle. These shapes together constitute a family of barrier systems. From formulas (43a) and (43b) in I for the transmission coefficient, and Eqs. (33a)–(33d) and (17) in the present paper, we conclude that the maximum one of the transmission coefficients (for the particular energy E), pertaining to those barrier systems that are members of the above-mentioned family, is given by

$$\max T(E) = \begin{cases} 1, & \text{if } G_n(E) \leqslant 0, \\ (P_N/S_N)^2, & \text{if } G_n(E) > 0, \end{cases} \quad (57)$$

and the minimum one of the transmission coefficients (for the same E) by

$$\min T(E) = (P_N/M_N)^2, \quad (58)$$

the quantities G_n, S_n, P_n, and M_n being defined by (13)–(16) and by (20). Equations (57) and (18) imply that $\max T(E) < 1$, in case $G_N(E) > 0$.

Since the reflection coefficient is given by the formula $R = 1 - T$, we immediately obtain, with the aid of (57), (58), (18), and (22),

$$\min R(E) = \begin{cases} 0, & \text{if } G_N(E) \leqslant 0, \\ (G_N/S_N)^2, & \text{if } G_N(E) > 0, \end{cases} \quad (59)$$

$$\max R(E) = (H_N/M_N)^2, \quad (60)$$

the quantity H_n being defined by (19).

We realize from (10) and (14) that the condition $G_N(E) > 0$, occurring in (57) and (59), can be regarded as a condition on the sizes of the single barriers. Since these sizes are functions of the energy E of the incident particle, it is clear that the condition $G_N(E) > 0$, although fulfilled in one energy interval, very well may be violated in another energy region.

For convenience let us call the quantity $\alpha_n(E)$, defined by (29b) in I, the size of the well between the barriers $B(n)$ and $B(n + 1)$. Except for the particle energies near the top of either of the barriers $B(n)$ and $B(n + 1)$, the quantity α_n is approximately equal to the quantity L_n, defined by (7) in I [cf. the text from (B8) to the end of Appendix B in I]. We realize that changing the distances between the barrier humps provides a means of varying the sizes α_n of the wells.

Using the new terminology, we can now say that the result (57) gives an exact expression for that maximum value of the transmission coefficient $T(E)$ (pertaining to a given system of N potential barriers and a fixed energy E of the incident particle) that can be obtained by varying the shape of the given potential $V(x)$ in all possible ways subject to the condition that the sizes of the single barriers, relative to the

fixed energy E, remain unchanged. The potential $V(x)$ involved in this maximum problem is thus specified to the extent that the sizes of the single barriers, relative to the fixed energy E, will have certain given values. Otherwise, the potential can be freely deformed, which means that the sizes of the single wells, relative to the fixed E, are freely variable. By means of the phase-integral expressions given in Appendix B of I, the exact formulas above can be converted into accurate approximate ones, suitable for evaluation.

VI. A NECESSARY CONDITION, ON THE SIZES OF THE SINGLE BARRIERS, FOR COMPLETE TRANSPARENCY

It follows from the six equations in I—(43a), (43b), (28), (29a), (29b), and (33)—that the transmission properties of a barrier system depend solely upon the sizes of the single barriers and the sizes of the wells between them. From (57) and (18) we infer that max $T(E)$ is equal to unity, only if $G_N(E) \leqslant 0$. Hence, for a system of N barriers and incident particles of energy E, complete transparency is impossible when $G_N(E) > 0$, whatever the sizes of the wells between the single barriers. In other words, formula (57) implies that $G_N(E) \leqslant 0$ is a necessary condition, on the sizes of the single barriers, for total transmission of incident particles of energy E through a system of N barriers of arbitrary shapes.

The result may be understood as follows. From (A4a) in Appendix A we realize that the relation $G_N(E) > 0$ holds when the size of the biggest barrier $B(m)$, i.e., $(B_m/A_m)^2$, exceeds the value $[H_N(m)/M_N(m)]^2$, which is a function of the sizes of the other barriers in the system. Formula (57) thus implies that if, for an energy E, the biggest barrier $B(m)$ is too big in comparison with the rest of the barriers, we cannot have complete transparency for this E, even if the sizes of the wells between the barriers assume their most favorable values. The physical meaning of the quantity $[H_N(m)/M_N(m)]^2$ is discussed in the text below (A6) in Appendix A. It is shown there that $[(H_N(m)/M_N(m)]^2$ is equal to the maximum reflection coefficient that can be attained, for a given energy E, by any barrier system consisting of $N-1$ barriers having, relative to this E, the same sizes as the barriers of the N barrier system considered, from which the barrier $B(m)$ has been excluded.

Transmission through a general system of two barriers is treated in Ref. 5. Some comments will be added here. From (57) and (A4b), with the aid of (A1), (A2), and (13), we find that the necessary condition for total transmission in this case, $G_2(E) \leqslant 0$, can be written equivalently as $B_1/A_1 = B_2/A_2$. We alternatively arrive at the same result by using (14) instead of (A4b). In the special case of a symmetric barrier system, it follows from (B10a) and (B10b) in Appendix B that the relation $B_1/A_1 = B_2/A_2$ actually holds for all values of E, if x_2 on the real axis is chosen to coincide with the point of symmetry, and in Ref. 5 it is shown that total transmission in fact occurs for certain discrete values of E, namely for those satisfying Eq. (31) in Ref. 5 (see pp. 639 and 640 in Ref. 5).

From Eqs. (43b), (33), and (28) in I we obtain for the transmission coefficient T, pertaining to a general system of two barriers, the exact formula

$$1/T = 1 + |A_2B_1 + A_1B_2 \exp(2i\beta_1)|^2 \exp(2K_1 + 2K_2), \tag{61}$$

which shows that even in the case of a nonsymmetric double barrier complete transparency is possible under special circumstances. Indeed, if the relation $B_1/A_1 = B_2/A_2$ incidentally happens to be fulfilled for any of the energies E that satisfy the equation $\exp(2i\beta_1) = -1$ [identical to (31) in Ref. 5], we will certainly have total transmission for this E.

ACKNOWLEDGMENTS

I wish to thank Professor N. Fröman and Professor P. O. Fröman for reading the manuscript, and for making valuable comments on the presentation.

APPENDIX A: THE PHYSICAL MEANING OF THE CONDITIONS $G_N(E) > 0$ AND $G_N(E) \leqslant 0$

Let us consider an arbitrary system of N potential barriers and an incident particle of energy E.

We define, for $i = 1,2,...,n$ and $n \leqslant N$,

$$H_n(i) = \frac{1}{2}\left(\frac{1}{A_i + B_i}\prod_{k=1}^{n}(A_k + B_k) - \frac{1}{A_i - B_i}\prod_{k=1}^{n}(A_k - B_k)\right), \tag{A1}$$

$$M_n(i) = \frac{1}{2}\left(\frac{1}{A_i + B_i}\prod_{k=1}^{n}(A_k + B_k) + \frac{1}{A_i - B_i}\prod_{k=1}^{n}(A_k - B_k)\right). \tag{A2}$$

In particular, the quantities $H_n(n)$ and $M_n(n)$ are identical to H_{n-1} and M_{n-1}, respectively, which were earlier defined in (19) and (20). With the aid of (A1) and (A2), we obtain from (13) and (14) the formula

$$\begin{aligned} G_N &= \frac{1}{2}(A_m + B_m)(M_N(m) - H_N(m)) \\ &\quad - \frac{1}{2}(A_m - B_m)(M_N(m) + H_N(m)) \\ &= -A_m H_N(m) + B_m M_N(m) \\ &= A_m M_N(m)\left(\frac{B_m}{A_m} - \frac{H_N(m)}{M_N(m)}\right), \end{aligned} \tag{A3}$$

where, according to (11), $A_m M_N(m) > 0$. Hence,

$$G_N > 0 \Leftrightarrow \frac{B_m}{A_m} > \frac{H_N(m)}{M_N(m)}, \tag{A4a}$$

$$G_N \leqslant 0 \Leftrightarrow \frac{B_m}{A_m} \leqslant \frac{H_N(m)}{M_N(m)}, \tag{A4b}$$

$$N \geqslant 1,$$

Using (A1), (A2), and (11), we can easily show that

$$\frac{B_k}{A_k} < \frac{H_N(m)}{M_N(m)} < 1, \quad \text{for } k = 1,2,...,N \text{ but } k \neq m, \tag{A5}$$

provided $N \geqslant 3$. For $N = 2$ we have instead

$$\frac{B_k}{A_k} = \frac{H_2(m)}{M_2(m)} < 1, \quad \text{for } k = 1,2 \text{ but } k \neq m. \tag{A6}$$

We recall that $(B_m/A_m)^2$ is the size of the biggest barrier $B(m)$ in the considered system of N potential barriers;

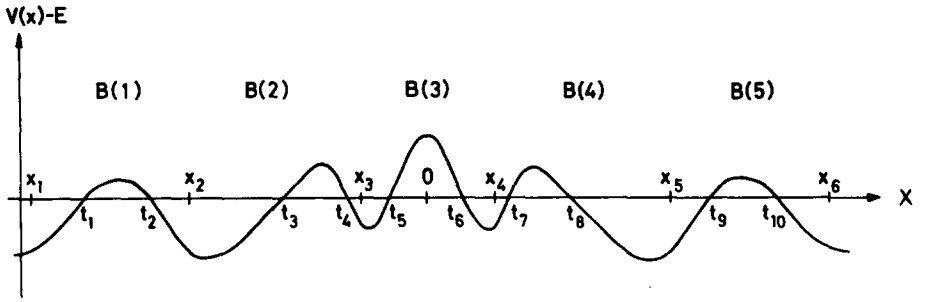


FIG. 4. An example of a symmetrical system of five potential barriers. Figures 1(a), 1(b) and 2(a)–2(d) in Ref. 1, illustrating the general case of a system of N potential barriers of arbitrary shapes, contain information concerning our application of the phase-integral method, which is entirely relevant also to this special case.

cf. (10) and (13). Furthermore, it follows from (60) and the definitions (19), (20), (A1), and (A2) that the quantity $[H_N(m)/M_N(m)]^2$ is equal to the maximum attainable value, for a fixed energy E, of the reflection coefficient pertaining to a variable system of $N-1$ barriers that is allowed to assume every possible shape compatible with the requirement that the single barriers shall have the same sizes, relative to this very E, as the barriers of the given N barrier system with the barrier $B(m)$ omitted.

Remembering also (6), we see that one can roughly understand the relation (A4b) as follows: $G_N(E)\leqslant 0$, if and only if the reflection coefficient (for E) pertaining to the biggest barrier in the system is less than or equal to the maximum one of the reflection coefficients (for the same E) pertaining to all those systems of $N-1$ barriers that, relative to this E, have the same sizes as the barriers of the N barrier system considered, from which the biggest barrier has been omitted.

APPENDIX B: SYMMETRY RELATIONS FOR TWO BARRIERS WHICH ARE MIRROR IMAGES OF EACH OTHER

We shall start by writing down, without proof, some relations that are valid for a symmetrical system of N potential barriers [see (29)], provided that the points $x_1,x_2,\ldots,x_{N+1}$ on the real axis are chosen to be symmetrically located with respect to the point of symmetry.

We have

$$w_{2n-1}(x_n) = \pm w_{2N-2n+2}(x_{N-n+2}), \quad n = 1,2,\ldots,N, \tag{B1a}$$

$$w_{2n}(x_{n+1}) = \pm w_{2N-2n+1}(x_{N-n+1}), \quad n = 1,2,\ldots,N, \tag{B1b}$$

where the upper sign pertains to the case when N is odd, and the lower sign to the case when N is even. The quantities $w_{2n+1}(x_{n+1})$ and $w_{2n}(x_{n+1})$ are defined by (8a)–(8c) in I.

If, in the symmetric barrier system, the barriers $B(i)$ and $B(k)$, $i<k$, are symmetrically situated with respect to the point of symmetry, it is true that

$$K_i = K_k, \tag{B2a}$$

$$L_i = L_{k-1}. \tag{B2b}$$

The relations to be derived below are independent of the number of barriers in the system. For the sake of simplicity we shall therefore derive these relations by considering a small, easily handled barrier system. Let us consider a system of five potential barriers that is symmetric with respect to a point, chosen as origin. The potential $V(x)$ thus satisfies, for all real values of x, the equation

$$V(x) = V(-x). \tag{B3}$$

By choosing $x_1 = -x_6$, $x_2 = -x_5$, and $x_3 = -x_4$, we achieve the result that the barriers B(2) and B(4), defined within the intervals (x_2,x_3) and (x_4,x_5), respectively, become mirror images of each other (mirror-symmetric). See Fig. 4. It is reasonable, although not necessary, to let the points x_2, x_3, x_4, and x_5 coincide with the minima of the potential function $V(x)$. If, however, we allow these symmetrically situated points to be chosen more freely, as in Fig. 4, we realize that the barriers B(2) and B(4), within their intervals (x_2,x_3) and (x_4,x_5), may assume shapes of a more general kind than is possible for any two mirror-symmetric barriers in a system consisting of less than five barriers. This is the motivation for the choice above of precisely five barriers.

Using Eqs. (A.5b) and (A.10) in Ref. 6, and the inversion formula (A15) in Appendix A of I, we obtain

$$F_{11}(x_2,x_3) = -F_{22}(x_4,x_5) \times \exp(i[w(x_5) - w(x_4) + w(x_3) - w(x_2)]), \tag{B4a}$$

$$F_{12}(x_2,x_3) = F_{12}(x_4,x_5) \times \exp(i[w(x_5) + w(x_4) - w(x_3) - w(x_2)]). \tag{B4b}$$

Inserting $N=5$ in (B1b), we obtain, for $n=1$,

$$w_2(x_2) = w_9(x_5) \tag{B5a}$$

and, for $n=2$, we get

$$w_4(x_3) = w_7(x_4). \tag{B5b}$$

From (B2a) and (B2b) we find that

$$K_2 = K_4, \tag{B6a}$$

$$L_1 = L_4. \tag{B6b}$$

With the aid of Eqs. (10b), (23b), and (9) in I, and (B5a), (B5b), (B6a), and (B6b) in the present paper, we can write (B4a) and (B4b) as follows:

$$F_{11}(x_2,x_3) = -F_{22}(x_4,x_5)\exp(-2K_2), \tag{B7a}$$

$$F_{12}(x_2,x_3) = F_{12}(x_4,x_5)\exp[i(2\Theta_3 - 2\Theta_1)]. \tag{B7b}$$

Utilizing Eqs. (27a), (27b), (27d), and (19) in I, we find from (B7a) and (B7b) the relations

$$F_{11}(t_3,2;x_2,x_3) = F^*_{11}(t_7,4;x_4,x_5), \tag{B8a}$$

$$F_{12}(t_3,2;x_2,x_3) = F_{12}(t_7,4;x_4,x_5), \tag{B8b}$$

$$F_{21}(t_3,2;x_2,x_3) = F_{21}(t_7,4;x_4,x_5), \tag{B8c}$$

$$F_{22}(t_3,2;x_2,x_3) = F^*_{22}(t_7,4;x_4,x_5). \tag{B8d}$$

The elements of the matrix $\mathbf{F}(t_{2n-1},n;x_n,x_{n+1})$ are

completely determined by the behavior of the function $V(x) - E$ within the interval (x_n, x_{n+1}) pertaining to the barrier $B(n)$. This follows from the definitions of the F-matrix elements by (3.22a)–(3.22d) in Ref. 4 and from the definition of the function $q(z)$ by (A36) in Appendix A in I, taking into consideration that the problems connected with the lower limit of integration in the integral defining $w(z)$ and the choice of phase of $q(z)$ on the real axis are taken care of by the definition of the matrix $\mathbf{F}(t_{2n-1}, n; x_n, x_{n+1})$.

We conclude that the validity of the relations (B8a)–(B8d) is entirely independent of the behavior of the potential $V(x)$ in the regions outside the intervals (x_2, x_3) and (x_4, x_5) pertaining to the barriers B(2) and B(4), respectively. Accordingly, the relations (B8a)–(B8d) are independent of the other barriers in the system, i.e., independent of their shapes and also of the number of other barriers. We thus realize that the relations (B8a)–(B8d) are due solely to the mirror symmetry of the two barriers B(2) and B(4), which are defined by the potential $V(x)$ within the intervals (x_2, x_3) and (x_4, x_5), respectively. The same relations must therefore be valid for any two barriers, which are mirror images of each other, whatever positions they may occupy in a barrier system, provided of course that the arguments of the pertinent F matrices are changed so as to indicate the actual positions, in the barrier system, of the two symmetric barriers.

Hence, if the barriers $B(i)$ and $B(k)$ in a barrier system are mirror images of each other, we have

$$F_{11}(t_{2i-1}, i; x_i, x_{i+1}) = F_{11}^{*}(t_{2k-1}, k; x_k, x_{k+1}), \quad \text{(B9a)}$$

$$F_{12}(t_{2i-1}, i; x_i, x_{i+1}) = F_{12}(t_{2k-1}, k; x_k, x_{k+1}), \quad \text{(B9b)}$$

$$F_{21}(t_{2i-1}, i; x_i, x_{i+1}) = F_{21}(t_{2k-1}, k; x_k, x_{k+1}), \quad \text{(B9c)}$$

$$F_{22}(t_{2i-1}, i; x_i, x_{i+1}) = F_{22}^{*}(t_{2k-1}, k; x_k, x_{k+1}). \quad \text{(B9d)}$$

From these formulas and (18a)–(18d) in I, we get the useful relations

$$A_i = A_k, \quad \text{(B10a)}$$

$$B_i = B_k, \quad \text{(B10b)}$$

$$\sigma_i = \sigma_k, \quad \text{(B10c)}$$

$$\tau_i = -\tau_k \quad \text{(B10d)}$$

between the barrier characteristics of two mirror-symmetric barriers $B(i)$ and $B(k)$.

Let us finally regard the particular case when the barrier $B(i)$ in itself is symmetric with respect to the center $\frac{1}{2}(x_i + x_{i+1})$ of the barrier interval (x_i, x_{i+1}), i.e., when $V(x) = V(x_i + x_{i+1} - x)$ for every x belonging to the barrier interval. Formulas (B10a)–(B10d) naturally remain valid also for this particular case. However, from (B10d), considering the fact that the barriers $B(i)$ and $B(k)$ in this case are of equal shape, which means that $\tau_i = \tau_k$ [see (18d) in I], we now simply obtain

$$\tau_i = 0. \quad \text{(B11)}$$

[1] Ö. Dammert, "Transmission through a system of potential barriers. I. Transmission coefficient," J. Math. Phys. **24**, 2163 (1983).

[2] Ö. Dammert, "Transmission through a system of N identical potential barriers of a general shape," to be published.

[3] Ö. Dammert, "Bound state energies for a general potential well containing N barrier humps of arbitrary shapes," to be published.

[4] N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) [Russian translation (MIR, Moscow, 1967)].

[5] N. Fröman and Ö. Dammert, Nucl. Phys. A **147**, 627 (1970).

[6] N. Fröman, Ark. Fys. **32**, 79 (1966).

Another identity among squares of eigenfunctions

R. L. Sachs
Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 14 January 1985; accepted for publication 3 October 1985)

A variant of an identity of H. P. McKean and E. Trubowitz [Commun. Pure Appl. Math. **29**, 143 (1976)] for Hill's equation is derived via contour integration. The identity is $1 = \Sigma_{j=0}^{\infty}(-1)^j |y_2(1,\lambda_j)| \cdot f_j^2(x)$.

I. INTRODUCTION

In this brief paper, a variant of an identity of McKean and Trubowitz[1] is derived via contour integration. The proof is therefore similar to that of Deift and Trubowitz.[2] Before proceeding we introduce notation.

The Hill's operator H is given by $Hf = -D^2 f + qf$, where $D = d/dx$ and $q(x)$ is a real, smooth function of period 1. The functions $y_{1,2}(x,\lambda)$ solve

$$Hy = \lambda y, \tag{1}$$

with initial conditions

$$y_1(0,\lambda) = 1, \quad y_1'(0,\lambda) = 0, \qquad y_2(0,\lambda) = 0, \quad y_2'(0,\lambda) = 1. \tag{2}$$

Define the discriminant

$$\Delta(\lambda) := y_1(1,\lambda) + y_2'(1,\lambda)$$

$$\left(y_2' \equiv \frac{\partial y_2}{\partial x}, \quad y_2 \equiv \frac{\partial y_2}{\partial \lambda}\right).$$

Here, $\Delta(\lambda)$ is the trace of the Floquet matrix and the roots of $\Delta^2 - 4$ correspond to Floquet multipliers ± 1. It is well known (Magnus–Winkler[3]) that these roots $\lambda_0 < \lambda_1 \leqslant \lambda_2 < \lambda_3 \leqslant \lambda_4 < \cdots$ are real and satisfy $\lambda_{2n}, \lambda_{2n-1} = n^2\pi^2 + O(1)$, while the eigenfunctions y_1, y_2 behave like $\cos\sqrt{\lambda}\cdot x$ and $\sin\sqrt{\lambda x}/\sqrt{\lambda}$, respectively. Hence $\Delta(\lambda) \sim 2\cos\sqrt{\lambda}$ for $|\lambda|$ large. Denote by $f_{\pm}(t,\lambda)$ the Bloch eigenfunctions $y_1(t,\lambda) + m_{\pm}(\lambda)\, y_2(t,\lambda)$. We are now ready to prove the main result.

II. THEOREM AND PROOF

Theorem: Let $f_j(x)$ denote the normalized (periodic or antiperiodic) eigenfunction for the eigenvalue λ_j. (If $\lambda_{2j-1} = \lambda_{2j}$ then choose any such eigenfunction.) Then the following identity holds:

$$1 = \sum_{j=0}^{\infty} (-1)^j |y_2(1,\lambda_j)|\, f_j^2(x).$$

Proof: Consider the contour integral

$$I_N = \frac{1}{2\pi i} \oint_{|\lambda| = (N+\frac{1}{2})^2\pi^2} \frac{y_2^2(1,\lambda) f_+(t,\lambda) f_-(t,\lambda)\, d\lambda}{(\Delta(\lambda)/2)^2 - 1}.$$

These estimates quoted above and the fact that $y_2(1,\lambda) f_+(t,\lambda) f_-(t,\lambda)$ corresponds to $y_2(1,\lambda)$ for the translation of q by t [i.e., $q(\cdot + t)$] (see Ref. 1) combine to show that the integrand behaves like $\lambda^{-1}(-1 + O(1))$ for $|\lambda|$ large. For integers $N \to +\infty$, we obtain the limit $I_N \to -1$. Now apply Cauchy's theorem to reduce the integral to a residue sum. The poles are at $\lambda = \lambda_j$ and since $\lambda_{2j} = \lambda_{2j-1}$ implies $y_2(1,\lambda_{2j}) = 0$, the double roots of $\Delta^2 - 4$ do not contribute. At simple roots λ_j the residue is

$$\frac{y_2^2(1,\lambda_j) f_+(t,\lambda_j) f_-(t,\lambda_j)}{\tfrac{1}{2}\Delta(\lambda_j)\dot{\Delta}(\lambda_j)} = \frac{y_2(1,\lambda_j)[-\dot{\Delta}(\lambda_j) f_j^2(t)]}{\tfrac{1}{2}\Delta(\lambda_j)\dot{\Delta}(\lambda_j)} = \pm y_2(1,\lambda_j) f_j^2(t), \quad \text{since } \Delta(\lambda_j) = \pm 2. \tag{3}$$

Note that we used the fact[1] that

$$y_2(1,\lambda_j) f_+(t,\lambda_j) = -\dot{\Delta}(\lambda_j) f_j^2(t).$$

In fact, it is well known that $\Delta(\lambda_j) = 2$ if and only if $j \equiv 0,3 \pmod 4$. The sign of $y_2(1,\lambda_j)$, when it is nonzero, is also easily determined since the roots μ_j of $y_2(1,\mu_j) = 0$ satisfy $\lambda_{2j-1} < \mu_j < \lambda_{2j}$. Thus $y_2(1,\lambda_0) > 0$, $y_2(1,\lambda_1) \geqslant 0$, $y_2(1,\lambda_2) < 0$, etc., so that $y_2(1,\lambda_j) \geqslant 0$ for $j \equiv 0,1 \pmod 4$ (and is < 0 for $j \equiv 2,3$). Combining these two facts we conclude that the coefficient of $f_j^2(t)$ is the identity (*) is nonnegative for even values of j and nonpositive for odd values of j.

III. REMARKS

(1) As in the analogous identity of McKean and Trubowitz, which is

$$1 = \sum_{j=0}^{\infty} \epsilon_j f_{2j}^2(x) \quad \begin{array}{l}(\text{where } \epsilon_j > 0 \text{ with equality} \\ \text{only at double roots } \lambda_{2j-1} = \lambda_{2j}).\end{array} \tag{4}$$

the identity (*) leads to a family of trace identities, the first of which expresses $q(x)$ as a sum of $f_j^2(x)$.

(2) In discussing the structure of isospectral manifolds for Hill's equation, McKean and Trubowitz[1] use as the "origin" the point where $\mu_j = \lambda_{2j-1}$. At such a point the odd terms vanish and we conjecture that (*) reduces to the identity (4).

(3) Periodicity imposes constraints on the λ_j's, so that the set $\{\lambda_{2j}\}$ determines all the λ_j's. It is not immediately clear how to absorb this dependency into the identity (*).

[1] H. P. McKean and E. Trubowitz, "Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points," Commun. Pure Appl. Math. **29**, 143 (1976).
[2] P. Deift and E. Trubowitz, "An identity among squares of eigenfunctions," Commun. Pure Appl. Math. **34**, 713 (1981).
[3] W. Magnus and S. Winkler, *Hill's Equation* (Wiley, New York, 1966).

 0022-2488/86/020471-01$02.50 © 1986 American Institute of Physics

Fermions in the space-time $R \times S^3$

Diptiman Sen
Physics Department, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 9 July 1985; accepted for publication 30 September 1985)

Quantum field theories on the surface of a four-dimensional sphere are considered. The Hamiltonian is rotation invariant and its eigenvalues are discrete. Scalar, vector, and spinorial functions on S^3 are discussed. The most general Lagrangians for Dirac, Weyl, and Majorana fermions are derived. They are different from the ones in existing literature. The wave functions and propagator are obtained and formulas for matrix elements involving spinors are presented. The discrete symmetries—parity, charge conjugation, and time reversal—are described. The Lagrangian in $R \times S^3$ transforms in a nontrivial way under these. Finally, the fermionic Lagrangian is rederived using the tetrad formalism, and conformal transformations are discussed. This leads to a generalization of the formalism to a time-dependent radius of curvature. As a particular case, a new Lagrangian for de Sitter space is obtained, which, however, is not invariant under the full de Sitter group.

I. INTRODUCTION

We have two reasons for considering a curved space-time manifold. First, a physical system in a nontrivial gravitational background may exhibit interesting behavior not obtained in flat space. Demanding that we recover the Minkowski theory in the limit of vanishing curvature may eliminate certain classes of theories.[1] Second, if the space is chosen to be compact, the particle modes are discrete. If a high-momentum cutoff is imposed, a quantum theory with a finite number of degrees of freedom is obtained. This is an alternative to making space discrete, which leads to problems in theories with chiral fermions and supersymmetry.

The simplest way to quantize a theory is through the Hamiltonian formalism. For this purpose, it is useful to keep the temporal dimension flat and compactify space alone. The sphere S^3 is maximally symmetric and is therefore preferable to a rectangular box with periodic boundary conditions. One might expect the spherical curvature to complicate the calculations but the special group theoretic properties of S^3 result in considerable simplification.

Scalar and vector fields have been described in Refs. 2 and 3. After discussing these, we consider spin-$\frac{1}{2}$ fermions and their Lagrangian in Sec. II. The wave equation can be derived solely from the requirement that the theory have the global symmetries of $R \times S^3$. The energy spectrum, wave function, and propagator for fermions are derived and major differences from flat space are pointed out. In Sec. III, formulas for spinorial matrix elements are presented. In Sec. IV, we discuss the three discrete symmetries—parity, charge conjugation, and time reversal. Though the theory is invariant under these, the Lagrangian itself transforms in a nontrivial way under the first two. In Sec. V, we return to fermions and rederive their Lagrangian using the tetrad formalism. Following this, conformal transformations are discussed and our formalism generalized to the case of a time-dependent radius of curvature. This enables us to obtain the Dirac and Weyl Lagrangians in de Sitter space. Although they do not have the full symmetry of the de Sitter group, they provide a good description of fermions, and the wave functions and spectra can be directly read off from Sec. II. Finally, we make a comment on the Weyl spectrum that is relevant to a forthcoming paper on supersymmetry in $R \times S^3$.

A. Notation and choice of tetrad

The sphere S^3 can be parametrized in different ways. One can use four Cartesian coordinates x_m, two complex variables (u,v), or three angles (θ,α,β). Setting the radius equal to unity,

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1 , \tag{1}$$

we define[2,4]

$$u = x_1 + ix_2 = \cos\theta\, e^{i\alpha} , \quad v = x_3 + ix_4 = \sin\theta\, e^{i\beta} , \tag{2}$$

where $0 \leqslant \theta \leqslant \pi/2$, $0 \leqslant \alpha, \beta < 2\pi$. The volume element is

$$d\Omega = (1/2\pi^2)\, d\alpha\, d\beta \sin\theta \cos\theta\, d\theta . \tag{3}$$

The manifold has an O(4) rotational symmetry, whose Lie algebra is SU(2)×SU(2). The lowering operators of the two SU(2) subgroups $\mathbf{L}_1$ and $\mathbf{L}_2$ are

$$L_{1-} = u^* \partial_v - v^* \partial_u , \quad L_{2-} = u^* \partial_{v^*} - v \partial_u . \tag{4}$$

For scalar functions on the sphere, the two $L_{\tau 3}$ can be diagonalized with eigenvalues s_τ, $\tau = 1,2$. As differential operators,

$$\begin{aligned} L_{13} &= \tfrac{1}{2}(u \partial_u + v \partial_v - u^* \partial_{u^*} - v^* \partial_{v^*}) , \\ L_{23} &= \tfrac{1}{2}(u \partial_u - v \partial_v - u^* \partial_{u^*} + v^* \partial_{v^*}) . \end{aligned} \tag{5}$$

We define certain combinations of these operators

$$\begin{aligned} J_i &= L_{1i} + L_{2i} = -i\epsilon_{ijk} x_j \partial_k , \\ K_i &= L_{1i} - L_{2i} = i(x_4 \partial_i - x_i \partial_4) , \\ J_3 &= -i \partial_\alpha , \quad \mathbf{K}_3 = -\mathrm{i}\, \partial_\beta , \end{aligned} \tag{6}$$

and the O(4) rotations

$$\begin{aligned} M_{ij} &= \tfrac{1}{2} \epsilon_{ijk} J_k , \\ M_{i4} &= \tfrac{1}{2} K_i , \\ M_{mn} &= -i/2(x_m \partial_n - x_n \partial_m) . \end{aligned} \tag{7}$$

 0022-2488/86/020472-11$02.50 © 1986 American Institute of Physics

The commutation relations of **J** and **K** are

$$[J_i, J_j] = [K_i, K_j] = i\epsilon_{ijk} J_k ,$$
$$[J_i, K_j] = i\epsilon_{ijk} K_k . \tag{8}$$

These correspond, in the limit of large radius, to rotations and spatial translations of Minkowski space. Near $x_4 = 1$, J_i rotates about the x_i axis and $-K_i$ translates along it. The three boost symmetries of M^4 are completely missing in $R \times S^3$.

Define a parity operation **P**, which interchanges $\mathbf{L}_1$ and $\mathbf{L}_2$, as

$$(x_1, x_2, x_3, x_4) \to (-x_1, -x_2, -x_3, x_4) ,$$
$$u \to -u , \quad v \to -v^* . \tag{9}$$

The metric for (t,θ,α,β) is $(-1, 1, \cos^2\theta, \sin^2\theta)$.

To write down a wave equation for fermions, one must construct a tetrad field.[5] This consists of a frame of four orthonormal vectors at each point in space-time. Since the time direction is flat in our case, we only have to find a right-handed triad on the sphere which is continuous everywhere. This is possible since S^3 is parallelizable. A convenient choice is

$$\hat{e}_1^{(1)} = (x_4, x_3, -x_2, -x_1) ,$$
$$\hat{e}_2^{(1)} = (-x_3, x_4, x_1, -x_2) , \tag{10}$$
$$\hat{e}_3^{(1)} = (x_2, -x_1, x_4, -x_3) .$$

Here (a,b,c,d) denotes the Cartesian components of a vector. Derivatives along these axes will be denoted by $\partial_i^{(1)}$ and the components of a vector **V** by $V_i^{(1)} = \mathbf{V} \cdot \hat{e}_i^{(1)}$. It is easily verified that

$$\partial_i^{(1)} = -2iL_{1i} , \quad [\partial_i^{(1)}, \partial_j^{(1)}] = 2\epsilon_{ijk}\, \partial_k^{(1)} . \tag{11}$$

The second equation shows these unit vectors have a non-zero twist due to the curvature. The divergence and curl of a vector are given by

$$\nabla \cdot \mathbf{V} = \partial_i^{(1)} V_i^{(1)} ,$$
$$[\nabla \times \mathbf{V}]_i^{(1)} = \tfrac{1}{2}\epsilon_{ijk} [\partial_j^{(1)} V_k^{(1)} - \partial_k^{(1)} V_j^{(1)}] - 2V_i^{(1)} . \tag{12}$$

The $-2V_i^{(1)}$ in the curl formula follows directly from (11) and is consistent with the curl of the divergence of a scalar function being zero.

It is important to note that a different frame might have been chosen:

$$\hat{e}_1^{(2)} = (x_4, -x_3, x_2, -x_1) ,$$
$$\hat{e}_2^{(2)} = (x_3, x_4, -x_1, -x_2) , \tag{13}$$
$$\hat{e}_3^{(2)} = (-x_2, x_1, x_4, -x_3) .$$

Then

$$\partial_i^{(2)} = 2i L_{2i} , \quad [\partial_i^{(2)}, \partial_j^{(2)}] = -2\epsilon_{ijk}\, \partial_k^{(2)} . \tag{14}$$

This triad is twisted in the opposite sense and the curl of a vector in terms of the components $V_i^{(2)} = \mathbf{V} \cdot \hat{e}_i^{(2)}$ is given by

$$[\nabla \times \mathbf{V}]_i^{(2)} = \tfrac{1}{2}\epsilon_{ijk} [\partial_j^{(2)} V_k^{(2)} - \partial_k^{(2)} V_j^{(2)}] + 2V_i^{(2)} . \tag{15}$$

The two frames are related by parity:

$$\hat{e}^{(2)} = -\mathbf{P} \cdot \hat{e}^{(1)} . \tag{16}$$

No matter which choice we make, some formulas later will look unsymmetric between $\mathbf{L}_1$ and $\mathbf{L}_2$. Of course, physics cannot depend on this. We will work with the first set and use Eqs. (10)–(12). The superscript (1) will be dropped from the next section.

We note the commutation relations between the $\hat{e}_i$ and the $L_{\tau i}$:

$$[L_{\tau' i}, \hat{e}_j^{(\tau)}] = i(-1)^{\tau+1}\delta_{\tau\tau'}\epsilon_{ijk}\hat{e}_k^{(\tau)} . \tag{17}$$

Hence the triads are spin-1 objects:

$$[\mathbf{L}^2, \hat{e}^{(\tau')}] = 2\delta_{\tau\tau'}\hat{e}^{\tau} . \tag{18}$$

The reason for this becomes clear in Sec. I C.

B. Scalar fields and Lagrangian

Scalar harmonic functions[2] on the sphere are traceless, symmetric tensors of rank s which belong to the representation $(s/2, s/2)$ of SU(2)×SU(2). The normalized function of highest eigenvalues s_τ is

$$S(s, s/2, s/2) = (s+1)^{1/2}(-u)^s \tag{19}$$

and the other $S(s,a,b)$ are obtained by applying the lowering operators (4). These functions satisfy

$$S^*(s, s_1, s_2) = (-1)^{s_1+s_2} S(s, -s_1, -s_2) ,$$
$$\mathbf{P}S(s, s_1, s_2) = (-1)^s S(s, s_2, s_1) . \tag{20}$$

The dimensionality is $(s+1)^2$ and the value of the Laplacian $\nabla^2 = \partial_i \partial_i$ is $-s(s+2)$. This corresponds to the squared momentum $-p^2$ in flat space.

These functions have the orthogonality property

$$\int d\Omega\, S^*(s, s_1, s_2)\; S(s', s_1', s_2') = \delta_{s,s'}\delta_{s_1,s_1'}\delta_{s_2,s_2'} \tag{21}$$

and the completeness relation

$$\sum_{n=s,s_1,s_2} S^*(n;r')\; S(n;r)$$
$$= (2\pi^2/\sin\theta\cos\theta)\delta(\theta-\theta')\delta(\alpha-\alpha')\delta(\beta-\beta') . \tag{22}$$

The free Lagrangian for a complex, massive, scalar field is

$$\mathcal{L} = \partial_0\Phi^*\,\partial_0\Phi - \partial_i\Phi^*\,\partial_i\Phi - m^2\Phi^*\Phi + 2e\Phi^* i\,\partial_0\Phi . \tag{23}$$

Since we only require invariance under spatial rotation, a term like $\Phi^* i\,\partial_0\Phi$ multiplied by the inverse radius and an arbitrary real coefficient can be present.

The dispersion relations for $\Phi(s,t) = S(s, s_1, s_2)e^{iEt}$ are

$$E(s) = e \pm [e^2 + s(s+2) + m^2]^{1/2} . \tag{24}$$

The spectrum is symmetric about zero without the e term. However, that is no reason for setting it equal to zero.

C. Vector fields and Lagrangian

Vector functions on the sphere are tensors symmetric in s indices and antisymmetric in one more. Hence $\mathbf{r} \cdot \mathbf{V}$ is zero, which is the property that vectors lying on the surface of a sphere should have. The antisymmetry also makes them divergenceless, so they form a basis for the transverse modes of a vector field. They are of two types: $\mathbf{V}^+(s)$, which belong to the representation $((s+1)/2, (s-1)/2)$ and $\mathbf{V}^-(s)$ belonging to $((s-1)/2, (s+1)/2)$. These are related by parity:

$$\mathbf{P}\mathbf{V}^{+}(s, s_1, s_2) = (-1)^{s+1}\mathbf{V}^{-}(s, s_2, s_1) . \tag{25}$$

Complex conjugation gives

$$\mathbf{V}^{\pm *}(s, s_1, s_2) = (-1)^{s_1+s_2+1}\mathbf{V}^{\pm}(s, -s_1, -s_2) . \tag{26}$$

The total dimensionality of $\mathbf{V}^{\pm}(s)$ is $2s(s+2)$. Their curl is given by[2]

$$\nabla \times \mathbf{V}^{\pm}(s) = \mp (s+1)\mathbf{V}^{\pm}(s) . \tag{27}$$

The triads $\hat{e}^{(1)}$ and $\hat{e}^{(2)}$ are the vectors (1,0) and (0,1).

We now make the vector functions explicit. One way of doing this is to introduce an auxiliary radial vector $\mathbf{r}'$ and take its scalar product with $\mathbf{V}$. The vector $\mathbf{V}^{+}(s)$ of maximum s_τ is given by

$$\mathbf{r}' \cdot \mathbf{V}^{+}\left(s, \frac{s+1}{2}, \frac{s-1}{2}\right) = \left(\frac{s}{2}\right)^{1/2}(-u)^{s-1}[u'v - v'u] . \tag{28}$$

The other $\mathbf{r}' \cdot \mathbf{V}^{+}(s, s_1, s_2)$ are obtained by acting on this with $L_{1-} + L'_{1-}$ and $L_{2-} + L'_{2-}$. The $\mathbf{V}^{-}(s, s_1, s_2)$ are found by (25).

We follow a different and more convenient procedure. Instead of the component along the fixed vector $\mathbf{r}'$, consider the projections on the three unit vectors. The quantities $V_i^{\pm} = \mathbf{V}^{\pm} \cdot \hat{e}_i$ are scalar functions. Define

$$V_{\pm}^{\tau} = V_1^{\tau} \pm iV_2^{\tau} . \tag{29}$$

For $\mathbf{V}^{+}(s)$, the $V_i(s)$ are in the representation $((s-1)/2,(s-1)/2)$. This must be so, since the $\hat{e}_i$ belong to (1,0), the $\mathbf{V}^{+}(s)$ to $((s+1)/2,(s-1)/2)$ and the $\hat{e}_i \cdot \mathbf{V}^{+}$ are scalars. Solving Eq. (27) gives

$$V_3^{+}(s, s_1, s_2) = [(s+2s_1+1)(s-2s_1+1)/2s(s+1)]^{1/2} \times S(s-1, s_1, s_2) , \tag{30}$$

$$V_{\pm}^{+}(s, s_1, s_2) = \pm [(s \mp 2s_1 + 1)(s \mp 2s_1 - 1)/2s(s+1)]^{1/2} \times S(s-1, s_1 \pm 1, s_2) ,$$

where $-(s+1)/2 \leqslant s_1 \leqslant (s+1)/2$. The end points $s_1 = \pm (s+1)/2$ means $V_3 = 0$ and $V_{\pm} = 0$ but $V_{\mp}$ is not zero. Since $-(s-1)/2 \leqslant s_2 \leqslant (s-1)/2$, the dimensionality is $s(s+2)$.

For $\mathbf{V}^{-}(s)$, the $V_i(s)$ are $((s+1)/2,(s+1)/2)$ scalars. From (27),

$$V_3^{-}(s, s_1, s_2) = [(s+2s_1+1)(s-2s_1+1)/2(s+2)(s+1)]^{1/2} S(s+1, s_1, s_2) , \tag{31}$$

$$V_{\pm}^{-}(s, s_1, s_2) = \mp [(s \pm 2s_1 + 3)(s \pm 2s_1 + 1)/2(s+2)(s+1)]^{1/2} S(s+1, s_1 \pm 1, s_2) .$$

Here $-(s-1)/2 \leqslant s_1 \leqslant (s-1)/2$ so V_3 and $V_{\pm}$ are never zero. Also $-(s+1)/2 \leqslant s_2 \leqslant (s+1)/2$, so the dimensionality is again $s(s+2)$.

We summarize the above results as follows: with $\tau = \pm 1$,

$$V_3^{\tau}(s, s_1, s_2) = \cos\alpha(s, s_1) S(s-\tau, s_1, s_2) ,$$

$$V_{+}^{\tau}(s, s_1, s_2) = 2^{1/2} \sin\alpha(s, s_1) \cos\beta^{\tau}(s, s_1) \times S(s-\tau, s_1+1, s_2) , \tag{32}$$

$$V_{-}^{\tau}(s, s_1, s_2) = 2^{1/2} \sin\alpha(s, s_1) \sin\beta^{\tau}(s, s_1) \times S(s-\tau, s_1-1, s_2) ,$$

$$\alpha(s, s_1) = \cos^{-1}\left[\frac{(s+2s_1+1)(s-2s_1+1)}{2s(s+1)}\right]^{1/2}$$

$$\beta^{+}(s, s_1) = -\tan^{-1}\left[\frac{\{(s-2s_1)^2 - 1\}}{\{(s+2s_1)^2 - 1\}}\right]^{1/2} , \tag{33}$$

$$\beta^{-}(s, s_1) = \pi - \tan^{-1}\left[\frac{\{s+2s_1+2)^2 - 1\}}{\{(s-2s_1+2)^2 - 1\}}\right]^{1/2} .$$

The normalization is fixed so $\int d\Omega\, \mathbf{V}_i^{*}\mathbf{V}_i = 1$.

If we had chosen the other frame $\hat{e}^{(2)}$, the V_i^{+} would have been $((s+1)/2,(s+1)/2)$ scalars and the V_i^{-} would have been $((s-1)/2,(s-1)/2)$ scalars.

The vector functions written above are complex. One can form combinations of them that are real. In the remainder of this section we assume this has been done.

The most general Lagrangian for a free, massive, Abelian, vector field is given by

$$\mathcal{L} = -\tfrac{1}{4}F_{mn}F^{mn} - \tfrac{1}{2}m^2V_mV^m + \tfrac{1}{2}f_1V_0^2 + if_2V_0\,\partial_iV_i + f_3mV_0 , \tag{34}$$

where

$$F_{0i} = \partial_0V_i - \partial_iV_0 , \quad F_{ij} = \partial_iV_j - \partial_jV_i - 2\epsilon_{ijk}V_k . \tag{35}$$

The $f_{1,2,3}$ are again arbitrary real numbers. Indeed, the last term $f_3 m V_0$ appears if the scalar Lagrangian (23) is gauged and the Higgs mechanism is invoked. It can be eliminated by a shift in V_0, so we ignore it. The following dispersion relations are obtained.

For transverse modes,

$$\mathbf{V} \text{ is a vector function and } V_0 = 0 , \quad E^2(s) = (s+1)^2 + m^2 . \tag{36}$$

The number of modes is $2s(s+2)$.

For longitudinal modes,

$$V_i = \partial_iS(s, s_1, s_2) , \quad V_0 = i[(E^2 - m^2)/(f_2 - E)]S(s, s_1, s_2) , \tag{37}$$

$$E^2(s) = s(s+2)[(m^2 - f_2^2)/(m^2 + f_1)] + m^2 . \tag{38}$$

For these modes, F_{ij} is zero. If $f_1 = f_2 = 0$, the dispersion relation simplifies to

$$E^2(s) = s(s+2) + m^2 . \tag{39}$$

In addition, if $m^2 = 0$, F_{0i} vanishes also. The dimensionality of these modes is $(s+1)^2$. They transform as vectors under rotations.

The spectra of transverse and longitudinal modes are different due to the curvature.

II. FERMIONS

We first describe our notation.[6] The Pauli matrices $\sigma^i = -\bar{\sigma}^i$ are well known. Define a fourth matrix $\sigma^0 = \bar{\sigma}^0 = -\mathbf{I}_2$. Then the gamma matrices are

$$\gamma^m = \begin{pmatrix} 0 & \sigma^m \\ \bar{\sigma}^m & 0 \end{pmatrix}, \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbf{I}_2 & 0 \\ 0 & -\mathbf{I}_2 \end{pmatrix}, \tag{40}$$

and the sigma matrices

$$\sigma^{mn} = \frac{i}{4}[\gamma^m, \gamma^n]\,. \tag{41}$$

For two components spinors, Ψ and $\bar{\Psi}$ will mean Ψ_α and $\bar{\Psi}^{\dot{\alpha}}$ with $\alpha, \dot{\alpha} = 1,2$. The Pauli matrices acting on these carry indices $\sigma^m_{\alpha\dot{\alpha}}$ and $\bar{\sigma}^{m\,\dot{\alpha}\alpha}$. Indices are raised and lowered by the antisymmetric tensor $\epsilon^{\alpha\beta} = -\epsilon_{\dot{\alpha}\dot{\beta}} = i\sigma^2$. The Ψ_α and $\bar{\Psi}^{\dot{\alpha}}$ are related by

$$\bar{\Psi}_a = -(\Psi_a)^+\,. \tag{42}$$

The Dirac equation in a general space-time is derived as follows. Introduce a tetrad field of one-forms,[7] which are coordinate systems that are locally inertial. Denoting the tetrad by e^m, $m = 1,2,3,4$, we have

$$e^m = e^m{}_\mu\, dx^\mu\,. \tag{43}$$

The "flat" index m is lowered by

$$\eta_{mn} = (-1,1,1,1)\,, \tag{44}$$

while the general coordinate index μ is lowered by

$$g_{\mu\nu} = \eta_{mn} e^m{}_\mu e^n{}_\nu\,. \tag{45}$$

The line element

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = e^m \cdot e_m\,. \tag{46}$$

The spin connection one-form $\omega^a{}_b$ is defined by

$$de^a = -\omega^a{}_b \wedge e^b\,, \quad \omega_{ab} = \omega_{\mu ab}\, dx^\mu\,. \tag{47}$$

then the equation for a massive spin-$\frac{1}{2}$ particle is

$$(i\gamma^m e_m{}^\mu D_\mu - m)\Psi = 0\,, \tag{48}$$

with the covariant derivative

$$D_\mu = \partial_\mu + \frac{i}{2}\omega_{\mu mn}\sigma^{mn}\,. \tag{49}$$

This equation arises from an action that is invariant under both general coordinate transformations and local Lorentz rotations of the tetrads.[5] This procedure will be followed in Sec. V.

Here we adopt a different method. This method works due to the special structure of the group of $R \times S^3$, which contains an invariant subgroup generated by $i\partial_0$ and $i\partial_i = 2L_{1i}$. This is just as in Minkowski space where the four translations form an invariant subgroup of the Poincaré group. We can therefore simply write down the equation

$$(i\gamma^m \partial_m - m)\Psi = 0\,, \tag{50}$$

and it is guaranteed that, under a global rotation, Ψ will transform as

$$\Psi'(x') = \exp(i\epsilon_{mn}\sigma^{mn})\Psi(x)\,, \tag{51}$$

where the ϵ_{mn} are constants. Restoring the radius of the sphere to its rightful place, it is easy to see that $-i\partial_i/R$ on S^3 reduces to translations in the neighborhood of $x_4 = R$ in the limit of infinite radius. Hence Eq. (50) describes fermions in $R \times S^3$ in the same way as the flat space equation.

Like the scalar equation, (50) can be slightly generalized, as we will see.

The requirement that the transformation group of a space-time should have an invariant subalgebra of four generators that have an unitary representation and reduce to translations of Minkowski space in the limit of vanishing curvature is very stringent. It rules out de Sitter and anti-de Sitter space, S^4 and $R \times H^{1,3}$, where $H^{1,3}$ is the hyperboloid $x_4^2 - x_1^2 - x_2^2 - x_3^2 = 1$. The argument above is therefore not applicable to these. However, de Sitter space has the topology of $R \times S^3$, and we take advantage of this in Sec. V A.

Spinorial functions on the hypersphere are constructed by taking direct products of scalar functions with the four spinors

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix}\quad\begin{pmatrix}0\\1\\0\\0\end{pmatrix}\quad\begin{pmatrix}0\\0\\1\\0\end{pmatrix}\quad\begin{pmatrix}0\\0\\0\\1\end{pmatrix}, \tag{52}$$

which will be shown to form the $\mathbf{J}^2 = \frac{3}{4}$ representation.

The SU(2)×SU(2) structure still holds, except that one of $\mathbf{L}_1$ and $\mathbf{L}_2$ remains unchanged, and the other becomes a sum

$$\mathbf{L}' = \mathbf{L} + \mathbf{S}\,. \tag{53}$$

Here $\mathbf{L}$ is the differential operator [(4) and (5)] and acts on scalar functions, whereas $\mathbf{S}$ is a $2d$-dimensional matrix that acts on spinors and satisfies

$$\mathbf{S}^2 = \tfrac{3}{4}\mathbf{I}_{2d}\,. \tag{54}$$

The first two spinors in (52) have the eigenvalue of γ^5 equal to $+1$ and describe right-handed Weyl particles given by the two-spinor Ψ_α. The most general Weyl Lagrangian is

$$\mathcal{L}_W = \bar{\Psi}(-i\bar{\sigma}^n \partial_n + d\bar{\sigma}^0)\Psi\,, \tag{55}$$

where d is real.

The operators that commute with $-i\bar{\sigma}^n \partial_n + d\bar{\sigma}^0$ are $i\partial_0, L_{1i} + \sigma^i/2 = i/2\,\partial_i + \sigma^i/2$ and L_{2i}. Under the rotations L_{2i}, Ψ does not change, $\Psi'(x') = \Psi(x)$. Under the rotations L_{1i}, Ψ must be multiplied by $\exp(i\epsilon \cdot \sigma)$. If we had chosen the frame $\hat{e}^{(2)}$ the commuting operators would have been L_{1i} and $L_{2i} + \sigma^i/2$. In either case, the angular momentum operator is

$$J_i = L_{1i} + L_{2i} + \sigma^i/2\,. \tag{56}$$

So the spin angular momentum is given by

$$S_i = \sigma^i/2\,, \quad \mathbf{S}^2 = \tfrac{3}{4}\mathbf{I}_2\,. \tag{57}$$

The number of Weyl modes for a given s is $2(s+1)^2$, as is also true for a Majorana particle. For a four-component Dirac particle, the dimension is $4(s+1)^2$.

For a Majorana particle with a mass term, the Lagrangian is

$$\mathcal{L}_M = \bar{\Psi}(-i\bar{\sigma}^n \partial_n + d\bar{\sigma}^0)\Psi + m(\Psi^\alpha\Psi_\alpha + \bar{\Psi}_{\dot{\alpha}}\bar{\Psi}^{\dot{\alpha}})\,. \tag{58}$$

This can be written as a four-spinor with a wave function of the special form

$$\Psi_M = \begin{pmatrix}\Psi_\alpha\\ \bar{\Psi}^{\dot{\alpha}}\end{pmatrix}. \tag{59}$$

In Minkowski space, the spatial dependence of Ψ_M factorizes off as a complex number $\exp(-ip^m x_m)$, and the remaining spinor can be made real by a suitable choice of gamma matrices. This cannot be done in $R \times S^3$, but Majorana spinors can be written as in (59).

The Dirac Lagrangian can now be written down, since such a fermion is a combination of two Weyl particles coupled by a mass term:

$$\begin{aligned}\mathscr{L}_{\mathrm{D}} &= \overline{\Psi}_1(-i\bar{\sigma}^n\partial_n + d_1\bar{\sigma}^0)\Psi_1 + \overline{\Psi}_2(-i\bar{\sigma}^n\partial_n + d_2\bar{\sigma}^0)\Psi_2 \\ &\quad + m(\Psi_1{}^{\alpha}\Psi_{2\alpha} + \overline{\Psi}_{1\dot{\alpha}}\overline{\Psi}^{2\dot{\alpha}}) \\ &= \overline{\Psi}(i\gamma^m\partial_m - m + e_1\gamma^5\gamma^0 + e_2\gamma^0)\Psi\,, \end{aligned} \tag{60}$$

where

$$\Psi = \begin{pmatrix}\Psi_{1\alpha}\\ \overline{\Psi}_2{}^{\dot{\alpha}}\end{pmatrix}, \qquad e_1 = (d_1+d_2)/2\,, \quad e_2 = (d_2-d_1)/2\,. \tag{61}$$

Once again e_1 and e_2 are arbitrary, though a theory with some additional symmetry may constrain them. For example, it is found[1] that with a nonzero mass term, the Lagrangian is supersymmetric if and only if $e_1 = -2$.

The symmetries of (60) are generated by $i\partial_0$, $L_{1i} + \frac{1}{2}\epsilon_{ijk}\sigma^{jk}$, and L_{2i}. Hence

$$J_i = \epsilon_{ijk}(-ix_j\partial_k + \sigma^{jk}/2)\,. \tag{62}$$

The spin angular momentum is

$$\mathbf{S}_i = \tfrac{1}{2}\epsilon_{ijk}\sigma^{jk}\,, \quad \mathbf{S}^2 = \tfrac{3}{4}\mathbf{I}_4\,. \tag{63}$$

A. Wave function, dispersion relation, and propagator

Consider first a Weyl fermion. The field Ψ_{w} can be expanded as

$$\Psi = \sum_n [b_n\Psi_+(n)e^{i\omega_+(n)t} + d_n{}^+\Psi_-(n)e^{-i\omega_-(n)t}]\,, \tag{64}$$

where ω_+ and ω_- are non-negative and n denotes the different momentum modes. The Hamiltonian

$$\begin{aligned} H &= \int d\Omega\, :\overline{\Psi}(i\bar{\sigma}^i\partial_i - d\bar{\sigma}^0)\Psi: \\ &= \sum_n \omega_+(n)b_n{}^+ b_n + \omega_-(n)d_n{}^+ d_n\,, \end{aligned} \tag{65}$$

provided the functions Ψ are normalized,

$$\int d\Omega\, \Psi_\sigma^+(n)\Psi_{\sigma'}(n') = \delta_{nn'}\delta_{\sigma\sigma'}\,, \tag{66}$$

and canonical quantization is invoked.

$$\{b_n{}^+, b_{n'}\} = \{d_n{}^+, d_{n'}\} = \delta_{nn'}\,. \tag{67}$$

To find the wave function, it is convenient to consider eigenvectors of $i\sigma^i\partial_i$. These are of two types. The $\chi_+(s, s_1, s_2)$ have non-negative eigenvalues and the $\chi_-(s, s_1, s_2)$ have negative eigenvalues. For a given s, there are $n_+(s) = (s+2)(s+1)$ of the χ_+ with eigenvalue s and $n_-(s) = s(s+1)$ of the χ_- with eigenvalue $-s-2$. The total number is $2(s+1)^2$. The $\chi(s)$ only contain $(s/2, s/2)$ scalar functions. Normalized expressions for them are

$$\begin{aligned} \chi_\pm(s, s_1, s_2) &= \begin{pmatrix}\cos\theta_\pm(s, s_1)S(s, s_1, s_2)\\ \sin\theta_\pm(s, s_1)S(s, s_1+1, s_2)\end{pmatrix}, \\ \cos\theta_\pm(s, s_1) &= \left[\frac{s+1\pm(2s_1+1)}{2(s+1)}\right]^{1/2}, \\ \sin\theta_\pm(s, s_1) &= \pm\left[\frac{s+1\mp(2s_1+1)}{2(s+1)}\right]^{1/2}. \end{aligned} \tag{68}$$

In each case, $-s/2 \leqslant s_2 \leqslant s/2$, but for χ_+, $-s/2-1 \leqslant s_1 \leqslant s/2$ and for χ_-, $-s/2 \leqslant s_1 \leqslant s/2-1$.

A few examples are presented:

$$\begin{aligned} i\sigma^i\partial_i = 0&: \chi_+(0) - \begin{pmatrix}1\\0\end{pmatrix}\ \begin{pmatrix}0\\1\end{pmatrix}, \\ i\sigma^i\partial_i = 1&: \chi_+(1) - \begin{pmatrix}2^{1/2}u\\0\end{pmatrix}\ \begin{pmatrix}2^{1/2}v\\0\end{pmatrix}\ \begin{pmatrix}0\\2^{1/2}u^*\end{pmatrix} \\ &\quad \begin{pmatrix}0\\2^{1/2}v^*\end{pmatrix}\ \begin{pmatrix}u^*\\v\end{pmatrix}\ \begin{pmatrix}v^*\\-u\end{pmatrix}, \\ i\sigma^i\partial_i = -3&: \chi_-(1) - \begin{pmatrix}u^*\\-v\end{pmatrix}\ \begin{pmatrix}v^*\\u\end{pmatrix}. \end{aligned} \tag{69}$$

The spinors have the completeness relation

$$\sum_{s_1,s_2}\chi_\pm(s, s_1, s_2)\chi_\pm^+(s, s_1, s_2) = \tfrac{1}{2}n_\pm(s)\mathbf{I}_2\,. \tag{70}$$

The energy of $\chi(s)$ is

$$E_\pm(s) = \pm(s+1) - d - 1\,. \tag{71}$$

In terms of these spinors, the Feynman propagator is much simpler than in flat space.

$$S_{\mathrm{F}}(E, s, \pm) = i/[-E - d - 1 \pm (s+1)]\,. \tag{72}$$

There are no Pauli matrices to worry about.

The reason for the notation $E_\pm$ is that $E_+ \geqslant 0$ and $E_- \leqslant 0$, if $-3 \leqslant d \leqslant 0$. It is convenient to make this assumption.

Now look at a Dirac fermion. The Hamiltonian

$$\begin{aligned} H &= \int d\Omega\, :\overline{\Psi}(-i\gamma^i\partial_i - e_1\gamma^5\gamma^0 - e_2\gamma^0)\Psi: \\ &= \sum_n \omega_+(n)b_n{}^+ b_n + \omega_-(n)d_n{}^+ d_n\,, \end{aligned} \tag{73}$$

provided the Ψ are normalized according to (66) and

$$\overline{\Psi} = -\Psi^+\gamma^0\,. \tag{74}$$

We define wave functions $\Psi_+(s, s_1, s_2, \sigma)$ and $\Psi_-(s, s_1, s_2, \sigma)$ with energies $E_+(s,\sigma)$ and $E_-(s,\sigma)$. Here $\sigma = \pm 1$, depending on whether $\Psi(s)$ is composed of $\chi_+(s)$ or $\chi_-(s)$. The dimensionalities of $\Psi_+(s,\sigma)$ and $\Psi_-(s,\sigma)$ are therefore both $2(s+1)^2$. Explicitly,

$$\Psi_\pm(s, s_1, s_2, \sigma) = \begin{pmatrix}\cos\theta_\pm(s,\sigma)\,\chi_\sigma(s, s_1, s_2)\\ \sin\theta_\pm(s,\sigma)\,\chi_\sigma(s, s_1, s_2)\end{pmatrix} \tag{75}$$

and

$$E_\pm(s,\sigma) = e_2 \pm [(\lambda_\sigma(s) - e_1)^2 + m^2]^{1/2}\,, \tag{76}$$

where $\lambda_\sigma(s)$ is the $i\sigma^i\partial_i$ eigenvalue of $\chi_\sigma(s)$. That is, $\lambda_\pm(s) = \pm(s+1) - 1$. The $\theta_\pm(s,\sigma)$ are given by

$$\begin{aligned} \cos\theta_\pm(s,\sigma) &= \frac{\pm\nu_\sigma(s) + \lambda_\sigma(s) - e_1}{[(\pm\nu_\sigma(s) + \lambda_\sigma(s) - e_1)^2 + m^2]^{1/2}}\,, \\ \sin\theta_\pm(s,\sigma) &= \frac{m}{[(\pm\nu_\sigma(s) + \lambda_\sigma(s) - e_1)^2 + m^2]^{1/2}}\,, \\ \nu_\sigma(s) &= [(\lambda_\sigma(s) - e_1)^2 + m^2]^{1/2}\,. \end{aligned} \tag{77}$$

The completeness relation for these spinors is more complicated but has a block form

$$\sum_{s_1,s_2,\sigma} \Psi_{\pm}(s,s_1,s_2,\sigma)\Psi_{\pm}^{+}(s,s_1,s_2,\sigma)$$

$$= \frac{s+1}{2}\begin{pmatrix} [(s+2)\cos^2\theta_{\pm}(s,+1) & [(s+2)\cos\theta_{\pm}(s,+1)\sin\theta_{\pm}(s,+1) \\ \quad +s\cos^2\theta_{\pm}(s,-1)]\mathbf{I}_2 & \quad +s\cos\theta_{\pm}(s,-1)\sin\theta_{\pm}(s,-1)]\mathbf{I}_2 \\ [(s+2)\cos\theta_{\pm}(s,+1)\sin\theta_{\pm}(s,+1) & [(s+2)\sin^2\theta_{\pm}(s,+1) \\ \quad +s\cos\theta_{\pm}(s,-1)\sin\theta_{\pm}(s,-1)]\mathbf{I}_2 & \quad +s\sin^2\theta_{\pm}(s,-1)]\mathbf{I}_2 \end{pmatrix}. \tag{78}$$

Just as in (72), the propagator has a simple form. From (75),

$$i\gamma^i\partial_i\Psi(s,\sigma) = -\lambda_\sigma(s)\gamma^5\gamma^0\Psi(s,\sigma), \tag{79}$$

and we find

$$S_F(E,s,\sigma) = i\frac{e_2 - E + (e_1 - \lambda_\sigma(s))\gamma^5}{(e_2-E)^2 - (e_1-\lambda_\sigma(s))^2}\cdot\gamma^0. \tag{80}$$

The energy, wave function, and propagator simplify if $e_1 = -1$ and $e_2 = 0$. Then the square of the Dirac operator is exactly the Klein–Gordon operator, and

$$E_{\pm}(s) = \pm[(s+1)^2 + m^2]^{1/2}, \tag{81}$$

$$\theta_{\pm}(s,\sigma) = \tan^{-1}\{m/[E_{\pm} + \sigma(s+1)]\}, \tag{82}$$

$$S_F(E,s,\sigma) = -i\{[E+\sigma(s+1)\gamma^5]/[E^2-(s+1)^2]\}\cdot\gamma^0. \tag{83}$$

B. Interactions

The interaction terms for fermions look exactly as in flat space,

$$\mathcal{L}_{\Psi\Phi} = g_s\Phi_s\bar{\Psi}\Psi + g_p\Phi_p\bar{\Psi}\gamma^5\Psi, \tag{84}$$

where Φ_s (Φ_p) is a scalar (pseudoscalar).

For vectors,

$$\mathcal{L}_{\Psi V} = -gV_m\bar{\Psi}\gamma^m\Psi, \tag{85}$$

$$\mathcal{L}_{\Phi V} = gV_m(\Phi^* i\partial^m\Phi - i\partial^m\Phi^*\Phi) - g^2V_mV^m\Phi^*\Phi - 2egV_0\Phi^*\Phi. \tag{86}$$

The last term is needed to make the Lagrangian (23) gauge invariant.

Without a vector mass and the nonminimal terms in (34), the theory is gauge invariant,

$$\Psi \to \exp(-igS)\Psi, \quad \Phi \to \exp(-igS)\Phi, \quad V_m \to V_m + \partial_m S, \tag{87}$$

where S is a scalar harmonic function with an arbitrary dependence on time.

Gauge invariance implies the existence of conserved currents. These are

$$J_F^m = \bar{\Psi}\gamma^m\Psi, \tag{88}$$

$$J_B^m = i(\Phi^*\partial^m\Phi - \partial^m\Phi^*\Phi) - 2gV^m\Phi^*\Phi + 2eg^{m0}\Phi^*\Phi. \tag{89}$$

They are conserved by the Euler–Lagrange equations of motion. One therefore expects Ward identities to hold as usual.

C. Regularization

For the purposes of numerical analysis, one must define a high-momentum cutoff that is parity and rotation invariant. Choose a large number s and keep all functions up to $S(s)$, $\Psi(s)$, and $V^{\pm}(s)$. This is a system with a finite number, about s^3, of degrees of freedom that can approximate the behavior of any flat-space theory, provided all mass parameters m_i in the theory satisfy

$$R^{-1} \ll m_i \ll sR^{-1}. \tag{90}$$

Perturbative renormalization can be done in a similar way. Dimensional regularization is inapplicable here since we want the special properties of $R\times S^3$, and Pauli–Villars regularization does not work in gauge theories. But since we have broken Lorentz symmetry to begin with, an energy cutoff can be imposed to make divergent integrals finite. There is no need for similarly cutting off the spatial momentum and, in fact, it might be convenient not to do so, in order to use the completeness relations of the spherical harmonics.

III. MATRIX ELEMENTS

In this section, we derive formulas for some typical matrix elements. They all reduce to spherical integrals of the product of three scalar functions.[2] This can be simplified to the product of two 3-j symbols[8] with a reduced matrix element. Define

$$\mathbf{I}(rst; r_1s_1t_1; r_2s_2t_2) = \int d\Omega\, S(r,r_1,r_2)S(s,s_1,s_2)S(t,t_1,t_2) = R(r,s,t)\begin{pmatrix} \frac{r}{2} & \frac{s}{2} & \frac{t}{2} \\ r_1 & s_1 & t_1 \end{pmatrix}\begin{pmatrix} \frac{r}{2} & \frac{s}{2} & \frac{t}{2} \\ r_2 & s_2 & t_2 \end{pmatrix}, \tag{91}$$

with

$$R(r,s,t) = (-1)^{\alpha}[(r+1)(s+1)(t+1)]^{1/2} \tag{92}$$

and $\alpha = (r+s+t)/2$ must be an integer for the above integral to be nonzero. Henceforth, whenever possible, we will simply write $\mathbf{I}(r_1s_1t_1)$, the other arguments being understood.

The matrix elements we calculate are

$$\mathbf{I}^{(1)} = \int d\Omega\,\bar{\Psi}(r\,r_1\,r_2)\Psi(s\,s_1\,s_2)\Phi(t\,t_1\,t_2),$$

$$\mathbf{I}^{(2)} = \int d\Omega\,\Psi^{+}(r)\Psi(s)\Phi(t),$$

$$\mathbf{I}^{(3)} = \int d\Omega\,\bar{\Psi}(r)\gamma^i\Psi(s)\partial_i\Phi(t), \tag{93}$$

$$\mathbf{I}^{(4)} = \int d\Omega\,\bar{\Psi}(r)\gamma^i\Psi(s)V_i^{\pm}(t).$$

The $\mathbf{I}^{(m)}$ will also have arguments $(rst,r_1s_1t_1;r_2s_2t_2;\rho\sigma\cdots)$ but we only write $\mathbf{I}^{(m)}(r_1s_1t_1;\rho\sigma\cdots)$ in the following.

Consider Weyl fermions first. Then, by Eqs. (42) and (68),

$$\begin{aligned}\mathbf{I}_w^{(2)}(r_1 s_1 t_1;\rho\sigma) &= -\mathbf{I}_w^{(1)}(r_1 s_1 t_1;\rho\sigma)\\ &= \int d\Omega\, \chi_\rho^+(r)\chi_\sigma(s)\Phi(t)\\ &= (-1)^{r_1+r_2}[\cos\theta_\rho(r,r_1)\cos\theta_\sigma(s,s_1)\mathbf{I}(-r_1 s_1 t_1)\\ &\quad + \sin\theta_\rho(r,r_1)\sin\theta_\sigma(s,s_1)\mathbf{I}(-r_1-1\, s_1+1\, t_1)]\,. \end{aligned} \tag{94}$$

To calculate $\mathbf{I}^{(3)}$, we need (11), $\partial_i S(t,t_1,t_2) = -2iL_{1i}S(t,t_1,t_2)$. Define

$$A(t,t_1) = [(t+2t_1+2)(t-2t_1)]^{1/2}\,. \tag{95}$$

Then

$$\sigma^j\partial_i S(t_1) = \begin{pmatrix} -2it_1 S(t_1) & -iA(t,-t_1)\,S(t_1-1)\\ -iA(t,t_1)S(t_1+1) & 2it_1 S(t_1)\end{pmatrix} \tag{96}$$

and

$$\begin{aligned}\mathbf{I}_w^{(3)}(r_1 s_1 t_1;\rho\sigma) &= \int d\Omega\, \chi_\rho^+(r)\sigma^j\chi_\sigma(s)\partial_i\Phi(t)\\ &= -(-1)^{r_1+r_2}2it_1[\cos\theta_\rho(r,r_1)\cos\theta_\sigma(s,s_1)\mathbf{I}(-r_1 s_1 t_1) - \sin\theta_\rho(r,r_1)\sin\theta_\sigma(s,s_1)\mathbf{I}(-r_1-1\, s_1+1\, t_1)]\\ &\quad -(-1)^{r_1+r_2}iA(t,-t_1)\cos\theta_\rho(r,r_1)\sin\theta_\sigma(s,s_1)\mathbf{I}(-r_1 s_1+1\, t_1-1)\\ &\quad -(-1)^{r_1+r_2}iA(t,t_1)\sin\theta_\rho(r,r_1)\cos\theta_\sigma(s,s_1)\mathbf{I}(-r_1 s_1 t_1+1)\,, \end{aligned} \tag{97}$$

$$\begin{aligned}\mathbf{I}^{w(4)}(r_1 s_1 t_1;\rho\sigma\tau) &= \int d\Omega\, \chi_\rho^+(r)\sigma^j\chi(s)V^\tau{}_i(t)\\ &= (-1)^{r_1+r_2}\cos\alpha(t,t_1)[\cos\theta_\rho(r,r_1)\cos\theta_\sigma(s,s_1)\mathbf{I}(rst-\tau;\,-r_1 s_1 t_1)\\ &\quad +\sin\theta_\rho(r,r_1)\sin\theta_\sigma(s,s_1)\mathbf{I}(rst-\tau;\,-r_1-1\, s_1+1\, t_1)]\\ &\quad +(-1)^{r_1+r_2}2^{1/2}\sin\theta_\rho(r,r_1)\cos\theta_\sigma(s,s_1)\sin\alpha(t,t_1)\cos\beta^\tau(t,t_1)\mathbf{I}(rst-\tau;\,-r_1-1\, s_1 t_1+1)\\ &\quad +(-1)^{r_1+r_2}2^{1/2}\cos\theta_\rho(r,r_1)\sin\theta_\sigma(s,s_1)\sin\alpha(t,t_1)\sin\beta^\tau(t,t_1)\mathbf{I}(rst-\tau;\,-r_1 s_1+1\, t_1-1)\,. \end{aligned} \tag{98}$$

Now $\mathbf{I}^{(3)}$ and $\mathbf{I}^{(4)}$ can be further simplified in that, instead of all four of the $\mathbf{I}(-r_1 s_1 t_1)$, it is enough to know only two, for example, $\mathbf{I}(-r_1 s_1 t_1)$ and $\mathbf{I}(-r_1-1\, s_1+1\, t_1)$. This is because of the relations

$$\begin{aligned}A(t,t_1)\mathbf{I}(-r_1-1\, s_1 t_1+1) &= -A(r,-r_1-1)\mathbf{I}(-r_1 s_1 t_1) - A(s,s_1)\mathbf{I}(-r_1-1\, s_1+1\, t_1)\,,\\ A(t,-t_1)\mathbf{I}(-r_1 s_1+1\, t_1-1) &= -A(s,-s_1-1)\mathbf{I}(-r_1 s_1 t_1) - A(r,r_1)\mathbf{I}(-r_1-1\, s_1+1\, t_1)\,,\end{aligned} \tag{99}$$

which follow from a recursion formula for 3-j symbols,

$$A(r,r_1)\mathbf{I}(r_1+1\, s_1 t_1) + A(s,s_1)\mathbf{I}(r_1 s_1+1\, t_1) + A(t,t_1)\mathbf{I}(r_1 s_1 t_1+1) = 0\,. \tag{100}$$

For Dirac fermions, the integrals (93) can be expressed in terms of the $\mathbf{I}_w^{(m)}$:

$$\begin{aligned}\mathbf{I}_D^{(1)}(r_1 s_1 t_1;\rho\sigma;\rho'\sigma') &= \int d\Omega\, \bar\Psi_\rho(r,r_1,r_2,\rho')\Psi_\sigma(s,s_1,s_2,\sigma')\Phi(t,t_1,t_2)\\ &= \sin[\theta_\rho(r,r_1,\rho') + \theta_\sigma(s,s_1,\sigma')]\\ &\quad\times\mathbf{I}_w^{(2)}(r_1 s_1 t_1;\rho'\sigma')\,,\\ \mathbf{I}_D^{(2)}(r_1 s_1 t_1;\rho\sigma;\rho'\sigma') &= -\cos[\theta_\rho(r,r_1,\rho') - \theta_\sigma(s,s_1,\sigma')]\\ &\quad\times\mathbf{I}_w^{(2)}(r_1,s_1,t_1;\rho'\sigma')\,,\\ \mathbf{I}_D^{(3)}(r_1,s_1,t_1;\rho\sigma;\rho'\sigma') &= -\cos[\theta_\rho(r,r_1,\rho') + \theta_\sigma(s,s_1,\sigma')]\\ &\quad\times\mathbf{I}_w^{(3)}(r_1,s_1,t_1;\rho'\sigma')\,,\\ \mathbf{I}_D^{(4)}(r_1,s_1,t_1;\rho\sigma;\rho'\sigma') &= -\cos[\theta_\rho(r,r_1,\rho') + \theta_\sigma(s,s_1,\sigma')]\\ &\quad\times\mathbf{I}_w^{(4)}(r_1,s_1,t_1;\rho'\sigma'\tau)\,.\end{aligned} \tag{101}$$

IV. DISCRETE SYMMETRIES

In flat space, we say a Lagrangian has a discrete symmetry if it remains invariant under the transformation. In $R\times S^3$, things are not as simple. For example, under parity

$$\partial_i^{(1)} \to -\partial_i^{(2)}\,, \tag{102}$$

so the fermionic Lagrangian *must* change in form. We require that physics should remain the same and that the quantum fields and the creation and destruction operators should transform in the same way as in flat space. The Lagrangian must then change if necessary. In other words, one is looking for symmetries of the theory, not the Lagrangian directly.

A. Parity

Under parity, the coordinates transform as

$$t, x_4, x_i \to t, x_4, -x_i\,. \tag{103}$$

For a scalar, we may define

$$\Phi(t,x_4,x_i) \to \Phi(t,x_4,-x_i) \tag{104}$$

so the Lagrangian (23) does not change. This is also true of a vector particle,

$$V_0(t, x_4, x_i) \to V_0(t, x_4, -x_i)\,,$$
$$V_i(t, x_4, x_i) \to -V_i(t, x_4, -x_i)\,. \tag{105}$$

For a Dirac particle,

$$\Psi(t, x_4, x_i) \to \gamma^0\Psi(t, x_4, -x_i) \tag{106}$$

and the Lagrangian transforms as

$$e_1 \to -e_1\,,$$
$$e_2 \to e_2\,,$$
$$\bar{\Psi}(x)(i\gamma^0\partial_0 + i\gamma^i\partial_i^{(1)} - m + e_1\gamma^5\gamma^0 + e_2\gamma^0)\Psi(x)$$
$$\to\bar{\Psi}(x')(i\gamma^0\partial_0 + i\gamma^i\partial_i^{(2)} - m - e_1\gamma^5\gamma^0 + e_2\gamma^0)\Psi(x')\,. \tag{107}$$

It is important to understand that the e_1 term does not violate parity, but simply that it appears with the opposite sign in the parity reversed coordinate system. The observer in the new system finds particles with the same energy but parity reversed wave function. The e_1 and e_2 terms have a R^{-1} hidden in them. They are due to the curvature of the space and the twist of the coordinate frame. So it is not surprising that they change sign under a discrete symmetry. A term like $m'\gamma^5$, where m' has the dimension of mass but is unrelated to R^{-1} and survives in the $R\to\infty$ limit, would certainly violate parity.

A similar situation arises with charge conjugation, this time even with the scalar Lagrangian.

B. Charge conjugation

Consider the charged scalar field expanded in terms of creation and destruction operators.

$$\Phi = \sum_n [a(n)S_+(n)e^{i\omega_+(n)t} + c^+(n)S_-(n)e^{-i\omega_-(n)t}]\,. \tag{108}$$

The $S_\pm(n)$ are scalar functions and a and c are the particle and antiparticle destruction operators. Under charge conjugation, we require

$$a \leftrightarrow c\,, \quad \Phi \leftrightarrow \Phi^*\,. \tag{109}$$

This is a symmetry if in (23),

$$e \to -e\,. \tag{110}$$

For the Dirac particle

$$\Psi = \sum_n [b(n)\Psi_+(n)e^{i\omega_+(n)t} + d^+(n)\Psi_-(n)e^{-i\omega_-(n)t}]\,,$$

we want

$$\Psi \to i\gamma^2\Psi^*\,. \tag{111}$$

This implies

$$e_1 \to e_1\,, \quad e_2 \to -e_2\,. \tag{112}$$

This can also be understood as follows. Charge conjugation is equivalent to interchanging the two Weyl particles in Ψ,

$$\Psi = \begin{pmatrix}\Psi_{1\alpha}\\ \bar{\Psi}_2{}^{\dot\alpha}\end{pmatrix} \to \begin{pmatrix}\Psi_{2\alpha}\\ \bar{\Psi}_1{}^{\dot\alpha}\end{pmatrix}, \quad d_1 \leftrightarrow d_2\,. \tag{113}$$

Hence

$$e_1 \to e_1 \quad \text{but} \quad e_2 \to -e_2\,.$$

For the vector particle,

$$V_m \to -V_m\,. \tag{114}$$

So in (34),

$$f_3 \to -f_3\,, \tag{115}$$

which is consistent with (110), assuming the f_3 term only appears by spontaneous symmetry breaking.

C. Time reversal

This means

$$t \to -t\,. \tag{116}$$

For the scalar field,

$$\Phi(t, x) \to \Phi(-t, x)\,,$$
$$a,c \to a,c\,, \tag{117}$$
$$e \to e\,.$$

Thus the Lagrangian remains unchanged. The Dirac and Weyl Lagrangians do not change either.

$$d \to d\,, \quad e_1,e_2 \to e_1,e_2\,. \tag{118}$$

The wave functions transform as

$$\Psi_D(t, x) \to i\gamma^1\gamma^3\Psi_D(-t, x)\,,$$
$$\Psi_W(t, x) \to \sigma^2\Psi_W(-t, x)\,. \tag{119}$$

For the vector particle,

$$V_0(t, x) \to V_0(-t, x)\,, \quad V_i(t, x) \to -V_i(-t, x)\,. \tag{120}$$

So

$$f_1, f_3 \to f_1, f_3\,, \quad f_2 \to -f_2\,. \tag{121}$$

D. CP

All the nonminimal terms in the scalar and Dirac Lagrangians change sign:

$$e \to -e\,, \quad e_1,e_2 \to -e_1, -e_2\,. \tag{122}$$

The latter also follows from (60) and (61), since the Weyl theory is CP invariant:

$$\Psi(t, x_4, x_i) \to i\sigma^2\Psi^*(t, x_4, -x_i)$$

but

$$d \to -d\,. \tag{123}$$

For the vector particle

$$f_1 \to f_1\,, \quad f_2, f_3 \to -f_2, -f_3\,. \tag{124}$$

The above rules can be summarized in a simple way if we set the f_1 and f_2 terms equal to zero in the vector Lagrangian, as in fact they must be, if one demands gauge invariance. Then the rule is that under CP, all nonminimal terms (e,d,e_1,e_2,f_3) change sign and under T, none of them change sign.

V. FERMIONS AGAIN

We now rederive the Dirac equation using the tetrad formalism. This will prove the invariance of the equation

under any coordinate transformation.

The sphere S^3 is the manifold of SU(2), as shown by the parametrization

$$U = x_4 + i\mathbf{x}\cdot\boldsymbol{\sigma}\,. \tag{125}$$

The SU(2)×SU(2) transformations act on U from the left and right

$$U' = U_1 U U_2^{+}\,. \tag{126}$$

Hence SO(4) is isomorphic to SU(2)×SU(2) mod Z_2. Further, J corresponds to the diagonal subgroup $U_1 = U_2$, which is SO(3). Since S^3 is a group manifold, one can take the Maurer–Cartan forms as a basis for a triad.[3] The coordinate frame we have chosen throughout is the left-invariant Maurer–Cartan form:

$$\Omega = g^{-1}dg = i\sigma^i e^i\,,$$

$$e_1 = -\sin(\alpha+\beta)\,d\theta + \sin\theta\cos\theta\cos(\alpha+\beta)[d\alpha - d\beta]\,,$$

$$e^2 = \cos(\alpha+\beta)\,d\theta + \sin\theta\cos\theta\sin(\alpha+\beta)[d\alpha - d\beta]\,,$$

$$e^3 = -\cos^2\theta\,d\alpha - \sin^2\theta\,d\beta\,,$$

$$e^0 = dt\,. \tag{127}$$

Then the spin-connection one-form has the simple structure

$$w^{0i} = 0\,,\quad w^{ij} = -\epsilon_{ijk}e^k\,. \tag{128}$$

Hence

$$\begin{aligned} D_0 &= \partial_0\,,\\ D_i &= e_i{}^{\mu}\left(\partial_\mu + \frac{i}{2}\omega_{\mu jk}\sigma^{jk}\right)\\ &= \partial_i - \frac{i}{2}e_i{}^{\mu}e_{jkl}e^l{}_{\mu}\sigma^{jk}\\ &= \partial_i + \tfrac{1}{8}\epsilon_{ijk}[\,\gamma^j,\gamma^k\,]\,. \end{aligned} \tag{129}$$

The Dirac operator is

$$i\gamma^m D_m = i\gamma^m\partial_m - \tfrac{3}{2}\gamma^5\gamma^0 \tag{130}$$

so the minimal Dirac Lagrangian has

$$e_1 = -\tfrac{3}{2}\,,\quad e_2 = 0\,. \tag{131}$$

For the Weyl equation,

$$\begin{aligned} \sigma^{ij} &= -\tfrac{1}{2}\epsilon^{ijk}\sigma^k\bar{\sigma}^0\,,\\ D_i &= \partial_i + \frac{i}{2}\sigma^i\bar{\sigma}^0\,, \end{aligned} \tag{132}$$

so one gets

$$-i\bar{\sigma}^m D_m = i\bar{\sigma}^m\partial_m - \tfrac{3}{2}\bar{\sigma}^0\,. \tag{133}$$

We use the two-component notation henceforth. The covariant derivatives are

$$\begin{aligned} D_\theta &= \partial_\theta + \frac{i}{2}\sin(\alpha+\beta)\sigma^1 - \frac{i}{2}\cos(\alpha+\beta)\sigma^2\,,\\ D_\alpha &= \partial_\alpha - \frac{i}{2}\sin\theta\cos\theta\,[\cos(\alpha+\beta)\sigma^1\\ &\quad + \sin(\alpha+\beta)\sigma^2\,] + \frac{i}{2}\cos^2\theta\,\sigma^3\,,\\ D_\beta &= \partial_\beta + \frac{i}{2}\sin\theta\cos\theta\,[\cos(\alpha+\beta)\sigma^1\\ &\quad + \sin(\alpha+\beta)\sigma^2\,] + \frac{i}{2}\sin^2\theta\,\sigma^3\,. \end{aligned} \tag{134}$$

For application to supersymmetry, we must consider the modified covariant derivative, $\mu = \theta,\alpha,\beta$,

$$\tilde{D}_\mu = D_\mu \pm \frac{i}{2}\sigma^j e_{i\mu}\,. \tag{135}$$

The extra term is proportional to the Ricci scalar $R_\mu{}^\mu$ in general.

In our case, $R_\mu{}^\mu = -6$. Global supersymmetry can only be defined if there exist two Killing spinors ξ satisfying

$$\tilde{D}_\mu\xi = 0\,. \tag{136}$$

The modified derivatives satisfy

$$[\tilde{D}_\mu, \tilde{D}_\nu] = 0\,, \tag{137}$$

but this does not guarantee the existence of the spinors. In fact, one can solve (135) and (136) with either sign. With the plus sign,

$$\tilde{D}_\mu = \partial_\mu\,,\quad \xi = \begin{pmatrix}1\\0\end{pmatrix} \text{ and } \begin{pmatrix}0\\1\end{pmatrix}. \tag{138}$$

With the minus sign,

$$\tilde{D}_\mu = 2D_\mu - \partial_\mu\,,\quad \xi = \begin{pmatrix}u^*\\-v\end{pmatrix} \text{ and } \begin{pmatrix}v^*\\u\end{pmatrix}. \tag{139}$$

This is another great advantage of our choice of axes as opposed to the usual one, which is

$$e^1 = d\theta\,,\quad e^2 = \cos\theta\,d\alpha\,,\quad e^3 = \sin\theta\,d\beta\,, \tag{140}$$

Then

$$\begin{aligned} D_\theta &= \partial_\theta\,,\\ D_\alpha &= \partial_\alpha + \frac{i}{2}\sin\theta\,\sigma^3\,,\\ D_\beta &= \partial_\beta + \frac{i}{2}\cos\theta\,\sigma^2\,. \end{aligned} \tag{141}$$

Once again, one can define modified derivatives

$$\begin{aligned} \tilde{D}_\theta &= D_\theta \pm \frac{i}{2}\sigma^1\,,\\ \tilde{D}_\alpha &= D_\alpha \pm \frac{i}{2}\cos\theta\,\sigma^2\,,\\ \tilde{D}_\beta &= D_\beta \pm \frac{i}{2}\sin\theta\,\sigma^3\,. \end{aligned} \tag{142}$$

These commute but there is no continuous Killing spinor corresponding to them. In fact, Eq. (136) implies

$$(\partial_\theta^2 + \tfrac{1}{4})\xi = (\partial_\alpha^2 + \tfrac{1}{4})\xi = (\partial_\beta^2 + \tfrac{1}{4})\xi = 0\,, \tag{143}$$

which means ξ has sines and cosines of $\alpha/2$ and $\beta/2$. This is unacceptable in S^3, which is simply connected.

Of course, (140) and (141) also give a Dirac equation with

$$\begin{aligned} D_1 &= \partial_\theta\,,\\ D_2 &= \sec\theta\,\partial_\alpha + \frac{i}{2}\tan\theta\,\sigma^3\,,\\ D_3 &= \csc\theta\,\partial_\beta + \frac{i}{2}\csc\theta\,\sigma^2\,. \end{aligned} \tag{144}$$

This is much more complicated than (60) and difficult to solve.

A. Conformal transformations

Conformal transformations[9] of the metric change the physical manifold, unlike coordinate transformations that merely amount to relabeling. A conformal transformation is defined by a continuous, nonzero, real function $\Omega(t, x)$,

$$g_{\mu\nu}(z) \to \bar{g}_{\mu\nu}(z) = \Omega^2(z)\, g_{\mu\nu}(z) . \tag{145}$$

The scalar Lagrangian is conformally invariant if

$$e = 0 \quad \text{and} \quad m^2 = 1 .$$

The vector Lagrangian is invariant since we are in four dimensions, provided m^2, f_1, f_2, f_3 are all zero.

The fermionic case is more interesting. We used the fact that the manifold is $R \times S^3$ and chose a particularly simple tetrad field. One might therefore think that even the minimal massless equation (130) would not be conformally invariant. This is in fact true, except for two special cases. In the first, Ω depends only on time, and in the second, Ω is a function of space alone which is such that it does not change the sign of the curvature.

More precisely, in the second case, Ω must be such that each sectional curvature remains strictly positive and bounded away from zero everywhere. The sectional curvature at any point is defined[10] as the Gaussian curvature of the surface generated by two vectors at that point. In our manifold, the time direction is flat so there are only three independent sectional curvatures to consider. If all three are positive, the spatial manifold locally looks like a region of S^3, so the previous analysis works. In particular, one can define a continuous tetrad field which satisfies (128). With the above property of the curvature, the manifold is bounded and has the topology of S^3 globally.

If Ω is a function of time alone,

$$ds^2 = R^2(\tau)[-d\tau^2 + d\theta^2 + \cos^2\theta\, d\alpha^2 + \sin^2\theta\, d\beta^2] . \tag{146}$$

Changing variables to

$$t = \int^{\tau} d\tau'\, R(\tau') , \tag{147}$$

we get the Robertson–Walker metric for a space of positive curvature

$$ds^2 = -dt^2 + R^2(t)[d\theta^2 + \cos^2\theta\, d\alpha^2 + \sin^2\theta\, d\beta^2] . \tag{148}$$

The Lagrangian and action are, of course, invariant,

$$\mathcal{L} = \bar{\Psi}\left(\frac{i}{2}\gamma^m \overleftrightarrow{\partial}_m - \frac{3}{2}\gamma^5\gamma^0\right)\Psi , \tag{149}$$

$$S = \int d\tau\, d\Omega(-g)^{1/2}\, \mathcal{L}(\tau,\Omega) ,$$

and the wave function is given by the conformal transformation

$$\Psi(\tau; R = 1) \to [R(\tau)]^{-3/2}\Psi(\tau) . \tag{150}$$

It remains normalized with the volume element $R^3(\tau)\, d\Omega$.

If Ω depends only on time, we may consider massive fermions also, and give d, e_1, and e_2 arbitrary values. The Lagrangian is the same as in Sec. II, except that the derivatives are

$$\partial_m = \frac{1}{R(\tau)}\partial_\tau , \quad \frac{1}{R(\tau)}\partial_i , \tag{151}$$

the four-dimensional volume element in the action is $R^4(\tau)\, d\tau\, d\Omega$, and the wave function has the factor $R^{-3/2}$.

As a special case, consider de Sitter space

$$R = \cosh t = \sec\tau . \tag{152}$$

This describes the manifold

$$-x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1 , \tag{153}$$

with the pseudo-Riemannian metric

$$ds^2 = -dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 . \tag{154}$$

To get back to (148), parametrize as in Sec. I A,

$$\begin{aligned} x_0 &= \sinh t , \\ x_1 + ix_2 &= \cosh t \cos\theta\, e^{i\alpha} , \\ x_3 + ix_4 &= \cosh t \sin\theta\, e^{i\beta} . \end{aligned} \tag{155}$$

This space is of great interest cosmologically. Quantum field theory in it has been studied by several people beginning, as often, with Dirac.[11] Its symmetry group has ten generators, six rotations, and four boosts.

The action (149) describes fermions adequately, but it is not symmetric under the full de Sitter group but only under the subgroup that leaves x_0 invariant. The boosts act as raising and lowering operators between the different SO(4) representations. The wave functions corresponding to a completely symmetric action will be complicated combinations of the ones derived in Sec. II.

From the point of view of this paper, the extra symmetry of de Sitter space is the accidental result of a special time dependence of the radius. Our formalism only takes advantage of the special features of S^3.

B. Weyl spectrum

We point out an interesting feature of the Weyl Lagrangian that is relevant to supersymmetry. As in Sec. II, restrict d to lie in the interval $[0, -3]$. Then the spectrum with d and $-3-d$ are very similar. The energy, for d in the above range, is a non-negative integer E. For any energy, the Lagrangian with d gives as many particles (antiparticles) as the Lagrangian with $-3-d$ gives antiparticles (particles). In terms of E, the precise number is $(E+d+1)(E+d+2)$ particles and $(E-d-1)(E-d-2)$ antiparticles for the first Lagrangian.

This symmetry resembles CP but is actually different from it. The particles and antiparticles that have the same energy, have different wave functions. In fact, the wave functions belong to adjacent representations of SO(4).

The minimal Lagrangian $d = -\frac{3}{2}$ is self-conjugate. For each energy, it has the same number of particles and antiparticles. This is also true for the scalar with $e = 0$ and is a general feature of minimal Lagrangians.

The conjugate cases with d equal to -1 and -2 are specially interesting. If, for each Lagrangian, we count the particles and antiparticles together, then for any energy E the number of modes is $2E^2$. Further, E is a positive integer. This last property is shared by the two conformal bosons, a massless complex scalar with $e = 0$ and a massless gauge particle with all $f_i = 0$. The fermions and the bosons have

the same spectrum, though not the same dimensionality.

It turns out that under supersymmetry, the conformal scalar and vector transform into the $d = -1$ and -2 fermions. Which boson turns into which fermion depends on whether the Killing spinors (138) or (139) are chosen as the anticommuting parameters. For reasons given elsewhere, the correct choice is (139). Then the scalar and the $d = -2$ fermion transform into each other and similarly, the vector and the $d = -1$ fermion.

ACKNOWLEDGMENTS

I thank R. Cutkosky for many stimulating discussions.

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-76ER0306.

[1]D. Sen. "Supersymmetry in $R \times S^3$," CMU preprint, 1985.
[2]R. E. Cutkosky, J. Math. Phys. **25**, 939 (1984).
[3]Y. Hosotani, Phys. Rev. D **30**, 1238 (1984).
[4]L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).
[5]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[6]J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, NJ, 1983).
[7]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[8]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[9]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U.P., Cambridge, 1982), Chap. 3.
[10]J. Milnor, *Morse Theory* (Princeton U.P., Princeton, NJ, 1969), pp. 100–105.
[11]P. A. M. Dirac, Ann. Math. NY **36**, 657 (1935); M. Gutzwiller, Helv. Phys. Acta **29**, 313 (1956); F. Gürsey, *Group Theoretical Concepts and Methods in Elementary Particle Physics* (Gordon and Breach, New York, 1964).

On the classical limit of phase-space formulation of quantum mechanics: Entropy

Lipo Wang
Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803

(Received 30 May 1985; accepted for publication 9 September 1985)

The classical limits of phase-space formulation of quantum mechanics are studied. As a special example, some properties of both quantum mechanical and classical entropies are discussed in detail.

I. INTRODUCTION

It has been common knowledge that quantum mechanics approaches classical mechanics when Planck's constant approaches zero. Rigorous investigations have been carried out during the last decade by various authors.[1-4] So far the methods employed are restricted to the quantum mechanical operator techniques and the questions considered are mainly partition function and ensemble average. The purpose of the present work is to examine the general problem of the classical limit $\hbar \to 0$ by means of the so-called phase-space formalism of quantum mechanics. With the help of the general results, the unsolved problem of the behavior of quantum mechanical entropies at the classical limit is discussed.

The phase-space formulation of quantum mechanics has found many applications, particularly in statistical mechanics and quantum optics. Its basic feature is to provide a framework for the treatment of quantum mechanical problems in terms of classical concepts. Following the appearance of the well-known Wigner distribution function,[5] many other distribution functions have been considered. For instance, the antinormal-ordered (Husimi[6]) and the normal-ordered distribution (P distribution) functions,[7] the anti-standard-ordered (Kirkwood[8]) and the standard-ordered distribution functions.[9] Each of those distribution functions was created for a particular purpose.[10]

Considering the properties of entropies, Wehrl stated, "It is usually claimed that in the limit $\hbar \to 0$, the quantum mechanical expression tends towards the classical one, however, a rigorous proof of this is nowhere found in the literature."[11] In a recent paper,[12] Beretta took the first attempt at this question. But some weak points can be found in Beretta's investigation, as shown in our paper. In fact, both quantum mechanical and classical entropies are singular at the classical limit, however, the difference between them does vanish at this limit.

The paper is organized as follows. In Sec. II we briefly review the concepts of the phase-space formalism of quantum mechanics. Some useful results are derived. In Sec. III the classical limit of quantum mechanical description is considered. We discuss the relation between quantum mechanical and classical entropies in Sec. IV. Conclusions and discussions are presented in Sec. V. Also, in the Appendix, we wish to make some comments on the problems of complete classical phase-space representation of quantum kinematics and spectral expansion in the classical limit $\hbar \to 0$ discussed in Ref. 12.

We are going to restrict our discussion to the case of one degree of freedom so that the Hilbert space is $\mathscr{H} = L^2(R)$ and phase space is $\mathscr{P} = R^2$. But we wish to emphasize that the arguments can be easily extended to the case of many degrees of freedom.

II. THE GENERAL CLASSICAL PHASE-SPACE REPRESENTATION OF QUANTUM MECHANICAL OPERATORS

The mathematical form of the general question about the classical phase-space representation of quantum mechanical operators is stated as follows. Suppose $\hat{A}$ and $\hat{B}$ are two Hermitian operators. Find a pair of mappings, Θ and Θ', say, on phase space, which have the following properties:

$$\Theta(\hat{A}) = a(q, p), \tag{1}$$

$$\Theta'(\hat{B}) = b'(q, p), \tag{2}$$

and

$$\mathrm{Tr}(\hat{A}) = \int\int a(q, p)dq\, dp, \tag{3}$$

$$\mathrm{Tr}(\hat{A}\hat{B}) = \int\int a(q, p)b'(q, p)dq\, dp. \tag{4}$$

This problem was solved satisfactorily by Agarwal and Wolf,[9] but their results were mainly presented in terms of c-number space, annihilation, and creation operators, which is convenient for applications in quantum optics. For the sake of statistical mechanics and discussions in the present paper, we will derive the similar results in terms of phase space, i.e., q and p, language.

Denote the inverse mapping of Θ by Ω, i.e.,

$$\Theta\Omega = \Omega\Theta = 1, \tag{5}$$

$$\hat{A} = \Omega[a(q, p)], \tag{6}$$

then[9]

$$\hat{A} = 2\pi\hbar \int\int a(q, p)\Delta^{(\Omega)}(q - \hat{q}, p - \hat{p})dq\, dp, \tag{7}$$

where the Δ operator is defined by

$$\Delta^{(\Omega)}(q' - \hat{q}, p' - \hat{p}) = \Omega[\delta(q' - q)\delta(p' - p)]. \tag{8}$$

Explicitly it can be written as follows:

$$\Delta^{(\Omega)}(q' - \hat{q}, p' - \hat{p}) = (2\pi\hbar)^{-2} \int\int \Omega(u, v) \times \exp\left[\frac{-i(u(q' - \hat{q}) + v(p' - \hat{p}))}{\hbar}\right] du\, dv. \tag{9}$$

 0022-2488/86/020483-05$02.50 © 1986 American Institute of Physics

TABLE I. The filter functions $\Omega(u,v)$ for the commonly used rules of association, where the symbol $(\hat{q}^n\hat{p}^m)_w$ denotes the Weyl-symmetrized form of the product $\hat{q}^n\hat{p}^m$, e.g., $(\hat{q}^2\hat{p})_w = (\hat{q}^2\hat{p} + \hat{q}\hat{p}\hat{q} + \hat{p}\hat{q}^2)/3$.

Rule of association		$\Omega(u,v)$
Weyl	$p^m q^n \to (\hat{q}^n\hat{p}^m)_w$	1
Standard	$p^m q^n \to \hat{q}^n\hat{p}^m$	$\exp(-iuv/2\hbar)$
Antistandard	$p^m q^n \to \hat{p}^n\hat{q}^m$	$\exp(iuv/2\hbar)$

The inversion is

$$a(q,p) = \mathrm{Tr}(\hat{A}\Delta^{(\tilde{\Omega})}(q-\hat{q}, p-\hat{p})). \tag{10}$$

Each mapping is characterized by a so-called filter function $\Omega(u,v)$ (Table I), which is chosen to satisfy the trivial normalization condition

$$\Omega(0,0) = 1. \tag{11}$$

The operator $\Delta^{(\tilde{\Omega})}(q-\hat{q}, p-\hat{p})$ is defined in the same fashion as Eq. (9) with filter function

$$\tilde{\Omega}(u,v) = [\Omega(-u,-v)]^{-1}. \tag{12}$$

The problem of expressing an operator in an ordered form according to a prescribed rule is equivalent to an appropriate mapping of the operator on phase space.

The second mapping Ω' is determined by

$$\hat{B} = \int\int b'(q,p)\Delta^{(\tilde{\Omega})}(q-\hat{q}, p-\hat{p})dq\,dp, \tag{13}$$

with inversion

$$b'(q,p) = 2\pi\hbar\,\mathrm{Tr}(\hat{B}\Delta^{(\Omega)}(q-\hat{q}, p-\hat{p})). \tag{14}$$

It is clear that

$$\Theta = (2\pi\hbar)^{-1}\Theta'. \tag{15}$$

Next we wish to find the relation between two different mappings Ω_1 and Ω_2 say. The Δ-operator can be expressed in a slightly different form

$$\Delta^{(\Omega_j)}(q-\hat{q}, p-\hat{p}) = (2\pi\hbar)^{-2}\int\int \Omega_j(u,v)\hat{D}(u,v)\exp\left(\frac{-i(qu+vp)}{\hbar}\right)du\,dv, \tag{16}$$

where $j = 1, 2$,

$$\hat{D}(u,v) = \exp(i(u\hat{q} + v\hat{p})/\hbar) \tag{17}$$

is the well-known displacement operator if we define

$$\hat{a} = (2\hbar)^{-1/2}(\hat{q} + i\hat{p}), \tag{18}$$

and

$$\alpha = (2\hbar)^{-1/2}(-u + iv). \tag{19}$$

We observe that

$$\Delta^{(\tilde{\Omega}_2)}(q-\hat{q}, p-\hat{p}) = (2\pi\hbar)^{-2}\int\int\left(\frac{\Omega_2(u,v)}{\Omega_1(u,v)}\right)\Omega_1(u,v)\hat{D}(u,v) \times\exp(-i(qu+pv)/\hbar)du\,dv$$
$$= (2\pi\hbar)^{-2}\,\frac{\Omega_2(i\hbar\partial/\partial q, i\hbar\partial/\partial p)}{\Omega_1(i\hbar\partial/\partial q, i\hbar\partial/\partial p)} \times\int\int\Omega_1(u,v)\hat{D}(u,v)\exp\left(\frac{-i(qu+pv)}{\hbar}\right)$$
$$\times du\,dv = \left[\frac{\Omega_2(i\hbar\partial/\partial q, i\hbar\partial/\partial p)}{\Omega_1(i\hbar\partial/\partial q, i\hbar\partial/\partial p)}\right] \times\Delta^{(\Omega_1)}(q-\hat{q}, p-\hat{p}). \tag{20}$$

From Eqs. (10) and (20) it follows that

$$a^{(\tilde{\Omega}_2)}(q,p) = L_{21}\left(i\hbar\frac{\partial}{\partial q}, i\hbar\frac{\partial}{\partial p}\right)a^{(\tilde{\Omega}_1)}(q,p), \tag{21}$$

where

$$L_{21}(x,y) = \Omega_2(x,y)/\Omega_1(x,y). \tag{22}$$

Letting $\tilde{\Omega}_j \to \Omega_j$, $j = 1, 2$, and using Eq. (9), we obtain the following differential relation between $a^{(\Omega_1)}(q,p)$ and $a^{(\Omega_2)}(q,p)$:

$$a^{(\Omega_2)}(q,p) = L_{12}\left(-i\hbar\frac{\partial}{\partial q}, -i\hbar\frac{\partial}{\partial p}\right)a^{(\Omega_1)}(q,p). \tag{23}$$

In particular, we choose $\hat{A} = \hat{\rho}$, which is the density operator describing the system of interest, then $a(q,p)$ serves as if it were a classical distribution function. Conventionally $b'(q,p)$ defined by Eq. (13) is called the $\tilde{\Omega}$-equivalence of operator $\hat{B}$, and $a(q,p)$ defined by Eq. (10) the $\tilde{\Omega}$-distribution function, which is usually denoted by $P^{(\tilde{\Omega})}(q,p)$. Thus the expectation value of a quantum mechanical observable $\hat{B}$ can be written in a classical form

$$\langle\hat{B}\rangle = \mathrm{Tr}(\hat{\rho}\hat{B}) = \int\int b'^{(\tilde{\Omega})}(q,p)P^{(\tilde{\Omega})}(q,p)dq\,dp. \tag{24}$$

Also the distribution thus defined satisfies the normalization condition

$$\int\int P^{(\tilde{\Omega})}(q,p)dq\,dp = \mathrm{Tr}(\hat{\rho}) = 1. \tag{25}$$

For example, if we consider the simplest case where

$$\tilde{\Omega}(u,v) = \Omega(u,v) = 1, \tag{26}$$

then it leads to the famous Wigner distribution function and the Wigner equivalence (denoted by suffix w)[5]:

$$b'^{(w)}(q,p) = \int\left\langle p - \frac{u}{2}\middle|\hat{B}\middle|p + \frac{u}{2}\right\rangle\exp\left(\frac{-iuq}{\hbar}\right)du. \tag{27}$$

The Wigner equivalence of an operator $\hat{F} = \hat{B}\hat{C}$ can be expressed in terms of those corresponding to $\hat{B}$ and $\hat{C}$ through the Groenewold theorem[13]

$$f'^{(w)}(q,p) = b'^{(w)}(q,p)\exp(\hbar\overleftrightarrow{G}/2i)c'^{(w)}(q,p), \tag{28}$$

where

$$\overleftrightarrow{G} = \left(\frac{\overleftarrow{\partial}}{\partial p}\right)\left(\frac{\overrightarrow{\partial}}{\partial q}\right) - \left(\frac{\overleftarrow{\partial}}{\partial q}\right)\left(\frac{\overrightarrow{\partial}}{\partial p}\right), \tag{29}$$

and the arrows indicate in which direction the derivatives act.

One of the major advantages of the Wigner equivalence is that it leads to the simplest forms for the quantum corrections to the corresponding classical quantity,[14] and therefore is very useful to the semiclassical calculations.[15] It can assume negative values, which makes it quite different from classical distribution functions.

Another choice of the filter function leads to the antistandard-ordered distribution function[6] (see Table I),

which has the important property that it is non-negative everywhere in phase space.[16] The class of non-negative quantum distribution functions has been shown to be rather small.[17]

III. CLASSICAL LIMIT $\hbar \to 0$

With the help of the formulas mentioned in the Sec. II, we now consider taking the classical limit $\hbar \to 0$.

First of all we observe from Eqs. (11), (22), and (23) that any phase-space distribution function that describes the same system, approaches the same limit at $\hbar \to 0$.[18] Also, any phase-space equivalence (resulting from any rule of association) of the same quantum mechanical operator approaches the same limit at $\hbar \to 0$. Explicitly we have

$$\lim_{\hbar \to 0} b'^{(\hat{\Omega})}(q,p) = b^{(\mathrm{cl})}(q,p), \tag{30}$$

and

$$\lim_{\hbar \to 0} P^{(\tilde{\Omega})}(q,p) = P^{(\mathrm{cl})}(q,p). \tag{31}$$

Of course the necessary and sufficient condition for any of Eqs. (30) and (31) to be true is that the appropriate limit exists, which will be assumed in the following discussions.

Equation (31) thus defines a classical distribution $P^{(\mathrm{cl})}(q,p)$. We can prove that $P^{(\mathrm{cl})}(q,p)$ is real and non-negative everywhere in phase space simply by choosing a real and non-negative quantum distribution function, e.g., the anti-standard-ordered distribution function, on the left-hand side of Eq. (31). In the case of a canonical ensemble, $P^{(\mathrm{cl})}(q,p)$ turns out to be the Maxwell–Boltzmann distribution.[5]

Now we would like to consider the properties that the "classical functions" $b^{(\mathrm{cl})}(q,p)$ and $P^{(\mathrm{cl})}(q,p)$ possess. The conclusions at which we just arrived make it enough to restrict ourselves within the Wigner equivalence and distribution function.

By using Eqs. (28) and (29) we get the Wigner equivalence of $\hat{B}^n$, where n is a positive integer,

$$\begin{aligned}(\hat{B}^n)^{(w)} &= b'^{(w)}(q,p)\exp(\hbar\overleftrightarrow{G}/2i)(\hat{B}^{n-1})^{(w)} \\ &= b'^{(w)}(q,p)\exp(\hbar\overleftrightarrow{G}/2i) \\ &\quad \times [b'^{(w)}(q,p)\exp(\hbar\overleftrightarrow{G}/2i)(\hat{B}^{n-2})^{(w)}].\end{aligned} \tag{32}$$

Obviously, $\exp(\hbar\overleftrightarrow{G}/2i)$ approaches its identity at $\hbar \to 0$. Hence

$$\lim_{\hbar \to 0} (\hat{B}^n) = (b^{(\mathrm{cl})}(q,p))^n. \tag{33}$$

It is easy to see that for any infinitely differentiable function $R(t)$, we have the following useful relation:

$$\lim_{\hbar \to 0} (R(\hat{B}))^{(\Omega)} = R(b^{(\mathrm{cl})}(q,b)), \tag{34}$$

where suffix Ω denotes an arbitrary Ω-equivalence.

On the other hand we have

$$\mathrm{Tr}(\hat{\rho}R(\hat{B})) = \int\int P^{(\Omega)}(q,p)(R(\hat{B}))^{(\Omega)}dq\,dp. \tag{35}$$

Let $\hbar \to 0$ at both sides

$$\lim_{\hbar \to 0} \mathrm{Tr}(\hat{\rho}R(\hat{B})) = \langle R(\hat{B})\rangle_{\mathrm{cl}}, \tag{36}$$

where

$$\langle R(\hat{B})\rangle_{\mathrm{cl}} = \int\int P^{(\mathrm{cl})}(q,p)R(b^{(\mathrm{cl})}(q,p))dq\,dp, \tag{37}$$

and its existence has been assumed.

IV. RELATION BETWEEN QUANTUM MECHANICAL AND CLASSICAL ENTROPY

Traditionally entropy is introduced in the phenomenological thermodynamical considerations based on the second law of thermodynamics. The conception of entropy thus defined frequently leads to some obscure ideals.[19] The well-known heat death provides a good example. In classical statistical mechanics the Boltzmann and the Gibbs entropies are not very good ones either. The reason is that they never lead to the third law of thermodynamics. Thus a correct definition of entropy is only possible in the framework of quantum mechanics.

If a system is described by a density operator $\hat{\rho}$, its entropy is then defined quantum mechanically by

$$\begin{aligned}S(\hat{\rho}) &= -k\,\mathrm{Tr}(\hat{\rho}\ln\hat{\rho}) \\ &= -k\int\int P^{(w)}(q,p)(\ln\hat{\rho})^{(w)}dq\,dp \\ &= -k\int\int P^{(w)}(q,p)\left(\ln\left(\frac{\hat{\rho}}{2\pi\hbar}\right)\right)^{(w)}dq\,dp - k\ln(2\pi\hbar).\end{aligned} \tag{38}$$

Noticing that

$$(\hat{\rho}/2\pi\hbar)^{(w)} = P^{(w)}(q,p), \tag{39}$$

we find, according to Eq. (36), that the first term in Eq. (38) approaches

$$-k\int\int P^{(\mathrm{cl})}(q,p)\ln P^{(\mathrm{cl})}(q,p)dq\,dp \tag{40}$$

in the limit $\hbar \to 0$. But the second term diverges to positive infinity. If the classical entropy functional is defined by

$$S^{(\mathrm{cl})}(P^{(\mathrm{cl})}) = -k\int\int P^{(\mathrm{cl})}(q,p)\ln P^{(\mathrm{cl})}(q,p)dq\,dp - k\ln(2\pi\hbar), \tag{41}$$

then the quantum mechanical entropy approaches the classical entropy functional in the limit $\hbar \to 0$, in the following sense:

$$\lim_{\hbar \to 0}(S(\hat{\rho}) - \ln(2\pi\hbar)) = \lim_{\hbar \to 0}(S^{(\mathrm{cl})}(P^{(\mathrm{cl})}) - \ln(2\pi\hbar)), \tag{42}$$

or

$$\lim_{\hbar \to 0}(S(\hat{\rho}) - S^{(\mathrm{cl})}(P^{(\mathrm{cl})})) = 0. \tag{43}$$

Let us consider a simple example, i.e., an ensemble of harmonic oscillators with a heat bath of temperature T. The easiest way to compute the quantum mechanical entropy is to use the Wigner phase-space equivalence and distribution function.

From Ref. 20,

$$\begin{aligned}b'^{(w)}(q,p) &= (\exp(-\beta\hat{H}))^{(w)} \\ &= \mathrm{sech}(\hbar\omega\beta/2)\exp(-2\tanh(\hbar\omega\beta/2)\mathrm{H}/\hbar\omega),\end{aligned} \tag{44}$$

and the partition function is

$$Z = \mathrm{Tr}(\exp(-\beta\hat{H})) = (2\pi\hbar)^{-1}\int\int b'^{(w)}(q,p)dq\,dp = (2\sinh(\hbar\omega\beta/2))^{-1}. \quad (45)$$

Hence the Wigner distribution function is an immediate result of Eqs. (44) and (45),

$$P^{(w)}(q,p) = (\hat{\rho}/2\pi\hbar)^{(w)} = b'^{(w)}(q,p)/(2\pi\hbar Z) = (\pi\hbar)^{-1}\tanh(\hbar\omega\beta/2) \times \exp((-2/\hbar\omega)\tanh(\hbar\omega\beta/2)H), \quad (46)$$

where the Hamiltonian is

$$\hat{H} = \hat{p}^2/2m + m\omega^2\hat{q}^2/2. \quad (47)$$

The quantum mechanical entropy can be obtained by

$$S(\hat{\rho}) = k\left(\ln Z - \beta\frac{\partial \ln Z}{\partial\beta}\right) = (k\beta\hbar\omega/2)/\tanh(\hbar\omega\beta/2) - k\ln(2\sinh(\hbar\omega\beta/2)). \quad (48)$$

The Wigner distribution function approaches the classical canonical distribution function at $\hbar \to 0$,

$$\lim_{\hbar\to 0} P^{(w)}(q,p) = (\omega\beta/2\pi)\exp(-\beta H), \quad (49)$$

as predicted by the general considerations. The classical partition function is, by definition,

$$Z^{(\mathrm{cl})} = (\hbar\omega\beta)^{-1}. \quad (50)$$

Finally, the classical entropy has the form

$$S^{(\mathrm{cl})} = k - k\ln(\hbar\omega\beta). \quad (51)$$

With the help of Eqs. (48) and (51), Eqs. (42) and (43) are maintained.

V. CONCLUSIONS

We discussed the general phase-space representation of quantum mechanics at the classical limit $\hbar \to 0$. We proved that every representation approaches the same "limit representation" at $\hbar \to 0$. The open question on the relation between the classical and quantum mechanical entropies was answered. The differences between the classical and quantal entropies are shown to approach zero at the classical limit $\hbar \to 0$.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. F. O'Connell for many helpful conversations.

This research was partially supported by the Division of Materials Science, U. S. Department of Energy under Grant No. DE-FG05-84ER45135.

APPENDIX: COMMENTS ON TWO OF THE PROBLEMS DISCUSSED IN REF. 12

In a recent paper,[12] Beretta gave a set of rather restrictive conditions defining a complete classical phase space representation of quantum kinematics for systems with both classical and quantum mechanical descriptions. With help of the general considerations of phase-space representation we wish to make some comments on Beretta's ideals and derivations. In order to keep consistent with the notations that we have been using, we quote those conditions in a slightly different form.

Given a system with quantum mechanical Hilbert space $\mathscr{H}$ and classical space $\mathscr{P}$, find two mappings $w(q,p;\hat{\rho})$ and $b(q,p;\hat{B})$ that satisfy the following conditions. For every density operator $\hat{\rho}$ on $\mathscr{H}$, every well-defined Hermitian operator $\hat{B}$ on $\mathscr{H}$, every point (q,p) in $\mathscr{P}$, and every continuous real function $R(t)$ of the real variable t,

(i) $w(q,p;\hat{\rho})$ is real and non-negative, (A1)

(ii) $b(q,p;\hat{B})$ is real, (A2)

(iii) $$\int\int R(w(q,p;\hat{\rho}))dq\,dp = \mathrm{Tr}(R(\hat{\rho})), \quad (A3)$$

(iv) $$\int\int w(q,p;\hat{\rho})R(b(q,p;\hat{B}))dq\,dp = \mathrm{Tr}(\hat{\rho}R(\hat{B})). \quad (A4)$$

The purpose of seeking this representation is to show that the quantum mechanical entropy is exactly equal to a "classical entropy functional," which is defined by

$$S^{(\mathrm{cl})}(w) = -k\int\int w(q,p;\hat{\rho})\ln 2\pi\hbar w(q,p;\hat{\rho})dq\,dp. \quad (A5)$$

If such a representation exists, then we choose $R(t) = -k\ln t$ and from (A10) obtain

$$S(\hat{\rho}) = -k\,\mathrm{Tr}(\hat{\rho}\ln\hat{\rho}) = S^{(\mathrm{cl})}(w). \quad (A6)$$

Beretta did not know whether this representation existed or not. After making a conjecture, Beretta tried to prove that this representation was just the one to which the Wigner, the Blokhintzev, and the Wehrl[21] phase-space representations (R_0) converge in the classical limit $\hbar \to 0$.

Although we do not know whether this representation exists, we are able to conclude that R_0 is an incorrect candidate for the representation, the reason being that in R_0, Eqs. (A3) and (A4) hold only after limit $\hbar \to 0$ are taken in the right-hand sides.

Now we turn to consider another problem discussed in Ref. 12: the behavior of the spectral expansions in the classical limit $\hbar \to 0$. The density operator can be written as follows:

$$\hat{\rho} = \sum_{i=0}^{\infty} w_j\hat{P}_j, \quad (A7)$$

where $\hat{P}_j = |\psi_j\rangle\langle\psi_j|$ is the projector onto the eigenspace $|\psi_j\rangle$ with eigenvalue

$$\omega_j = [\exp(-\beta E_j)]/Z \quad (A8)$$

and

$$\hat{H}|\psi_j\rangle = E_j|\psi_j\rangle. \quad (A8')$$

By definition we have

$$\sum_{j=0}^{\infty}\hat{P}_j = \hat{I}, \quad (A9)$$

where $\hat{I}$ denotes the identity operator.

The Wigner equivalences of Eqs. (A7) and (A8) are

$$P^{(\omega)}(q,p)=\sum_{j=0}^{\infty} w_j P_j^{(\omega)}(q,p). \tag{A10}$$

The relation between $P_j^{(\omega)}(q,p)$ and $r(q,p;\hat{P}_j)$ in Ref. 12 is

$$P_j^{(\omega)}(q,p)=(2\pi\hbar)^{-1}r(q,p;\hat{P}_j). \tag{A11}$$

Next we consider letting $\hbar\to 0$. It has been shown that[22]

$$\lim_{\hbar\to 0} P_j^{(\omega)}(q,p)=\delta(I(q,p)-I_j)/2\pi, \tag{A12}$$

where I_j is the semiclassical action associated with $|\psi_j\rangle$ [i.e., $I=(j+\gamma)\hbar$ with γ the Maslov index].

While quantization disappears in the classical limit $\hbar\to 0$, we expect

$$\lim_{\hbar\to 0} w_j=0, \tag{A13}$$

since w_j is the probability of the system being in state $|\psi_j\rangle$.

Thus when $\lim_{\hbar\to 0}$ is applied to both sides of Eq. (A10), the order of $\lim_{\hbar\to 0}$ and $\Sigma_{n=0}^{\infty}$ cannot be exchanged. Furthermore it is easily verified that

$$\alpha_j(q,p)=\lim_{\hbar\to 0} r(q,p;\hat{P}_j)=0. \tag{A14}$$

Therefore Eqs. (34), (35), and (39), the conjecture, in Ref. 12 are not valid.

[1] K. Hepp, Commun. Math. Phys. **35**, 265 (1974).
[2] J. M. Combes, R. Schrader, and R. Seiler, Ann. Phys. **111**, 1 (1978).
[3] B. Simon, Commun. Math. Phys. **71**, 247 (1980).
[4] H. Hogreve, J. Potthoff, and R. Schrader, Commun. Math. Phys. **91**, 573 (1983).
[5] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[6] K. Husimi, Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
[7] K. E. Cahill, and R. J. Glauber, Phys. Rev. **177**, 1857, 1882 (1968).
[8] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[9] G. S. Agarwal, and E. Wolf, Phys. Rev. D **2**, 2161 (1970).
[10] See the recent reviews, e.g., R. F. O'Connell, Found. Phys. **13**, 83 (1983); V. I. Tatarskii, Sov. Phys. Usp. **26**, 311 (1983); M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, Phys. Rep. **106**, 123 (1984).
[11] A. Wehrl, Rep. Math. Phys. **16**, 353 (1979).
[12] G. P. Beretta, J. Math. Phys. **25**, 1507 (1984).
[13] H. J. Groenewold, Physica **12**, 405 (1946).
[14] R. F. O'Connell and Lipo Wang, Phys. Rev. A **30**, 2187 (1984); **31**, 1707 (1985); Phys. Lett. A **107**, 9 (1985).
[15] Lipo Wang and R. F. O'Connell (to be published).
[16] N. D. Cartwright, Physica A **83**, 210 (1976).
[17] A. J. E. M. Janssen, J. Math. Phys. **25**, 2240 (1984).
[18] G. C. Summerfield and P. F. Zweifel, J. Math. Phys. **10**, 233 (1969).
[19] For recent reviews see, e.g., A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[20] K. Imre, E. Ozizmir, M. Rosenbaum, and P. Zweifel, J. Math. Phys. **8**, 1097 (1967).
[21] It is worth noticing that the Wehrl phase-space representation that Beretta referred to is actually the antinormal-ordered (Husimi) one. See, e.g., Ref. 6 and Ref. 7.
[22] See, e.g., W. H. Miller, Adv. Chem. Phys. **25**, 69 (1974); M. V. Berry, in *Topics In Nonlinear Dynamics*, edited by S. Jorna (A. I. P., New York, 1978); M. V. Berry, Philos. Trans. R. Soc. London **287**, 237 (1977).

Numerical integration in many dimensions. III

Charles Schwartz
Department of Physics, University of California, Berkeley, California 94720

(Received 3 June 1985; accepted for publication 27 September 1985)

Extending a previous line of work, a powerful computational method is found for numerical integration in many dimensions of functions of the form $F(f_1(x_1,x_2)+f_2(x_2,x_3)+f_3(x_3,x_4)+\cdots+f_d(x_d,x_1))$.

I. INTRODUCTION

In a previous paper[1] a method for fast and accurate machine computation of d-dimensional integrals, where the integrand was of the form $F(f_1(x_1)+f_2(x_2)+\cdots+f_d(x_d))$ was presented. The first step was to introduce an integral transform representation of the function F so that its argument appeared in an exponential, then each of the d integrations over the coordinates x_i could be done separately, with the final product then numerically integrated over the transform variable. If n lattice points were needed for the adequate numerical evaluation of each one-dimensional integral, then this method would require of the order of n^2d operations: This is enormously better than the n^d operations that would be required in a direct integration method. Now this method of approach is extended to a more complicated integrand, in which the argument of the general function F has the coordinates x_i linked together in a chain.

II. THE METHOD

Consider the integral over the d-dimensional product space

$$I=\left(\prod_{i=1}^{d}\int g_i(x_i)dx_i\right)F\left(\sum_{i=1}^{d}f_i(x_i,x_{i+1})\right), \tag{1}$$

where $x_{d+1}=x_1$. Start, as before, with some integral transform

$$F(s)=\int d\sigma\,\hat{F}(\sigma)e^{su(\sigma)}, \tag{2}$$

where the integration takes place along some suitable contour. Then we have

$$I=\int d\sigma\,\hat{F}(\sigma)J(\sigma), \tag{3}$$

where

$$J(\sigma)=\left(\prod_{i=1}^{d}\int g_i(x_i)dx_i\right)\exp\left(u(\sigma)\sum_{i=1}^{d}f_i(x_i,x_{i+1})\right). \tag{4}$$

Now introduce the numerical quadrature rule of choice for each x_i:

$$\int h(x)dx\cong\sum_{j=1}^{n}w_jh(z_j). \tag{5}$$

We assume, only for simplicity of notation, that we use the same quadrature rule (points z_j and weights w_j) for each x_i integration variable.

Now comes the coup. Notice, that with the definition

$$A^i_{jj'}(\sigma)=w_j\,g_i(z_j)\exp[u(\sigma)f_i(z_j,z_{j'})], \tag{6}$$

we can write the multiple integration in terms of the matrices A^i:

$$J(\sigma)=\text{Trace}\,A^1(\sigma)A^2(\sigma)\cdots A^d(\sigma). \tag{7}$$

There are n^2 elements in each of d matrices, and these must be evaluated for each of n values of σ. The multiplication of two matrices requires n^3 multiplications of numbers. Therefore the total amount of computer time for this method is of the order of n^3d function evaluations plus n^4d additional multiplications. For n of the order of 10, this means that we can evaluate integrals with d into the hundreds or more for pennies.

Once again, a problem that seemed to increase exponentially with the number of dimensions has been reduced to a procedure that increases only linearly. The choice of the integral transform is of course important, and the reader is referred to Ref. 1, where several examples are given.

III. FURTHER COMMENTS

If the entire integrand is symmetric in all variables (all functions f_i and g_i given by a single f and g), then there is only a single matrix A; and then

$$J=\sum_{j=1}^{n}(\lambda_j)^d, \tag{8}$$

where the λ_j are the eigenvalues of the matrix A (for each value of σ). Thus we can even take the limit as d goes to infinity, with the answer given in terms of the largest eigenvalue of A.

If the structure of the integrand is that of an *open* chain [i.e., if the function $f_d(x_d,x_1)$ is absent in (1)], then the problem is simplified a bit. The work of multiplying the matrices is reduced by a factor of n.

The technique used here for handling the multiple sum over chain-linked variables leads to the study of some well-known problems in statistical mechanics. I have applied this approach to the Ising model in one, two, and three dimensions; and these results will be published separately.

[1] C. Schwartz, J. Math. Phys. **26**, 951 (1985).

 0022-2488/86/020488-01$02.50 © 1986 American Institute of Physics

On the Poisson brackets of differentiable generators in classical field theory

J. David Brown
Center for Theoretical Physics, The University of Texas at Austin, Austin, Texas 78712

Marc Henneaux[a)]
Faculté des Sciences, Université Libre de Bruxelles, Campus Plaine C. P. 231, B-1050, Bruxelles, Belgium[b)]
and Centro de Estudios Científicos de Santiago, Casilla 16443, Santiago 9, Chile

(Received 10 June 1985; accepted for publication 16 September 1985)

The canonical formulation of field theory on open spaces is considered. It is proved, under appropriate assumptions, that the Poisson bracket of two differentiable generators is also a differentiable generator.

I. INTRODUCTION

The following question naturally arises in the study of the Hamiltonian formulation of Einstein theory on open spacelike sections. Given the generators of two asymptotic symmetry transformations, supplemented by the appropriate surface terms at spatial infinity that make them differentiable (i.e., that make their functional derivatives well defined),[1,2] can it be taken for granted[3] that their Poisson bracket is also a differentiable generator? In this paper, we analyze this question and answer it affirmatively.

Because the theorem to be derived here can be applied to more general situations than just general relativity, we will adopt notations that do not explicitly refer to that case. The canonical variables will be denoted by $y^A(x)$ and the Poisson bracket of two differentiable functionals $F[y]$, $G[y]$ of the canonical variables will be denoted by

$$[F, G] = \int \frac{\delta F}{\delta y^A(x)} \frac{\delta G}{\delta y^B(x)} \sigma^{AB} d^n x, \tag{1}$$

where σ^{AB} is, for simplicity, a constant anstisymmetric matrix. In (1), $\delta F/\delta y^A(x)$ is the functional derivative of F with respect to $y^A(x)$, defined by

$$\delta F = \int \frac{\delta F}{\delta y^A(x)} \delta y^A(x) d^n x \tag{2}$$

for any allowed variation δy^A of the fields. The functional F is said to be differentiable if its variation can indeed be brought to the form (2), with functional derivatives $\delta F/\delta y^A(x)$ that are regular functions of x.

The formula (1) can be rewritten in various useful ways as follows:

$$[F, G] = \delta_F G = -\delta_G F \tag{3a}$$

$$= \int \frac{\delta G}{\delta y^A(x)} \delta_F y^A(x) d^n x, \tag{3b}$$

with

$$\delta_F y^A(x) = \frac{\delta F}{\delta y^B(x)} \sigma^{BA}. \tag{4}$$

In the open case considered here, the field configurations $y^A(x)$ are restricted by appropriate boundary conditions at spatial infinity, which read

$$\chi^\alpha(y^A, y^A_i, \ldots y^A_{i_1\cdots i_q}, x^j) \to 0, \tag{5}$$

for some given functions χ^α of the fields and their derivatives $y^A_{i_1 i_2 \cdots i_k} \equiv \partial y^A/\partial x^{i_1} \partial x^{i_2} \cdots \partial x^{i_k}$, and for some specified rate of approach.

If the differentiable functional G is such that $y^A(x) + \delta_G y^A(x)$ obeys the boundary conditions whenever $y^A(x)$ does, G is said to be a differentiable generator. In that case the infinitesimal canonical transformation generated by G maps an allowed configuration on to another allowed configuration.

From now on, F and G will always stand for differentiable generators which are given by integrals of local densities,

$$F = \int f(y^A, y^A_k, y^A_{kl}, \ldots, y^A_{k_1\cdots k_m}, x^i) d^n x, \tag{6a}$$

$$G = \int g(y^A, y^A_k, y^A_{kl}, \ldots, y^A_{k_1\cdots k_p}, x^i) d^n x. \tag{6b}$$

Moreover, we assume that f and g are differentiable (as many times as needed) functions of their arguments. This case covers all field theoretical models of common interest.

From (6), it is easy to obtain[4]

$$\delta F = \int d^n x \frac{\delta f}{\delta y^A} \delta y^A + \oint d^{n-1} S_k V^k_f(\delta y), \tag{7}$$

where the following conventions have been adopted:

$$\frac{\delta f}{\delta y^A} = \frac{\partial f}{\partial y^A} - \partial_k \frac{\partial f}{\partial y^A_K} + \cdots + (-)^m \partial_{k_1\cdots k_m} \frac{\partial f}{\partial y^A_{k_1\cdots k_m}}, \tag{8a}$$

$$V^k_f(\delta y) = \frac{\delta f}{\delta y^A_k} \delta y^A + \frac{\delta f}{\delta y^A_{kl}} \delta y^A_l + \cdots + \frac{\delta f}{\delta y^A_{kl_2\cdots l_m}} \delta y^A_{l_2\cdots l_m}, \tag{8b}$$

$$\frac{\delta f}{\delta y^A_{k_1\cdots k_j}} = \frac{\partial f}{\partial y^A_{k_1\cdots k_j}} - \partial_{k_{j+1}} \frac{\partial f}{\partial y^A_{k_1\cdots k_{j+1}}} + \cdots + (-)^{m-j} \partial_{k_{j+1}\cdots k_m} \frac{\partial f}{\partial y^A_{k_1\cdots k_m}}. \tag{8c}$$

The functions $\delta f/\delta y^A_{k_1\cdots k_j}$ are the "variational derivatives" of f with respect to $y^A_{k_1\cdots k_j}$.

Since the (already improved) functional F is differentiable by hypothesis, its variation (7) must reduce to the form written in (2) for any δy^A that preserves the boundary conditions (5). For the case in which the variations δy^A are of compact support but otherwise arbitrary, the surface inte-

[a)] Chercheur qualifié au Fonds National Belge de la Recherche Scientifique.
[b)] Permanent address.

 0022-2488/86/020489-03$02.50 © 1986 American Institute of Physics

gral in (7) vanishes and the variation δF indeed reduces to the form (2). It then follows that at each point x,

$$\frac{\delta F}{\delta y^A(x)} = \frac{\delta f}{\delta y^A}, \tag{9}$$

where the right-hand side of (9) is understood to be evaluated at x. Now by comparing (7) with (2) for arbitrary variations δy^A compatible with the boundary conditions, the surface integral is seen to vanish,

$$\oint d^{n-1}S_k \; V_f^k(\delta y) = 0, \tag{10}$$

for all allowed δy^A. Similar expressions hold for G.

We are now in a position to formulate the main theorem of our paper.

Theorem: The bracket $[F, G]$ of two differentiable generators is itself a differentiable generator.

This theorem contains two assertions. The first one is that $[F, G]$ is differentiable. The second one is that $\delta_{[F,G]} y^A$ has the correct asymptotic behavior, i.e., that $[F, G]$ generates a canonical transformation that preserves the boundary conditions.

II. PROOF OF THE THEOREM

In order to prove the theorem, we first demonstrate the following lemmas.

Lemma 1: Let $\mu^A(y^A, y^A_i, \ldots, y^A_{i_1\cdots i_s}, x^j)$ and $\nu^A(y^A, y^A_i, \ldots, y^A_{i_1\cdots i_t}, x^j)$ be two allowed, field-dependent, "infinitesimal" variations of y^A (i.e., let $y^A + \mu^A$ and $y^A + \nu^A$ obey the boundary conditions). Then, the bracket $[\mu, \nu]^A$ defined by

$$[\mu, \nu]^A = \delta_\mu \nu^A - \delta_\nu \mu^A \tag{11}$$

is also an allowed infinitesimal variation of y^A. In (11), $\delta_\mu \nu^A$ stands for the change in ν^A induced by the variation μ^A,

$$\delta_\mu \nu^A = \sum_{k=0}^{t} \frac{\partial \nu^A}{\partial y^c_{i_1\cdots i_k}} \mu^c_{i_1\cdots i_k}. \tag{12}$$

The proof of this lemma is straightforward. Infinitesimal variations δy^A are indeed characterized by

$$\delta\chi^\alpha = \sum_{k=0}^{q} \frac{\partial \chi^\alpha}{\partial y^A_{i_1\cdots i_k}} \delta y^A_{i_1\cdots i_k} \to 0. \tag{13}$$

Also, note that $\delta_{[\mu,\nu]}\chi^\alpha$ is equal to $\delta_\mu\delta_\nu\chi^\alpha - \delta_\nu\delta_\mu\chi^\alpha$ [use $\delta_\mu h_{i_1\cdots i_k} = (\delta_\mu h)_{i_1\cdots i_k}$]. Then since $\delta_\nu\chi^\alpha$ and $\delta_\mu\chi^\alpha$ fall off for all admissible field configurations, it follows successively that $\delta_\mu\delta_\nu\chi^\alpha$, $\delta_\nu\delta_\mu\chi^\alpha$, and $\delta_{[\mu,\nu]}\chi^\alpha$ also go to zero at infinity and thus, that $[\mu, \nu]^A$ preserves the boundary conditions. □

Lemma 2: Let μ^A and ν^A be as above. In performing the integrations by parts needed to transform

$$\int \delta_\nu\left(\frac{\delta f}{\delta y^B}\right) \cdot \mu^B \, d^n x$$

into a volume piece containing only undifferentiated ν^A plus a surface integral, that surface integral is simply given by

$$\oint d^{n-1}S_k \; A_f^k(\mu, \nu), \tag{14a}$$

with

$$A_f^k(\mu, \nu) = \delta_\mu V_f^k(\nu) - \delta_\nu V_f^k(\mu) + V_f^k([\nu, \mu]). \tag{14b}$$

This can be seen by using again the simple identity

$$\delta_\nu\delta_\mu \; f - \delta_\mu\delta_\nu f = \delta_{[\nu,\mu]} f. \tag{15}$$

By integrating (15) over x and taking into account (7) and the definition of the bracket $[\nu, \mu]^A$, this yields

$$\int \delta_\nu\left(\frac{\delta f}{\delta y^A}\right)\mu^A \, d^n x = \int \delta_\mu\left(\frac{\delta f}{\delta y^A}\right)\nu^A \, d^n x + \oint d^{n-1}S_k \; A_f^k(\mu, \nu), \tag{16}$$

where $A_f^k(\mu, \nu)$ is given by (14). Since the volume integral on the right-hand side of (16) contains no derivative of ν^A, the second lemma is proved. □

Incidentally, (16) shows that the polynomials $\delta_\nu(\delta f/\delta y^A)$ are "self-adjoint," as indeed they should be (see, e.g., Ref. 4, pp. 202–204).

Lemma 3: Let μ^A and ν^A be given by

$$\mu^A = \delta_F y^A = \frac{\delta f}{\delta y^B}\sigma^{BA}, \quad \nu^A = \delta_G y^A = \frac{\delta g}{\delta y^B}\sigma^{BA}. \tag{17a}$$

The bracket $[\mu, \nu]^A$ is then equal to

$$[\mu, \nu]^A = \frac{\delta}{\delta y^B}\left(\frac{\delta f}{\delta y^C}\frac{\delta g}{\delta y^D}\sigma^{CD}\right)\sigma^{BA}. \tag{17b}$$

This is a straightforward consequence of the following identity of the calculus of variations:

$$\frac{\delta}{\delta y^A}\left(\frac{\delta f}{\delta y^B}\frac{\delta g}{\delta y^C}\sigma^{BC}\right) = \delta_F\frac{\delta g}{\delta y^A} - \delta_G\frac{\delta f}{\delta y^A} \tag{18}$$

(where $\delta_F = \delta_\mu$, $\delta_G = \delta_\nu$).

To prove this identity, integrate the right-hand side of (18), multiplied by an arbitrary function ϵ^A of compact support, over the whole space and make all necessary integrations by parts. The resulting surface integrals will all vanish since ϵ^A is identically zero sufficiently close to spatial infinity. This gives, using the equality (16),

$$\begin{aligned}\int \epsilon^A\left(\delta_F\frac{\delta g}{\delta y^A} - \delta_G\frac{\delta f}{\delta y^A}\right)d^n x &= \int\left(\delta_F y^A\,\delta_\epsilon\frac{\delta g}{\delta y^A} - \delta_G y^A\,\delta_\epsilon\frac{\delta f}{\delta y^A}\right)d^n x \\ &= \int \delta_\epsilon\left(\frac{\delta f}{\delta y^A}\frac{\delta g}{\delta y^B}\sigma^{AB}\right)d^n x.\end{aligned} \tag{19}$$

Again making an integration by parts and taking the arbitrariness of ϵ^A into account, (18) is easily seen to hold true. □

We now turn to the demonstration of the theorem itself. The first step in the demonstration amounts to showing that $[F, G]$ is differentiable. This is done by relating the surface term that appears in the variation of $[F, G]$ to the surface terms that arise in the variations of F and G, and that are known to vanish.

The definition (1) and the equality (9) may be used to write the variation of $[F, G]$ for an arbitrary (allowed) variation ν^A of y^A as

$$\delta_\nu[F, G] = \int \delta_\nu\left(\frac{\delta g}{\delta y^B}\right)\delta_F y^B \, d^n x - \int \delta_\nu\left(\frac{\delta f}{\delta y^B}\right)\delta_G y^B \, d^n x. \tag{20}$$

We want to prove that the surface integrals arising from the appropriate integrations by parts in each term of the right-hand side of (20) are separately zero. Since these terms have the same structure, it is enough to show this property for one of them.

From the second lemma, we know that the surface integral arising in $\int \delta_\nu(\delta f/\delta y^B)\,\delta_G y^B\, d^n x$ is given by $\oint A^k_f(\delta_G y, \nu) d^{n-1}S_k$ [formula (14)], where

$$A^k_f(\delta_G y, \nu) = \delta_G V^k_f(\nu) - \delta_\nu V^k_f(\delta_G y) + V^k_f([\nu, \delta_G y]). \tag{21}$$

But each of the terms in $A^k_f(\delta_G y, \nu)$ contribute zero, because (i) $\oint V^k_f(\nu) d^{n-1}S_k$ vanishes for any allowed variation performed about any allowed configuration, and so also does

$$\oint \delta_G V^k_f(\nu) d^{n-1}S_k = \delta_G \oint V^k_f(\nu) d^{n-1}S_k;$$

(ii) similary, $\oint \delta_\nu V^k_f(\delta_G y)\, d^{n-1}S_k$ vanishes for an identical reason; and (iii) $[\nu, \delta_G y]^A$ is an allowed variation by virtue of the first lemma, and so its flux integral $\oint V^k_f([\nu, \delta_G y]) d^{n-1}S_k$ is also zero. Hence, there is no unwanted surface term in $\delta_\nu[F, G]$.

This proves that the Poisson bracket $[F, G]$ is differentiable; its functional derivatives are equal to [see (1) and (9)],

$$\frac{\delta[F, G]}{\delta y^A(x)} = \frac{\delta}{\delta y^A}\left(\frac{\delta f}{\delta y^B}\frac{\delta g}{\delta y^C}\right)\sigma^{BC}. \tag{22}$$

Then the result of the third lemma can be rewritten as

$$\delta_{[F, G]} y^A = [\delta_F, \delta_G] y^A, \tag{23}$$

and it follows from the first lemma that $\delta_{[F, G]} y^A$ is an allowed variation. This shows that the canonical transformation generated by $[F, G]$ preserves the boundary conditions and completes the demonstration of the theorem.

The conclusion of our paper is that as soon as the canonical generators of a field theory are supplemented, in the spatially open case, by the appropriate surface terms needed to make them well defined, then their Poisson brackets may be taken as many times as desired (provided these generators are given by the integrals of C^∞ local densities). In that sense, the problem of the surface terms arises only once. In addition, the Poisson brackets possess the nice properties expected from classical mechanics with a finite number of degrees of freedom [cf. (23)].

As a final point, we note that the results of our paper can be easily extended to the case when some of the variables $y^A(x)$ are fermionic.

ACKNOWLEDGMENTS

It is a pleasure to thank Jules Géhéniau and Claudio Teitelboim for useful discussions.

This work has been supported in part by the National Science Foundation under Grant No. PHY-8216715 and by research funds from the University of Texas at Austin, Center for Theoretical Physics.

[1]The need to supplement the Dirac–A.D.M. Hamiltonian of general relativity by appropriate surface terms at spatial infinity has been pointed out, in the asymptotically flat case, by Regge and Teitelboim [T. Regge and C. Teitelboim, Ann. Phys. N.Y. **88**, 286 (1974)]. As shown by the authors, the surface terms yield the A.D.M. energy-momentum. These ideas have been further discussed in Y. Choquet-Bruhat, A. Fischer, and J. Marsden, "Maximal hypersurfaces and positivity of mass," in *Proceedings of the Enrico Fermi Summer School of the Italian Physical Society*, Course LXVII, edited by J. Ehlers (North-Holland, Amsterdam, 1979).

[2]The extension of the results of Ref. 1 to the asymptotically anti-de Sitter case has been done in M. Henneaux and C. Teitelboim, Phys. Lett. B **142**, 355 (1984); M. Henneaux and C. Teitelboim, Commun. Math. Phys. **98**, 391 (1985).

[3]This question is important for deriving the Poisson bracket algebra of the canonical generators of asymptotic symmetries, and has been particularly emphasized in J. D. Brown and M. Henneaux, "Central charges in the canonical realization of asymptotic symmetries: An example from three-dimensional gravity," to be published in Commun. Math. Phys. See also A. Ashtekar, "Asymptotic properties of isolated systems: Recent developments," in *General Relativity and Gravitation*, edited by B. Bertotti, F. de Felice, and A. Pascolini (Reidel, Dordrecht, 1984).

[4]Th. De Donder, "*Théorie Invariantive du Calcul des Variations* (Gauthier–Villars, Paris, 1935), 2nd. ed.

Remarks on the Yang–Mills equations in four dimensions: CP^2

C. Nash
Department of Mathematical Physics, St. Patrick's College, Maynooth, Ireland

(Received 27 August 1984; accepted for publication 3 October 1985)

This paper examines properties of solutions to the Yang–Mills equations in four dimensions and in particular on the manifold CP^2: Two solutions are found: one is neither self-dual nor anti-self-dual but is a solution of the full Yang–Mills equations, the other is a self-dual solution.

I. INTRODUCTION

The attention of both mathematicians and mathematical physicists in recent years has been drawn to Yang–Mills theories. There are deep results of a nontrivial mathematical kind, cf. Ref. 1; and, of course, physics is making an increasing use nowadays of Yang–Mills and Yang–Mills–Higgs theories. These theories are mathematically interesting in various numbers of dimensions, and physicists similarly vary the number of dimensions so as to be able to discuss vortices, monopoles, instantons, and Kaluza–Klein theories, to mention a few of the possibilities considered (cf. also the selective list of Refs. 2–6 and references therein). We shall limit ourselves in this paper to four dimensions.

In four dimensions there are extra things that can happen: there is the vital and fundamental relationship between an oriented four-manifold M and its associated twistor space Z, and also the fact that the Yang–Mills curvature F, being a two-form, has the possibility of being self-dual. The most studied case in four dimensions is that when $M = S^4$ (see Ref. 1) for which the corresponding twistor space Z is CP^3—the complex projective three-space. Another example that has received considerable attention[7] is the torus T^4, this being the manifold that is natural when periodic boundary conditions arise. In this paper we look at the manifold CP^2 complex projective two-space. This manifold is a two-dimensional *complex* manifold and so can be regarded as an orientable manifold of *real* dimension 4. An additional motivation of choosing $M = CP^2$ is that, as well as CP^2 being a well-known compact four-dimensional manifold, it has a twistor space Z, which is a Kähler manifold. Kähler manifolds are complex manifolds endowed with a metric that is Hermitian and possess the further invariance property that the Hermiticity operations are preserved under parallel transport. For an introduction to Kähler manifolds and physics, cf. Ref. 8. In this respect Z is a surprising rarity—the only other compact four-manifold M whose twistor space Z is Kählerian is S^4 (see Ref. 9).

On CP^2 we wish to study the G-invariant Yang–Mills action S where

$$S = \|F\|^2 = -\int_{CP^2} \mathrm{tr}(F \wedge {}^*F), \tag{1.1}$$

with G some compact Lie group, which in this paper is either SU(3) or SU(2), and the * operation on the Yang–Mills curvature is defined with respect to the so-called Fubini–Study metric on CP^2 (see Ref. 10). This means that if we have a two-form ω on CP^2 then

$$\omega = \omega_{AB}\, dx^A \wedge dx^B, \tag{1.2}$$

with $A,B = 1,\ldots,4$, then ${}^*\omega$ is given by

$$\omega = (\sqrt{\tilde{g}}/2)\epsilon_{AB}{}^{CD}\omega_{CD}\, dx^A \wedge dx^B, \tag{1.3}$$

and $\tilde{g}$ is the determinant of the Fubini–Study metric $\tilde{g}_{AB}$.

The organization of the rest of this paper is as follows: In the next section we give two extremal connections A and A' for S. The first of these connections A is su(3)-valued, while the second A' is su(2)-valued. It turns out that the curvature F of A is neither self-dual nor anti-self-dual, but satisfies the second-order Yang–Mills equation

$$D {}^*F = 0. \tag{1.4}$$

This solution has been independently found in CP^n context by Hogan.[11] However, neither the Chern class c_2 of A nor the action S were calculated there. Here we calculate c_2 and S. We shall show that c_2 is zero and that S is infinite. This means that A is actually a singular, noninstanton, solution of the Yang–Mills equations. Atiyah[12] has recently discussed the significance of some solutions to the Yang–Mills equations that possess singularities. The second connection A' has topological charge $k = 1$ and is self-dual with action S equal to $8\pi^2$. Finally the last section contains our concluding remarks.

II. THE YANG–MILLS EQUATIONS ON CP^2: THE CONNECTION A

As coordinates on CP^2 we use the homogeneous coordinates (z^0,z^1,z^2) and work for the most part on the patch $z^0 = 1$. The Fubini–Study metric on CP^2 that we use is denoted by $\tilde{g}_{AB}$ with $\tilde{g}_{AB}$ given in the block-off-diagonal form

$$\tilde{g}_{AB} = \begin{pmatrix} 0 & g_{a\bar{b}} \\ g_{\bar{a}b} & 0 \end{pmatrix}. \tag{2.1}$$

The capital letters A, B run over 1, 2, $\bar{1}$, $\bar{2}$, the lowercase letters a, b run over 1, 2, and the Hermitian matrix $g_{a\bar{b}}$ is given by

$$g_{a\bar{b}} = \frac{1}{2}\left[\frac{\delta^{ab}}{(1+z^e\bar{z}^e)} - \frac{\bar{z}^a z^b}{(1+z^e\bar{z}^e)^2}\right] = \partial_a\,\partial_{\bar{b}}\,\ln\sqrt{(1+z^e\bar{z}^e)}, \tag{2.2}$$

where $\partial_a \equiv \partial/\partial z^a$ and $\partial_{\bar{a}} \equiv \partial/\partial \bar{z}^a$.

Turning to the Yang–Mills action we have more explicitly

$$S = -\int_{CP^2} \mathrm{tr}(F \wedge {}^*F) = -\frac{1}{2}\int_{CP^2} \mathrm{tr}(F_{AB}F_{A'B'})\tilde{g}^{AA'}\tilde{g}^{BB'}\sqrt{\tilde{g}}\omega, \tag{2.3}$$

 0022-2488/86/020492-03$02.50 © 1986 American Institute of Physics

where ω is the volume form on CP^2 and

$$F=(F_{AB}/2)dz^A\wedge dz^B. \tag{2.4}$$

As we said in Sec. I the Euler–Lagrange equations for S are

$$D*F=0. \tag{2.5}$$

A solution to (2.5) is given by taking an ansatz of the form

$$A=f_1\,\partial F_1+f_2\,\bar{\partial}F_2, \tag{2.6}$$

where f_1 and f_2 are functions, F_1 and F_2 are matrices belonging to the Lie algebra su(3) of SU(3), both are homogeneous of degree zero in $z^0,\ldots,z^2$, and ∂F_1 and $\bar{\partial}F_2$ stand for the one-forms $(\partial F_1/\partial\eta^a)d\eta^a$ and $(\partial F_2/\partial\bar{\eta}^a)d\bar{\eta}^a$, respectively, with η^a, $\bar{\eta}^a$ local complex coordinates on CP^2. It turns out that we can satisfy the Yang–Mills equations for the case when f_1 and f_2 are both constant. To this end let α,β run from 0,...,2 and consider nine 3×3 matrices $T_{\alpha\beta}$, where $T_{\alpha\beta}$ is a matrix that is zero everywhere except at the (α,β) location, which has a nonzero entry equal to unity. Now if we write

$$\tilde{F}=\sum_{\alpha,\beta}f_{\alpha\beta}(z)\bar{z}^\alpha z^\beta T_{\alpha\beta}, \tag{2.7}$$

with $f_{\alpha\beta}$ a real function homogeneous of degree -2, then the choice

$$A=\partial\tilde{F}-\bar{\partial}\tilde{F} \tag{2.8}$$

gives a connection. A straightforward computation shows that this connection obeys the Yang–Mills equations (2.5) provided $f_{\alpha\beta}=(-1)^{\alpha+\beta}/\sqrt{1+\bar{z}^e z^e}$, where we now work in the patch $z^0=1$. Note that $A\in$su(3) although $\tilde{F}\notin$su(3). The connection (2.8) has a curvature F that is not self-dual or anti-self-dual. This can be seen directly from F itself but it will also be a consequence of the fact that we will find the action S and the second Chern class $c_2(F)$ to be different.

Of interest first is what is the value c_2 of $c_2(F)$, where

$$c_2=\frac{-1}{8\pi^2}\int_{CP}\mathrm{tr}(F\wedge F) \tag{2.9}$$

for su(3). The integrand $\mathrm{tr}(F\wedge F)$ is readily calculable and one obtains, with $f=(1+z^e\bar{z}^e)$,

$$\begin{aligned}\mathrm{tr}(F\wedge F)=&f^{-4}(\delta^{bc}+\bar{z}^b z^c)(\delta^{ef}+\bar{z}^e z^f)\\&\times(f\delta^{cg}-\bar{z}^c z^g)(f\delta^{fh}-\bar{z}^f z^h)\\&\times(2f^{-2}\bar{z}^a z^b\bar{z}^d z^e-f^{-1}\bar{z}^a z^b\delta^{de}\\&-f^{-1}\bar{z}^a z^e\delta^{bd}-f^{-1}\bar{z}^d z^e\delta^{ab}\\&-f^{-1}\bar{z}^d z^b\delta^{ae}+\delta^{ab}\delta^{de}+\delta^{ae}\delta^{bd})\\&\times d\bar{z}^g\wedge dz^a\wedge d\bar{z}^h\wedge dz^d,\end{aligned} \tag{2.10}$$

and this eventually simplifies to, with $G=f^{-2}(\bar{z}^a dz^a)\wedge(z^b d\bar{z}^b)$,

$$\mathrm{tr}(F\wedge F)=2G\wedge G, \tag{2.11}$$

and this is immediately zero. Thus c_2 is zero and the next question, since F is not self-dual, what is the action S? This means that we need to calculate

$$\mathrm{tr}(F\wedge *F)=\tfrac{1}{2}\mathrm{tr}(F_{AB}F_{CD})\tilde{g}^{AC}\tilde{g}^{BD}\sqrt{\tilde{g}}\,d^4z, \tag{2.12}$$

and d^4z the volume form on CP^2. In (2.12), $\tilde{g}$ is the determinant of the Fubini–Study metric $\tilde{g}_{AB}$ of (2.1) and $\tilde{g}^{AC}$ denotes the inverse of $\tilde{g}_{AC}$. This leads to $\tilde{g}$ being expressible in terms of g, the determinant of the 2×2 matrix $g_{a\bar{b}}$, via

$$\tilde{g}=g^2, \tag{2.13}$$

while $\tilde{g}^{AC}$ may be calculated directly or by using the explicit off-diagonal form (2.1) and the characteristic polynomial for $g_{a\bar{b}}$. With these facts one finds, after a lengthy computation, that

$$\begin{aligned}\mathrm{tr}(F\wedge *F)=&-2f^{-2}(2f^{-2}\bar{z}^a z^g\bar{z}^d z^h-f^{-1}\bar{z}^a z^g\delta^{dh}\\&-f^{-1}\bar{z}^a z^h\delta^{gd}-f^{-1}\bar{z}^d z^h\delta^{ag}\\&-f^{-1}\bar{z}^d z^g\delta^{ah}+\delta^{ag}\delta^{dh}+\delta^{ah}\delta^{gd})\\&\times g^{a\bar{h}}g^{\bar{g}d}g\,d^4z.\end{aligned} \tag{2.14}$$

If we introduce the notation $t=\mathrm{tr}\,g^{a\bar{b}}$, $h=\bar{z}^a z^b g^{a\bar{b}}$, $l=\bar{z}^a z^g g^{a\bar{d}}g^{\bar{g}d}$, and $m=\mathrm{tr}(g^{a\bar{b}}g^{\bar{a}b})$, we obtain

$$\begin{aligned}\mathrm{tr}(F\wedge *F)=&-2f^{-2}(2f^{-2}h^2-2f^{-1}l\\&-2f^{-1}ht+m+t^2)g\,d^4z.\end{aligned} \tag{2.15}$$

However, using the characteristic polynomial $p(\lambda)=\lambda^2-t\lambda-g$ with $\lambda=g_{a\bar{b}}$, it is possible to relate the functions l, m, and t and express them all as functions of f. We simply quote the result of doing all this. It is that (2.15) is equal to

$$-2\left(\frac{1}{f^5}+\frac{1}{f^3}+\frac{1}{f}-\frac{2}{f^5}(1+f^2)X+\frac{X^2}{f^5}\right)d^4z, \tag{2.16}$$

where $X=(\bar{z}^a z^a)(z^b\bar{z}^b)$.

It remains to integrate the expression in (2.16) over the manifold CP^2. In theory[13] this integration could require specifying a partition of unity subordinate to some covering $\{U_\alpha\}$ of CP^2. However in the case CP^2 may be covered by the three patches U_α specified by $z_\alpha\neq0$, respectively. Further each of the patches U_α is *dense* in CP^2 and hence we need only integrate over one of them—U_0 say. Hence

$$\begin{aligned}S=&-\int_{CP^2}\mathrm{tr}(F\wedge *F)\\=&2\int_{U_0}\left(\frac{1}{f^5}+\frac{1}{f^3}+\frac{1}{f}-\frac{2}{f^5}(1+f^2)X+\frac{X^2}{f^5}\right)d^4z.\end{aligned} \tag{2.17}$$

This means that S is reducible to a sum of four-dimensional integrals—the integrals involving X need some considerable attention. When this is given we obtain the formulas [$\Gamma(z)$ is the gamma function]

$$\begin{aligned}&\int_{U_0}\frac{X}{f^\alpha}d^4z\\&\quad=2^2\pi^2\frac{\Gamma(\alpha-6)(\alpha-6)(\alpha-5)}{\Gamma(\alpha-4)(\alpha-1)(\alpha-2)(\alpha-3)(\alpha-4)},\\&\int_{U_0}\frac{X^2}{f^\alpha}d^4z=2\cdot4!\pi^2\frac{\Gamma(\alpha-10)(\alpha-10)\cdots(\alpha-7)}{\Gamma(\alpha-8)(\alpha-8)\cdots(\alpha-1)}.\end{aligned} \tag{2.18}$$

For the action S we eventually find the expression $5\pi^2/12+S_I$, where S_I stands for the expression

$$\begin{aligned}\lim_{\alpha\to1}\Bigg[&2\pi^2\frac{\Gamma(\alpha-2)}{\Gamma(\alpha)}+\frac{2^2\cdot4!\pi^2}{(\alpha-2)(\alpha-1)\alpha\cdots(\alpha+3)}\\&-16\pi^2\frac{\Gamma(\alpha-2)}{\Gamma(\alpha+2)}\Bigg],\end{aligned} \tag{2.19}$$

which is, in fact, infinite so that S diverges. This infinity has

arisen on integrating over the patch $z_0 \neq 0$, where one can easily verify using (2.7) and (2.8) that A and F are regular. Since CP^2 is compact the infinity of S shows that the singularity is at $z_0 = 0$.

The connection A': We now take the group G to be SU(2) rather than SU(3). Our ansatz for A' is

$$A' = \partial\alpha - \bar{\partial}\alpha, \tag{2.20}$$

where in this case a is given by (α and β run from 0 to 1),

$$\alpha = \alpha_{\alpha\beta} T_{\alpha\beta},$$

and $\alpha_{\alpha\beta}$ is a homogeneous real-valued function given by $\alpha_{\alpha\beta} = \frac{1}{2}(-1)^{(\alpha+\beta)/2}\delta_{\alpha\beta} \ln(1 + z^e\bar{z}^e)$. The curvature F of A' is straightforward to evaluate and is given by expression

$$F = \left[\frac{dz^1 \wedge d\bar{z}^1 + dz^2 \wedge d\bar{z}^2}{(1 + z^e\bar{z}^e)} - \frac{z^1\bar{z}^1\, dz^1 \wedge d\bar{z}^1 + z^2\bar{z}^2 dz^2 \wedge d\bar{z}^2 + \bar{z}^1 z^2\, dz^1 \wedge d\bar{z}^2}{(1 + z^e\bar{z}^e)^2} - \frac{z^1\bar{z}^2\, dz^2 \wedge d\bar{z}^1}{(1 + z^e\bar{z}^e)^2} \right](T_{00} - T_{11}).$$

Using (1.3), we do a routine calculation of $^*F$, which verifies that F is self-dual so that $F = {}^*F$. Of interest next is the value of the topological charge k. Since $k = -c_2$ we have

$$k = \frac{1}{8\pi^2} \int_{CP^2} \mathrm{tr}(F \wedge F). \tag{2.21}$$

The techniques developed for the SU(3) solution suffice also to evaluate this expression and we obtain finally for k the integral

$$\frac{-1}{8\pi^2} \int_{U_0} \frac{4\omega}{(1 + z^e\bar{z}^e)^3},$$

where $\omega = dz^1 \wedge d\bar{z}^1 \wedge dz^2 \wedge d\bar{z}^2$. This integral is now straightforward to perform and we confirm that $k = 1$ as claimed. We also calculated the action S and find that $S = 8\pi^2$, as it should. Thus A' is a self-dual solution with $k = 1$.

III. CONCLUSIONS

We have carried out an investigation of the Yang–Mills equations in four-dimensions, and have considered solutions to the Yang–Mills equations for the groups SU(2) and SU(3). In particular we have shown that a solution A to the SU(3) equations on CP^2 endowed with the Fubini–Study metric is trivial topologically and possesses a singularity on CP^2. For the SU(2) case we have found a self-dual solution with unit topological charge. An interesting question also is the geometry and topology of the solution space on CP^2. A further point of interest in the study of CP^2 is that its twistor space[9] $Z = CF(3)$ and is one of the only two Kählerian twistor spaces possible. The number of non-instanton-like solutions to the Yang–Mills equations on CP^2 is an interesting topic[14]; in two dimensions all the topology of the space of solutions can be fully analyzed using an equivariant Morse theory,[15] but in four dimensions, where the criteria for applicability of the Morse theory are not met, the situation is open even for manifolds such as S^4. There are interesting calculational results,[16] not touched on here, for taking a step nearer to the quantum field theory; the significance of the solution discussed here may be in its relation to the properties of singular solutions as discussed by Atiyah,[12] or in its embedding in a space of higher dimension, these are points under further study.

ACKNOWLEDGMENT

This paper was written in part while the author was a visitor at CERN. The hospitality of the CERN theory division is gratefully acknowledged.

[1] M. F. Atiyah, *The Geometry of Yang–Mills Fields* (Leziane Fermiane, Pisa, 1979).

[2] A. Jaffe and C. H. Taubes, *Vortices and Monopoles* (Birkhauser, Boston, 1980).

[3] E. Corrigan and P. Goddard, Commun. Math. Phys. **80**, 575 (1981); P. Forgacs, Z. Horvath, and L. Palla, Phys. Lett. B **102**, 131 (1981); N. J. Hitchin, Commun. Math. Phys. **83**, 579 (1982); **89**, 145 (1983); M. K. Prasad, Commun. Math. Phys. **80**, 137 (1981); R. S. Ward, Commun. Math. Phys. **79**, 317 (1981).

[4] M. F. Atiyah, *The Geometry of Yang–Mills Fields* (Leziane Fermiane, Pisa, 1979); M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977); E. Corrigan, C. Devchand, D. B. Fairlie, and J. Nuyts, Nucl. Phys. B **214**, 452 (1983); E. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Phys. Lett. B **72**, 139 (1977); C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic, New York, 1983); R. S. Ward, Phys. Lett. A **61**, 81 (1977).

[5] E. Cremmer and J. Scherk, Nucl. Phys. B **103**, 399 (1976); E. Witten, Nucl. Phys. B **186**, 412 (1981); A. Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982).

[6] D. Olive and P. West, Nucl. Phys. B **217**, 248 (1983); K. Stelle, in *Proceedings of the International Conference on High Energy Physics*, Brighton, U.K., 1983, edited by J. Guy and C. Costain (Rutherford Appleton Laboratory, 1984).

[7] P. van Baal, Commun. Math. Phys. **85**, 529 (1982); Utrecht preprint, 1984; G. 't Hooft, Commun. Math. Phys. **81**, 267 (1980); M. Lüscher, Commun. Math. Phys. **85**, 39 (1982); C. Nash, Commun. Math. Phys. **88**, 319 (1983).

[8] T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980); E. J. Flaherty, *Hermitian and Kählerian Geometry in Relativity* (Springer, Berlin, 1976).

[9] N. J. Hitchin, Proc. Lond. Math. Soc. (3) **43**, 133 (1981). For reference we give here the twistor space Z of CP^2 it is the so-called flag manifold CF(3) of flags in C^3; the flag manifold CF(3) is six dimensional as is CP^3 the twistor space Z of S^4, $CF(3) = Gl(3,C)/\Delta(3,C)$; for more details on CF(3), cf. R. Wells, *Differential Analysis on Complex Manifolds* (Springer, Berlin, 1976).

[10] T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980); E. J. Flaherty, *Hermitian and Kählerian Geometry in Relativity* (Springer, Berlin, 1976); R. Wells, *Differential Analysis on Complex Manifolds* (Springer, Berlin, 1976).

[11] P. A. Hogan, J. Math. Phys. **24**, 674 (1983).

[12] M. F. Atiyah, Commun. Math. Phys. **93**, 437 (1984); M. F. Atiyah, *Proceedings of the International Colloquium on Vector Bundles* (Tata Institute, Bombay, 1984).

[13] For example, C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic, New York, 1983).

[14] C. H. Taubes, J. Diff. Geom. **17**, 139 (1982).

[15] M. F. Atiyah and R. Bott, Philos. Trans. R. Soc. London Ser. A **308**, 523 (1982); M. F. Atiyah and J. D. S. Jones, Commun. Math. Phys. **61**, 97 (1978).

[16] I. T. Drummond, S. Duane, and R. R. Horgan, Nucl. Phys. B **220**, 119 (1983); P. Goddard, D. Mansfield, and H. Osborn, Nucl. Phys. B **98**, 59 (1981); H. Osborn and G. P. Moody, Nucl. Phys. B **173**, 427 (1980).

Hamiltonian and non-Hamiltonian perturbation theory for nearly periodic motion

Jonas Larsson
Department of Plasma Physics, Umeå University, S-901 87 Umeå, Sweden

(Received 21 May 1985; accepted for publication 6 September 1985)

Kruskal's asymptotic theory of nearly period motion [M. Kruskal, J. Math. Phys. **4**, 806 (1962)] (with applications to nonlinear oscillators, guiding center motion, etc.) is generalized and modified. A new more natural recursive formula, with considerable advantages in applications, determining the averaging transformations and the drift equations is derived. Also almost quasiperiodic motion is considered. For a Hamiltonian system, a manifestly Hamiltonian extension of Kruskal's theory is given by means of the phase-space Lagrangian formulation of Hamiltonian mechanics. By performing an averaging transformation on the phase-space Lagrangian for the system ($L \to \bar{L}$) and adding a total derivative $dS/d\tau$, a nonoscillatory Lagrangian $\Lambda = \bar{L} + dS/d\tau$ is obtained. The drift equations and the adiabatic invariant are now obtained from Λ. By truncating Λ to some finite order in the small parameter ϵ, manifestly Hamiltonian approximating systems are obtained. The utility of the method for treating the guiding-center motion is demonstrated in a separate paper.

I. INTRODUCTION

A nearly periodic Hamiltonian system was shown by Kruskal to possess an all-order adiabatic invariant.[1] The theory may be used as a general method of finding the adiabatic invariants of various Hamiltonian nonlinear oscillators. It may also be applied to the motion of a gyrating charged particle in a strong magnetic field. It provides a systematic method for deriving to arbitrary order the (possibly) three adiabatic invariants—the magnetic moment, the longitudinal, and the flux invariants—and the associated drift equations. The method of Kruskal has, however, certain less satisfactory features that will be discussed later on in this section, and that motivates the present paper.

Following Kruskal we shall first consider nearly periodic but not necessarily Hamiltonian systems. Without the Hamiltonian property we do not obtain any adiabatic invariant, but we still derive the drift equations to arbitrary order, which determines the asymptotic solutions. Let us therefore consider a system of $m+1$ ordinary differential equations

$$\frac{d\mathbf{y}}{d\tau} = \mathbf{g}(\mathbf{y},\nu;\epsilon), \tag{1.1a}$$

$$\frac{d\nu}{d\tau} = \psi(\mathbf{y},\nu;\epsilon), \tag{1.1b}$$

where $\mathbf{y} \in R^m$, $\nu \in R$, and

$$\mathbf{g}(\mathbf{y},\nu;\epsilon) = \epsilon\mathbf{g}^{(1)}(\mathbf{y},\nu) + \epsilon^2\mathbf{g}^{(2)}(\mathbf{y},\nu) + \cdots, \tag{1.1c}$$

$$\psi(\mathbf{y},\nu;\epsilon) = \psi^{(0)}(\mathbf{y}) + \epsilon\psi^{(1)}(\mathbf{y},\nu) + \cdots \tag{1.1d}$$

are formal series in the small parameter ϵ. The functions are periodic in ν with period 2π. An ordinary differential equation with all solutions nearly periodic may, in appropriate coordinates, be written in the form (1.1).[1]

The generalization to the nearly *quasiperiodic* case is given in Appendix A. The *nonautonomous* case when $\mathbf{g}$ and ψ depend on $\epsilon\tau$ is considered in Appendix B.

We are interested in solving (1.1) for times of order $1/\epsilon$, i.e., for many periods of the rapid-angle variable ν. Direct iteration in (1.1) produces secular terms $\epsilon\tau$, $(\epsilon\tau)^2$, etc., and is not useful. Instead we may derive new equations approximating (1.1) with the rapid-angle variable averaged away. Straightforward averaging of (1.1) yields the equation

$$\dot{\mathbf{y}} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{g}(\mathbf{y},\nu;\epsilon)d\nu, \tag{1.2}$$

which turns out to be valid only to lowest order in an asymptotic expansion. In order to obtain an all-order theory we make use of an *averaging transformation* approach. The basic idea then is to find new coordinates such that all the ν dependence enters in the coordinate transformation leaving the transformed system independent of the rapid angle. Formally, a coordinate transformation means no approximation, but in our case, the transformation is a formal, usually not convergent, power series in ϵ. New constants of motion for the transformed equations may exist as formal series in ϵ and are called *adiabatic invariants*. The way we usually deal with these series makes the question of convergence less important; only terms of low order in ϵ are calculated, so the most important thing is that these few terms have good approximating properties.

Various methods for deriving the averaging transformations have been developed. When (1.1) are the canonical equations, we may use the Poincaré–von Zeipel method. By means of canonical averaging transformations we then eliminate the rapid-angle coordinate from Hamiltonian. Alternatively we may, without any Hamiltonian structure, find an averaging transformation directly from the dynamic equations (1.1). This is done by Kruskal[1] and by Krylov, Bogoliubov, and Mitropolsky.[2] An important contribution of Kruskal was the inclusion of an adiabatic invariant which exists when (1.1) is a Hamiltonian system.[3]

If (1.1) is a Hamiltonian system but $(\mathbf{y},\nu)$ are not canonical variables, the application of the Poincaré–von Zeipel method may be far from straightforward. For example, considering the guiding-center motion of a charged particle in an inhomogeneous magnetic field, this happens to be a non-

 0022-2488/86/020495-07$02.50 © 1986 American Institute of Physics

trivial problem.[4] The method of Kruskal may be used in place of canonical perturbation theory, but has the unsatisfactory property of producing non-Hamiltonian approximations of (1.1) to each finite order in ϵ when (1.1) is a Hamiltonian system. This problem was recently solved by Littlejohn by the development of two versions of noncanonical Hamiltonian perturbation theory. Both methods are basically generalizations of the Poincaré–von Zeipel method. In the first method,[4,5] the Poisson bracket formulation of noncanonical Hamiltonian mechanics is used and symplectic averaging transformations are used to eliminate the angle variable from the Hamiltonian. In the second method[6–8] arbitrary (nonsymplectic) averaging transformations is allowed. This is possible due to the very convenient phase-space Lagrangian formulation of noncanonical Hamiltonian mechanics (see Appendix C).[9] The averaging transformation is, in this case, used to eliminate the rapid-angle variable from the phase-space Lagrangian. Littlejohn makes use of Lie transforms to find the averaging transformations from the condition of a nonoscillating phase-space Lagrangian.[6] Remarkably nice looking formulas for the guiding-center motion have been derived in this manner.[7]

In the present paper we are going to improve and generalize the Kruskal method rather than the Poincaré–von Zeipel method. The former method has the advantages of a simple mathematical structure and an explicit, compact, and easily derived recursion relation determining the asymptotic expansion to all orders. Furthermore, non-Hamiltonian systems also may be considered. However, Kruskal's theory also has quite serious drawbacks, which makes it less attractive in applications. We have already mentioned that it produces non-Hamiltonian approximations to Hamiltonian systems. Furthermore, the recursion relation gives inconvenient new variables. It is probably no accident that Northrop, in his well-known book[10] on the guiding-center motion of a gyrating charged particle, prefers to use a WKB ansatz (also due to Kruskal) when deriving the guiding-center drift equations and only make use of Kruskal's general theory[1] for deriving the adiabatic invariants. The problem is evident in Ref. 11 where a straightforward application of Kruskal's recursion formula is seen not to give the usual lowest-order guiding-center position.

These problems with Kruskal's method are solved in the present paper.[12] The class of averaging transformations is quite large and we determine this class explicitly in order to make possible a favorable choice in each specific problem. Further, by considering the *inverse* averaging transformation in our recursion formulas (2.5), we avoid unnatural initial conditions like (B20) in Ref. 1. The relations (2.5) not only determine the whole class of averaging transformations, but the subclass that possesses certain gyrogauge invariance properties (see Sec. IV B) is also easily identified. The class of averaging transformations is determined by the functions $\langle \mathbf{Y} \rangle$ and $\langle \Upsilon \rangle$, both formal power series in ϵ, serving as free parameters. The gyrogauge invariance of the slow drift variable $\mathbf{z}$ is obtained if $\langle \mathbf{Y} \rangle$ is chosen in a "physical–geometric" way. These free parameters are in practice chosen, order by order, to simplify, for example, the transformed equations and the constants of motion. When dealing with a Hamiltonian system we try to get a simple formula for the phase-space Lagrangian. Compared with Kruskal's original approach, we have now achieved additional possibilities for obtaining convenient asymptotic formulas first by a more natural form of the recursion relation, and second by carefully displaying the free parameters, which are chosen in the end to make the results simple.

The phase-space Lagrangian formulation of Hamiltonian mechanics is a perfect tool for developing a noncanonical, but still manifestly Hamiltonian, perturbation theory from Kruskal's theory. In terms of the phase-space Lagrangian L for (1.1) and an averaging transformation determined by (2.5) we get an explicit expression for a nonoscillating phase-space Lagrangian Λ. The adiabatic invariant and the drift equations are now obtained directly from Euler's equations. We get Hamiltonian approximating systems by truncating Λ to finite order in ϵ. The adiabatic invariant may be used to reduce the order of the system by 1, and since we already have eliminated the rapid-angle variable, we have a reduced system two orders lower than (1.1). Kruskal proved that this reduced system is also Hamiltonian.[1] This result follows much more easily with the present method, but more important is that we now also get finite-order Hamiltonian approximations of the reduced system.

It is straightforward to obtain a systematic all-order theory for the guiding-center motion of a charged particle by means of this method. We have investigated[13] how the free parameters may be chosen in order to simplify the form of the guiding-center Lagrangian. In particular the algebraic results in Ref. 7 are confirmed apart from a claimed gyrogauge invariance property of the guiding-center Lagrangian. What is of interest, we believe, is the gyrogauge invariance of the slow guiding-center variables. However, the requirement of this invariance property probably forces the all-order Lagrangian to take a more complicated form than in Ref. 7.

The new recursion relation determining all averaging transformations is given in Sec. II. The manifestly Hamiltonian extension of Kruskal's theory is presented in Sec. III, and some invariance properties are considered in Sec. IV. In Sec. V we summarize the procedure of obtaining Λ, and also give a direct method of deriving Λ from L without explicit use of (1.1). This alternative is close to Littlejohn's second method but it is mathematically more elementary and Lie transforms are used. Section V also includes a discussion comparing the methods for obtaining Λ.

II. THE AVERAGING TRANSFORMATIONS

We are going to determine the class of near-identity coordinate transformations $(\mathbf{y},\nu) \to (\mathbf{z},\phi)$ such that (1.1) in terms of $(\mathbf{z},\phi)$ transforms into

$$\dot{\mathbf{z}} = \mathbf{h}(\mathbf{z};\epsilon), \quad \dot{\phi} = \omega(\mathbf{z};\epsilon), \tag{2.1a}$$

$$\mathbf{h}(\mathbf{z};\epsilon) = \epsilon \mathbf{h}^{(1)}(\mathbf{z}) + \cdots, \tag{2.1b}$$

$$\omega(\mathbf{z};\epsilon) = \omega^{(0)}(\mathbf{z}) + \epsilon\omega^{(1)}(\mathbf{z}) + \cdots, \tag{2.1c}$$

where we note that $\mathbf{h}$ and ω, to all orders, are independent of the rapid-angle variable ϕ. These transformations are called "averaging transformations" and $(\mathbf{z},\phi)$ are called "nice variables." Let us write an inverse averaging transformation

$$\mathbf{y} = \mathbf{z} + \mathbf{Y}(\mathbf{z},\phi;\epsilon),\ \nu = \phi + \Upsilon(\mathbf{z},\phi;\epsilon), \tag{2.2a}$$

where

$$\mathbf{Y}(\mathbf{z},\phi;\epsilon) = \epsilon\mathbf{Y}^{(1)}(\mathbf{z},\phi) + \cdots, \tag{2.2b}$$

$$\Upsilon(\mathbf{z},\phi;\epsilon) = \epsilon\Upsilon^{(1)}(\mathbf{z},\phi) + \cdots. \tag{2.2c}$$

Here $\mathbf{Y}$ and Υ are periodic with period 2π in ϕ.

Let the function $F(\mathbf{y},\nu)$ be periodic in ν with period 2π. We define the operations ∇, ∂, $\sim$, $\hat{}$, $\mathbf{I}$, and j by[14]

$$(\nabla F)(\mathbf{y},\nu) = \frac{\partial F(\mathbf{y},\nu)}{\partial \mathbf{y}}, \tag{2.3a}$$

$$(\partial F)(\mathbf{y},\nu) = \frac{\partial F(\mathbf{y},\nu)}{\partial \nu}, \tag{2.3b}$$

$$\langle F\rangle(\mathbf{y}) = \frac{1}{2\pi}\int_0^{2\pi} F(\mathbf{y},\nu)d\nu, \tag{2.3c}$$

$$\tilde{F} = F - \langle F\rangle, \tag{2.3d}$$

$$\hat{F}(\mathbf{y},\nu) = (\psi^{(0)}(\mathbf{y}))^{-1}\left(\int_0^{\nu}\tilde{F}(\mathbf{y},\nu')d\nu'\right)^{\sim}, \tag{2.3e}$$

$$\mathbf{I}(\mathbf{y},\nu) = \mathbf{y},\ j(\mathbf{y},\nu) = \nu. \tag{2.3f}$$

The inverse averaging transformation may now be written $(\mathbf{I}+\mathbf{Y}, j+\Upsilon)$. We use the symbol $\circ$ to denote composition of functions. We have, for example,

$$\mathbf{g}(\mathbf{z}+\mathbf{Y}(\mathbf{z},\phi;\epsilon),\phi+\Upsilon(\mathbf{z},\phi;\epsilon);\epsilon) = \mathbf{g}\circ(\mathbf{I}+\mathbf{Y}, j+\Upsilon)(\mathbf{z},\phi;\epsilon). \tag{2.4}$$

The formal series $\langle\mathbf{Y}\rangle(\mathbf{z};\epsilon)$ and $\langle\Upsilon\rangle(\mathbf{z};\epsilon)$ are now regarded as free parameters. To each choice of these we have $\mathbf{Y}$, Υ, $\mathbf{h}$, and ω determined recursively, order by order in ϵ, by means of the relations

$$\mathbf{h} + \psi^{(0)}\partial\tilde{\mathbf{Y}} = \mathbf{g}\circ(\mathbf{I}+\mathbf{Y}, j+\Upsilon) - \mathbf{h}\cdot\nabla\mathbf{Y} - (\omega-\psi^{(0)})\partial\mathbf{Y}, \tag{2.5a}$$

$$\omega + \psi^{(0)}\partial\tilde{\Upsilon} = \psi\circ(\mathbf{I}+\mathbf{Y}, j+\Upsilon) - \mathbf{h}\cdot\nabla\Upsilon - (\omega-\psi^{(0)})\partial\Upsilon. \tag{2.5b}$$

Furthermore, every averaging transformation may be obtained in this way by correspondingly choosing $\langle\mathbf{Y}\rangle$ and $\langle\Upsilon\rangle$.

The proof of these results concerning (2.5) is merely a matter of simple checking which is conveniently performed in the following steps.

(a) Given $\mathbf{g}^{(k)}$, $\psi^{(k)}$, $\langle\mathbf{Y}^{(k)}\rangle$, and $\langle\Upsilon^{(k)}\rangle$ for $k = 0,\ldots,n$, then Eqs. (2.5) determine $\mathbf{h}^{(n)}$, $\omega^{(n)}$, $\mathbf{Y}^{(n)}$, and $\Upsilon^{(n)}$.

(b) Let $\mathbf{h}$, ω, $\mathbf{Y}$, and Υ satisfy (2.5). Then (1.1) transforms into (2.1) by means of the coordinate transformation $(\mathbf{y},\nu)\to(\mathbf{z},\phi)$ defined by (2.2).

(c) If (2.2) is an inverse averaging transformation so that Eqs. (1.1) transform into Eqs. (2.1) then the relations (2.5) follow.

From (a) and (b) it is clear that (2.5) produces averaging transformations and corresponding transformed equations in the wanted fashion. From (c) it then follows that all averaging transformations are obtained in this way.

Let us now prove (a), (b), and (c). We use induction for (a). First take $n = 0$. Substituting $\mathbf{g}^{(0)} = 0$, $\psi^{(0)}(\mathbf{y},\nu) = \psi^{(0)}(\mathbf{y})$, $\langle\mathbf{y}^{(0)}\rangle = 0$, and $\langle\Upsilon^{(0)}\rangle = 0$ in the ϵ^0 part of (2.5) gives $\mathbf{h}^{(0)} = 0$, $\omega^{(0)} = \psi^{(0)}$, $\mathbf{Y}^{(0)} = 0$, and $\Upsilon^{(0)} = 0$. Assume (a) is true for $n = p-1$. Consider the ϵ^p terms in (2.5)

$$\mathbf{h}^{(p)} + \psi^{(0)}\partial\tilde{\mathbf{Y}}^{(p)} = \mathbf{g}^{(p)} + \mathbf{R}^{(p)}, \tag{2.6a}$$

$$\omega^{(p)} + \psi^{(0)}\partial\tilde{\Upsilon}^{(p)} = \psi^{(p)} + \mathbf{Y}^{(p)}\cdot\nabla\psi^{(0)} + r^{(p)}, \tag{2.6b}$$

where $\mathbf{R}^{(p)}$ and $r^{(p)}$ are known by the assumption above. Now given $\mathbf{g}^{(p)}$, $\psi^{(p)}$, $\langle\mathbf{Y}^{(p)}\rangle$, and $\langle\Upsilon^{(p)}\rangle$ we want to show that (2.6) determines $\mathbf{h}^{(p)}$, $\psi^{(p)}$, $\mathbf{Y}^{(p)}$, and $\Upsilon^{(p)}$. Then (a) is proved for $n = p$ and is, by induction, true for all n. Now $\tilde{\mathbf{Y}}^{(p)}$ is the only unknown quantity on the right-hand side of (2.6). From (2.6a) we obtain

$$\tilde{\mathbf{Y}}^{(p)} = \hat{\mathbf{g}}^{(p)} + \hat{\mathbf{R}}^{(p)}, \tag{2.7}$$

$$\mathbf{h}^{(p)} = \langle\mathbf{g}^{(p)}\rangle + \langle\mathbf{R}^{(p)}\rangle, \tag{2.8}$$

and from (2.6b)

$$\tilde{\Upsilon}^{(p)} = \hat{\psi}^{(p)} + \hat{\mathbf{Y}}^{(p)}\cdot\nabla\psi^{(0)} + \hat{r}^{(p)}, \tag{2.9}$$

$$\omega^{(p)} = \langle\psi^{(p)}\rangle + \langle\mathbf{Y}^{(p)}\rangle\cdot\nabla\psi^{(0)} + \langle r^{(p)}\rangle. \tag{2.10}$$

Thus $\mathbf{Y}^{(p)}$, $\mathbf{h}^{(p)}$, and $\omega^{(p)}$ are determined and so is $\Upsilon^{(p)}$ after substitution of (2.7) in (2.9). To demonstrate (b) we substitute (2.2) in (1.1) and eliminate the $\mathbf{g}$ and ψ terms by means of (2.5). Then, order by order, (2.1) follows. The demonstration of (c) is completely straightforward.

III. HAMILTONIAN PERTURBATION THEORY

We now consider the case when (1.1) is a Hamiltonian system but $(\mathbf{y},\nu)$ are not necessarily canonical coordinates. Then Eqs. (1.1) are the Euler equations for a phase-space Lagrangian function

$$L(\mathbf{y},\nu,\dot{\mathbf{y}},\dot{\nu};\epsilon) = \dot{\mathbf{y}}\cdot\gamma(\mathbf{y},\nu;\epsilon) + \dot{\nu}\eta(\mathbf{y},\nu;\epsilon) - H(\mathbf{y},\nu;\epsilon), \tag{3.1}$$

as shown in Appendix C. Let us now transform (3.1) by (2.2) and write the phase-space Lagrangian in $(\mathbf{z},\phi)$ coordinates as [see (4.32)]

$$\bar{L}(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon) = \dot{\mathbf{z}}\cdot\bar{\gamma}(\mathbf{z},\phi;\epsilon) + \dot{\phi}\bar{\eta}(\mathbf{z},\phi;\epsilon) - \bar{H}(\mathbf{z},\phi;\epsilon). \tag{3.2}$$

The Euler equations of (3.2) are (2.1) which are ϕ-independent. However, $\bar{L}$ is in general ϕ-dependent, but we prove in Appendix E that a rapid-angle independent phase-space Lagrangian Λ may be defined by

$$\Lambda = \bar{L} + \frac{d}{d\tau}(S + S_1), \tag{3.3a}$$

$$S_1(\mathbf{z},\phi;\epsilon) = -\left[\int_0^{\phi}(\bar{\eta}(\mathbf{z},\phi'))^{\sim}d\phi'\right]^{\sim}, \tag{3.3b}$$

where $S(\mathbf{z};\epsilon)$ is a free parameter. We note that Λ contains the free parameters $\langle\mathbf{Y}\rangle$, $\langle\Upsilon\rangle$, and S which are to be chosen as conveniently as possible. Of course Λ and $\bar{L}$ have the same Euler equations (2.1). A useful expression for Λ is

$$\Lambda = \langle\bar{L}\rangle + \frac{dS}{d\tau}. \tag{3.4}$$

The averaging in (3.4),

$$\langle\bar{L}\rangle(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon) = \frac{1}{2\pi}\int_0^{2\pi}\bar{L}(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon)d\phi, \tag{3.5}$$

simplifies the algebra in applications by eliminating a lot of terms. From (3.2), (3.3a), and (3.4) we note that

$$\bar{H} = \langle\bar{H}\rangle. \tag{3.6}$$

Let us write

$$\Lambda(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon)=\Gamma(\mathbf{z};\epsilon)\cdot\dot{\mathbf{z}}+J(\mathbf{z};\epsilon)\dot{\phi}-K(\mathbf{z};\epsilon). \tag{3.7}$$

The Euler equations of Λ give the drift equations to all orders. By truncating the series $\Lambda=\Lambda^{(0)}+\epsilon\Lambda^{(1)}+\cdots$, we obtain Hamiltonian approximations of these equations to any desired order. We also find the adiabatic invariant $J(\mathbf{z};\epsilon)$; $\dot{J}=0$ to all orders in ϵ as a consequence of $\partial\Lambda/\partial\phi=0$ in Euler's equations.

Let us now consider the reduced system obtained by eliminating ϕ and J. We then first change coordinates, $(\mathbf{z},\phi)\rightarrow(\mathbf{w},J,\phi)$, so that J becomes one of the independent variables. To simplify the notations, we assume that $(\mathbf{z},\phi)$ in (3.7) already has been chosen so that $z_m=J$ and we denote $\mathbf{z}=(\mathbf{w},J)$. The phase-space Lagrangian for the reduced system may now be written

$$\Lambda_R(\mathbf{w},\dot{\mathbf{w}};J,\epsilon)=\sum_{i=1}^{m-1}\Gamma_i(\mathbf{w};J,\epsilon)\dot{w}_i-K(\mathbf{w};J,\epsilon), \tag{3.8}$$

where $\mathbf{w},\dot{\mathbf{w}}\epsilon R^{m-1}$. Since here J also is considered to be a parameter, we now write J after the semicolon in (3.8). The Hamiltonian property of the reduced system thus follows almost directly by the phase-space Lagrangian formulation of Hamiltonian mechanics. The proof of this fact in Ref. 1 was quite complicated. Furthermore, we now obtain Hamiltonian approximating systems by truncating the series $\Lambda_R=\Lambda_R^{(0)}+\epsilon\Lambda_R^{(1)}+\cdots$.

IV. GAUGE INDEPENDENCE

We consider the nonuniqueness in the choices of coordinates and of phase-space Lagrangians and investigate some invariance properties. Let us, like Kruskal,[1] consider an ordinary differential equation

$$\dot{x}=F(x;\epsilon),\quad x\epsilon R^{m+1}, \tag{4.1a}$$

$$F(x;\epsilon)=F^{(0)}(x)+\epsilon F^{(1)}(x)+\cdots, \tag{4.1b}$$

such that

$$\dot{x}=F^{(0)}(x) \tag{4.2}$$

has all solutions periodic. There exists a coordinate transformation $x\rightarrow(\mathbf{y},\nu)$ so that (4.1) transforms into (1.1).[1] We take (4.1) to be our physical–geometrical equation which we express in terms of the nonuniquely determined coordinates $(\mathbf{y},\nu)$ and $(\mathbf{z},\phi)$. We have coordinate transformations

$$C\colon R^{m+1}\rightarrow R^{m+1},\quad x\rightarrow(\mathbf{y},\nu), \tag{4.3}$$

where C depends on some free parameters which we do not display explicitly. The averaging transformations A,

$$A\colon R^{m+1}\rightarrow R^{m+1},\quad(\mathbf{y},\nu)\rightarrow(\mathbf{z},\phi), \tag{4.4a}$$

$$A^{-1}=(\mathbf{I}+\mathbf{Y},j+\Upsilon), \tag{4.4b}$$

where $\mathbf{Y}$ and Υ are determined recursively by (2.5), depend on the free parameters $\langle\mathbf{Y}\rangle$ and $\langle\Upsilon\rangle$. When (4.1) is a Hamiltonian system we write the phase-space Lagrangian

$$\mathscr{L}(x,\dot{x};\epsilon)=\mathscr{G}(x,\dot{x};\epsilon)\cdot\dot{x}-\mathscr{H}(x;\epsilon). \tag{4.5}$$

Without changing the Euler equations we may add the total derivative of an arbitrary function $\mathscr{S}(x;\epsilon)$ to $\mathscr{L}$, we write

$$\mathscr{L}\cong\mathscr{L}+\frac{d}{d\tau}\mathscr{S}. \tag{4.6}$$

For the system (1.1) we choose the phase-space Lagrangian L obtained from $\mathscr{L}$ by the coordinate transformation C. We write

$$\mathscr{L}\overset{C}{\rightarrow}L. \tag{4.7}$$

The averaging transformation A yields

$$L\overset{A}{\rightarrow}\overline{L}, \tag{4.8}$$

where $\overline{L}$ is related to the ϕ-independent phase-space Lagrangian Λ by (3.4).

A. The adiabatic invariant

The adiabatic invariant $\mathscr{J}(x;\epsilon)$ of a Hamiltonian system (4.1) may be written

$$\mathscr{J}=J\circ A\circ C, \tag{4.9}$$

where $J(\mathbf{z};\epsilon)$ is determined by (3.7). In order to show that $\mathscr{J}$ is independent of the free parameters in A and C we introduce a second choice of these indicated by primed quantities

$$\mathscr{J}'=J'\circ A'\circ C', \tag{4.10}$$

and we show that

$$\mathscr{J}(x;\epsilon)=\mathscr{J}'(x;\epsilon). \tag{4.11}$$

Define B by

$$B=A'\circ C'\circ C^{-1}\circ A^{-1}, \tag{4.12}$$

so that B is the transformation between the unprimed and primed sets of nice variables

$$B(\mathbf{z},\phi;\epsilon)=(\mathbf{z}',\phi'). \tag{4.13}$$

Then B is of the form[1]

$$B(\mathbf{z},\phi;\epsilon)=(\mathbf{B}(\mathbf{z};\epsilon),\phi+\beta(\mathbf{z};\epsilon)). \tag{4.14}$$

From

$$\overline{L}\overset{B}{\rightarrow}\overline{L}' \tag{4.15}$$

and (4.14) we obtain

$$\langle\overline{L}\rangle\overset{B}{\rightarrow}\langle\overline{L}'\rangle. \tag{4.16}$$

From (3.4) we note that $\langle\overline{L}\rangle$ and Λ have the same $\dot{\phi}$ term and according to (3.7) it is $J\dot{\phi}$. This fact together with (4.16) and (4.14) gives

$$J=J'\circ B, \tag{4.17}$$

and (4.11) follows.

B. Gyrogauge invariance of the z-variable

The transformation

$$A\circ C\colon x\rightarrow(\mathbf{z},\phi) \tag{4.18}$$

depends on the free parameters of A and C. We shall demonstrate an invariance property of the $\mathbf{z}$ variable, i.e., of the mapping $\mathscr{Z}$,

$$\mathscr{Z}=\mathbf{I}\circ A\circ C,\quad\mathscr{Z}(x;\epsilon)=\mathbf{z}, \tag{4.19}$$

when we restrict the freedom of choosing A and C. Let there be some natural choice for the y-variable, i.e., we are given a function $\mathscr{Y}\colon R^{m+1}\rightarrow R^m$ and consider only coordinate transformations C such that

$$\mathbf{I}\circ C = \mathcal{I}. \tag{4.20}$$

We then choose the free parameters $\langle \mathbf{Y} \rangle$ of A independently of C, i.e., we choose a function $\mathbf{k} = \epsilon \mathbf{k}^{(1)} + \epsilon^2 \mathbf{k}^{(2)} + \cdots, \mathbf{k}^{(n)}$: $R^m \to R^m$ and take

$$\langle \mathbf{Y} \rangle = \mathbf{k}. \tag{4.21}$$

The conditions (4.20) and (4.21) restrict the allowed choices of C and A. The remaining freedom is associated with the choice of rapid-angle variable. We will show $\mathcal{I} = \mathbf{I}\circ A \circ C$ is uniquely determined by (4.20) and (4.21) and thus independent of the choice of rapid-angle variable, we call this property "gyrogauge invariance."[7]

Let us, as in Sec. IV A above, introduce primed variables A', C', etc., associated with an alternative choice of free parameters, still requiring (4.20) and (4.21). We obtain

$$\mathbf{I}\circ C = \mathbf{I}\circ C', \tag{4.22}$$

$$\langle \mathbf{Y} \rangle = \langle \mathbf{Y}' \rangle, \tag{4.23}$$

and want to prove that these imply

$$\mathcal{I} = \mathcal{I}', \text{ i.e., } \mathbf{I}\circ A \circ C = \mathbf{I}\circ A' \circ C'. \tag{4.24}$$

From (4.12) and (4.22) we have

$$\mathbf{I}\circ A^{-1} = \mathbf{I}\circ (A')^{-1} \circ B; \tag{4.25}$$

inserting (4.4b), we obtain

$$\mathbf{I} + \mathbf{Y} = (\mathbf{I} + \mathbf{Y}')\circ B, \tag{4.26}$$

and then averaging with use of (4.14)

$$\mathbf{I} + \langle \mathbf{Y} \rangle = (\mathbf{I} + \langle \mathbf{Y}' \rangle)\circ \mathbf{B}. \tag{4.27}$$

Inserting $\langle \mathbf{Y} \rangle = \langle \mathbf{Y}' \rangle = \mathbf{k}$ yields

$$\mathbf{I} + \mathbf{k} = (\mathbf{I} + \mathbf{k})\circ \mathbf{B}. \tag{4.28}$$

An order by order calculation now gives

$$\mathbf{B} = \mathbf{I} \tag{4.29}$$

and (4.24) follows.

C. Independence of gauge in $\mathcal{L}$

Let $u \in R^n$ and $a = (a_i), A = (A_i)$ be mappings from R^n to R^n. Then we can prove the identity

$$\begin{aligned} A(u + a(u)) \cdot \frac{d}{d\tau}(u + a(u)) &= A(u) \cdot \dot{u} - \sum_{i,j=1}^{n} \int_0^1 a_i(u) B_{ij}(u + \alpha a(u)) \\ &\times \frac{d}{d\tau}(u_j + \alpha a_j(u)) d\alpha \\ &+ \frac{d}{d\tau} \int_0^1 A(u + \alpha a(u)) \cdot a(u) d\alpha, \end{aligned} \tag{4.30}$$

where B_{ij} is the antisymmetric derivative of A, i.e., in usual index notation

$$B_{ij} = A_{i,j} - A_{j,i}. \tag{4.31}$$

This identity is useful for eliminating certain gauge-dependent terms in the perturbation series. The phase-space Lagrangian L for the system (1.1) may contain gauge-dependent terms $A(u) \cdot \dot{u}$, where for notational convenience we have denoted $u = (\mathbf{y}, \nu)$, such that the antisymmetric derivative of A is gauge independent. These terms reflect the arbitrariness in $\mathcal{L}$ and L due to (4.6). The construction of the rapid-angle independent phase-space Lagrangian Λ involves $\overline{L}$

$$\begin{aligned} \overline{L}(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon) &= L(\mathbf{z} + \mathbf{Y}(\mathbf{z},\phi),\ \phi + \Upsilon(\mathbf{z},\phi), \\ &\quad \dot{\mathbf{z}} + \dot{\mathbf{z}} \cdot \nabla \mathbf{Y} + \dot{\phi} \partial \mathbf{Y},\ \dot{\phi} + \dot{\phi}\, \partial \Upsilon + \dot{\mathbf{z}} \cdot \nabla \Upsilon; \epsilon). \end{aligned} \tag{4.32}$$

Straightforward Taylor expansion of the right-hand side around $(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi})$ would produce a lot of gauge-dependent terms involving derivatives of A. This problem is avoided by first applying the identity (4.30). Note that the $d/d\tau$ part of (4.30) may be neglected without changing the dynamics.

The equality (4.30) may be rewritten as

$$\begin{aligned} &\int_0^1 \frac{d}{d\alpha}\left[A(u + \alpha a(u)) \cdot \frac{d}{d\tau}(u + \alpha a(u))\right] d\alpha \\ &\quad - \frac{d}{d\tau}\int_0^1 A(u + \alpha a(u)) \cdot a(u) d\alpha \\ &\quad = -\sum_{i,j=1}^{n} \int_0^1 a_i(u) B_{ij}(u + \alpha a(u)) \\ &\qquad \times \frac{d}{d\tau}(u_j + \alpha a_j(u)) d\alpha. \end{aligned} \tag{4.33}$$

This equality is easily proved by performing the α and τ derivations on the left-hand side of (4.33).

V. SUMMARY AND A DIRECT METHOD OF FINDING Λ

In this section we first recapitulate, step by step, the procedure of constructing Λ given in Secs. I–IV. We include the method in Sec. IV C of avoiding gauge-dependent terms in the perturbation series. We then suggest an alternative method of deriving Λ from L without explicit use of (1.1) and (2.5). Finally the two methods are compared and discussed.

Let us now consider the derivation of Λ.

(a) The phase space Lagrangian $L(\mathbf{y},\nu,\dot{\mathbf{y}},\dot{\nu};\epsilon)$ has (1.1) as Euler's equations. We substitute (2.2) in L and obtain $\overline{L}(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon)$.

(b) If $\overline{L}$ contains gauge-dependent terms, they may be eliminated as explained in Sec. IV C. We get

$$\overline{\overline{L}} = \overline{L} + \frac{dS_3}{d\tau}, \tag{5.1}$$

where $S_3(\mathbf{z},\phi;\epsilon)$ is obtained from the last term in (4.30).

(c) Let $\overline{\eta}(\mathbf{z},\phi;\epsilon)\dot{\phi}$ be the $\dot{\phi}$ term in $\overline{L}$. We eliminate the oscillating part of it by the transformation

$$\Lambda = \overline{\overline{L}} + \frac{dS_1}{d\tau} + \frac{dS}{d\tau}, \tag{5.2a}$$

where

$$S_1(\mathbf{z},\phi;\epsilon) = -\left[\int_0^{\phi} (\overline{\eta}(\mathbf{z},\phi;\epsilon))^{\sim} d\phi'\right]^{\sim}, \tag{5.2b}$$

and $S(\mathbf{z};\epsilon)$ is a free parameter.

(d) Define $\mathbf{Y}$ and Υ by the recursion relation (2.5). Then, as shown in Appendix D, Λ is ϕ-independent. We thus have

$$\Lambda = \langle \overline{\overline{L}} \rangle + \frac{dS}{d\tau}, \tag{5.3}$$

which is a more convenient formula than (5.2)

(e) The so-constructed Λ contains the free parameters

$\langle \mathbf{Y} \rangle$, $\langle \Upsilon \rangle$, and S. This freedom may be used to simplify the expressions for Λ and the constants of motion. We may obtain a gyrogauge-invariant $\mathbf{z}$ variable, by means of a physical–geometrical choice of $\langle \mathbf{Y} \rangle$ as explained in Sec. IV B.

We now present an alternative method of deriving Λ without explicit use of (1.1) and (2.5). The steps (a), (b), (c), and (e) above are unaltered, but (d) is replaced by the following.

(d′) Require Λ to be ϕ-independent. This gives constraints on $\mathbf{Y}$ and Υ since Λ depends on them. Then $\tilde{\mathbf{Y}}^{(n)}$ and $\tilde{\Upsilon}^{(n)}$ must be uniquely determined for $n = 1$ while for $n \geqslant 2$ they may depend on the free parameters $\langle \mathbf{Y}^{(p)} \rangle$, and $\langle \Upsilon^{(p)} \rangle$ for $p \leqslant n-1$.

This direct method is not a satisfactory theory by itself, but depends on the results obtained by the explicit use of (1.1) and (2.5). It follows from these results that if Eqs. (1.1) are Euler's equations for L, then $\mathbf{Y}$ and Υ must be determined by (d′). We might hope that the direct method would avoid some unnecessary algebra since the use of (2.5) requires the explicit calculation of (1.1) and of the functions h and ω. We have rederived Littlejohn's results for the guiding-center motion valid to one order higher than usual with both methods.[15] The price we have to pay for a rather moderate decrease in the amount of algebra needed in the direct method is a considerably less straightforward calculation. While the use of (1.1) and (2.5) requires only straightforward algebra and the whole structure with free parameters is clear from the outset, we must, with the direct method, work very carefully so that each choice of terms we make in $\mathbf{Y}$ and Υ is forced by (d′).

Littlejohn has developed a direct method of deriving the averaging transformations by requiring Λ to be nonoscillating. He makes use of the technique of Lie transforms. The result for the guiding-center motion was, however, given without derivations, and from the published work on his method it seems to be difficult to compare it with the other two methods above.

APPENDIX A: ALMOST QUASIPERIODIC SYSTEMS

The generalization to quasiperiodic systems is straightforward. We then replace (1.1) by a system of $m + k$ ordinary differential equations where k is the number of rapid-angle variables. We make the replacements $\nu \to \boldsymbol{\nu}$ and $\psi \to \boldsymbol{\psi}$ in (1.1) and require $\mathbf{g}$ and $\boldsymbol{\psi}$ to be periodic with period 2π in each of the k-angle variables. All relations in Sec. II may be used with minor modifications. We make the obvious replacements $\omega \to \boldsymbol{\omega}$, $\partial \to \boldsymbol{\partial}$, $\phi \to \boldsymbol{\phi}$, $j \to \mathbf{j}$, where $\mathbf{j}(\mathbf{y},\boldsymbol{\nu}) = \boldsymbol{\nu}$ and

$$\langle F \rangle(\mathbf{y}) = \frac{1}{(2\pi)^k} \int_0^{2\pi} d\nu_1 \cdots \int_0^{2\pi} F(\mathbf{y},\boldsymbol{\nu}) d\nu_k, \tag{A1}$$

$$\hat{F}(\mathbf{y},\boldsymbol{\nu}) = [(\boldsymbol{\psi}^{(0)}(\mathbf{y}) \cdot \boldsymbol{\partial})^{-1} \tilde{F}(\mathbf{y},\boldsymbol{\nu})]^{\sim}. \tag{A2}$$

We may express $\hat{F}$ in terms of F's Fourier components $F_{\mathbf{n}}(\mathbf{y})$, where $\mathbf{n} \in \mathbf{Z}^k$ and $\mathbf{Z}$ is the set of all real integers. We have

$$F(\mathbf{y},\boldsymbol{\nu}) = \sum_{\mathbf{n} \in \mathbf{Z}^k} F_{\mathbf{n}}(\mathbf{y}) \exp(i\mathbf{n} \cdot \boldsymbol{\nu}), \tag{A3}$$

$$\hat{F}(\mathbf{y},\boldsymbol{\nu}) = -i \sum_{\mathbf{n} \in \mathbf{Z}^k - 0} \frac{F_{\mathbf{n}}(\mathbf{y})}{\mathbf{n} \cdot \boldsymbol{\psi}^{(0)}(\mathbf{y})} \exp(i\mathbf{n} \cdot \boldsymbol{\nu}). \tag{A4}$$

In (A4) the problem of small denominators appears, this will not be discussed in the present paper.

The recursion relation (2.5) may be used with the minor changes in notation introduced above.

APPENDIX B: NONAUTONOMOUS SYSTEMS

The nonautonomous case when $\mathbf{g}$ and ψ in (1.1) are weakly dependent on τ is easy to include in the theory. We will treat this τ dependence as a parameter, avoiding the obvious but less practical alternative of increasing the order of the system by 1 and treat τ as a variable on the same footing as $(\mathbf{y},\nu)$. In place of (1.1) we have

$$\frac{d\mathbf{y}}{d\tau} = \mathbf{g}(\mathbf{y},\nu;\epsilon\tau,\epsilon), \tag{B1a}$$

$$\frac{d\nu}{d\tau} = \psi(\mathbf{y},\nu;\epsilon\tau,\epsilon), \tag{B1b}$$

$$\mathbf{g}(\mathbf{y},\nu;\epsilon\tau,\epsilon) = \epsilon\mathbf{g}^{(1)}(\mathbf{y},\nu;\epsilon\tau) + \cdots, \tag{B1c}$$

$$\psi(\mathbf{y},\nu;\epsilon\tau,\epsilon) = \psi^{(0)}(\mathbf{y};\epsilon\tau) + \epsilon\psi^{(1)}(\mathbf{y},\nu;\epsilon\tau) + \cdots. \tag{B1d}$$

The recursion relation (2.5) is replaced by

$$\mathbf{h} + \psi^{(0)}\,\partial\tilde{\mathbf{Y}} = \mathbf{g}^{\circ}(\mathbf{I} + \mathbf{Y}, j + \Upsilon) - \mathbf{h} \cdot \nabla\mathbf{Y} - (\omega - \psi^{(0)})\partial\mathbf{Y} - \epsilon\frac{\partial\mathbf{Y}}{\partial(\epsilon\tau)}, \tag{B2a}$$

$$\omega + \psi^{(0)}\,\partial\tilde{\Upsilon} = \psi^{\circ}(\mathbf{I} + \mathbf{Y}, j + \Upsilon) - \mathbf{h} \cdot \nabla\Upsilon - (\omega - \psi^{(0)})\partial\Upsilon - \epsilon\frac{\partial\Upsilon}{\partial(\epsilon\tau)}. \tag{B2b}$$

The transformed Eq. (2.1) and the inverse averaging transformation (2.2) are changed only by the inclusion of the parameter $\epsilon\tau$ in $\mathbf{Y}$, Υ, $\mathbf{h}$, and ω just as in (B1).

APPENDIX C: PHASE-SPACE LAGRANGIAN FORMULATION OF HAMILTONIAN MECHANICS

Let us consider a Lagrangian function L of the form

$$L(x,\dot{x};\tau) = \gamma_i(x;\tau)\dot{x}_i - H(x;\tau), \tag{C1}$$

where $x = (x_1,\ldots,x_n)$ and we sum over the index i, $1 \leqslant i \leqslant n$, in (C1). Such a Lagrangian is associated with first- rather than second-order systems. The Euler equations are

$$\left(\frac{\partial\gamma_i}{\partial x_j} - \frac{\partial\gamma_j}{\partial x_i}\right)\dot{x}_j = \frac{-\partial H}{\partial x_i}. \tag{C2a}$$

If

$$\det\left(\frac{\partial\gamma_i}{\partial x_j} - \frac{\partial\gamma_j}{\partial x_i}\right) \neq 0, \tag{C2b}$$

then (C2a) may be written in the standard form

$$\dot{x} = f(x;\tau). \tag{C3}$$

A Lagrangian (C1), which satisfies (C2b), is said to be *nondegenerate*.

The class of equations (C3) that are derivable from a Lagrangian function of the form (C1) are precisely the class of Hamiltonian systems.[3] Thus we get the phase space Lagrangian formulation of Hamiltonian mechanisms.

It is observed in the standard textbooks on classical mechanisms that the Lagrangian L defined by

$$L(q,p,\dot{q},\dot{p};\tau) = p \cdot \dot{q} - H(q,p;\tau) \tag{C4}$$

has the canonical equations as its Euler equations. Thus the class of systems (C3) obtained from Lagrangians of the form (C1) clearly includes all Hamiltonian systems. Conversely, we also have to prove that if (C3) is derived from (C1), then (C3) is a Hamiltonian system. The proof is a simple application of Darboux's theorem. Consider the first-order differential form

$$\gamma = \gamma_i \, dx_i. \tag{C5}$$

Now $d\gamma$ is a closed nondegenerate two-form and so, according to Darboux's theorem, n is an even number and there exist new coordinates $q,p \in R^{n/2}$ such that $d\gamma \to \Sigma_{i=1}^{n/2} dp_i \wedge dq_i$ and thus

$$\gamma \to \sum_{i=1}^{n/2} p_i \, dq_i + dS(q,p;\tau), \tag{C6}$$

for some function S and

$$L \to \sum_{i=1}^{n/2} p_i q_i + \frac{dS}{d\tau} - \left(K + \frac{\partial S}{\partial \tau}\right), \tag{C7}$$

where $H \to K$. Here the arrows $\to$ stand for "transform into when we make the coordinate change $x \to (q,p)$." The Euler equations of (C7) are Hamilton's equations with the Hamiltonian $K + \partial S/\partial \tau$.

APPENDIX D: PROOF THAT Λ IS ϕ-INDEPENDENT

The phase-space Lagrangian Λ defined by (3.3) may be written

$$\Lambda(\mathbf{z},\phi,\dot{\mathbf{z}},\dot{\phi};\epsilon) = \Gamma(\mathbf{z},\phi;\epsilon)\cdot\dot{\mathbf{z}} + J(\mathbf{z};\epsilon)\dot{\phi} - K(\mathbf{z},\phi;\epsilon), \tag{D1}$$

and has the Euler equations (2.1). We note, by the construction of Λ, that J in (D1) is ϕ-independent. We will now show that Γ and K are also independent of ϕ so that Λ may be written as in (3.7). The $((d/d\tau)(\partial/\partial\dot{\phi}) - (\partial/\partial\phi))$ component of Euler's equations for (D1) together with (2.1a) implies

$$\tilde{K} = \mathbf{h}\cdot\tilde{\Gamma}. \tag{D2}$$

Then from $((d/d\tau)(\partial/\partial\dot{\mathbf{z}}) - (\partial/\partial\mathbf{z}))$ part of Euler's equations together with (2.1) and (D2) we obtain the following formal differential equation for $\tilde{\Gamma}$:

$$\frac{d}{d\tau}\tilde{\Gamma} = -\nabla\mathbf{h}\cdot\tilde{\Gamma}. \tag{D3}$$

The coefficient on the right-hand side in (D3) is $O(\epsilon)$. A direct order by order calculation or the application of Kruskal's theorem of phase independence[1] gives $\tilde{\Gamma} = 0$. From (D2) we note that $\tilde{K} = 0$, also, and the proof is complete.

[1] M. Kruskal, J. Math. Phys. **4**, 806 (1962).
[2] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach, New York, 1961).
[3] We call $\dot{x} = f(x)$ a Hamiltonian system, if by a change of coordinates $x \to (q,p)$ it may be written as Hamilton's canonical equations.
[4] R. G. Littlejohn, "Hamiltonian theory of guiding-center motion," Ph.D. thesis, LBL-12942, University of California, 1980.
[5] R. G. Littlejohn, J. Math. Phys. **20**, 2445 (1979).
[6] R. G. Littlejohn, J. Math. Phys. **23**, 742 (1982).
[7] R. G. Littlejohn, J. Plasma Phys. **29**, 111 (1983).
[8] R. G. Littlejohn, Ann. Phys. NY **151**, 1 (1983).
[9] Or equivalently by use of the fundamental one-form. We prefer, however, to use the (for most physicists) familiar language of Lagrangians rather than the differential geometric language of forms.
[10] T. G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience, New York, 1963).
[11] T. G. Northrop, C. S. Liu, and M. Kruskal, Phys. Fluids **9**, 1504 (1966).
[12] Some results of the present paper are summarized in *Proceedings and Contributed Papers of the International Conference on Plasma Physics*, Lausanne, 1984 (CRPP-EPFL, Lausanne, 1984), Vol. 2, p. 240.
[13] J. Larsson, "The guiding center Lagrangian," submitted to Phys. Scr.
[14] Note that F stands for a function with domain in R^{m+1} ∇ is the m-dimensional gradient operator on the first m variables, and ∂ is the derivative with respect to the $(m+1)$th variable. The notation $(\mathbf{y},\nu)$ in (2.3) for a point in R^{m+1} is arbitrary, and we could equally well have used $(\mathbf{z},\phi)$. We never assume that the transformation (2.2) has been performed without explicitly writing it out: $F(\mathbf{z},\phi) \neq F^{\circ}(\mathbf{I}+\mathbf{Y}, j+\Upsilon)(\mathbf{z},\phi)$.
[15] E. Näslund has performed the calculations with this direct method.

Stability of forced nonlinear oscillators via Poincaré map

K. L. Liu and K. Young
Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

(Received 8 May 1985; accepted for publication 20 September 1985)

The behavior of nonlinear oscillators $x(t)$ driven by a periodic external force is completely determined by the corresponding Poincaré map, which loses stability only in certain well-known ways. These translate into different classes of perturbations $\xi(t)$ of $x(t)$ that must be considered. By choosing simple representatives in each class, the stability of approximate solutions can be studied analytically. The Duffing equation is considered as an example. An extra island of stability is predicted for a range of driving forces and this is confirmed by numerical computation.

I. INTRODUCTION

The stability of the asymptotic oscillatory solutions of externally driven nonlinear oscillators $x(t)$ has been extensively studied. Although it is straightforward to obtain an approximate solution $x(t)$, e.g., by suitably truncating the higher harmonics of the response to the external force, the issue of stability against arbitrary small perturbations $x(t) \to x(t) + \xi(t)$ is more complicated. A recently developed technique well suited to the question of stability is a variational formulation,[1-4] in which a trial function with slowly varying coefficients $A(t)$, $B(t)$,... is put into the variational integral, and evolution equations are obtained for $A(t)$, $B(t)$,... . The possibility of subharmonic response has been treated in a similar fashion.[5] Such analytic studies are complemented with a host of numerical[6,7] and experimental[8] results for various systems.

However, these oscillators are described by a nonholonomic ordinary differential equation (ODE) with a number of parameters α_i. Numerical computation or experiments can yield a wealth of information at specific values of α_i, but is less useful in characterizing the global properties in the space of α_i. In particular, unstable solutions to the ODE, though inaccessible to numerical computation or experiment, are nevertheless important. For example, the collision of an unstable solution with a stable one causes the latter to disappear; two stable solutions in disjoint regions of parameter space may be linked by an unstable one. Thus the approximate analytic study of nonlinear oscillators remains useful.

On the other hand, the analytic studies[1-5] have in the main considered only instabilities caused by the growth of slowly varying coefficients. It must, however, be realized that such instabilities are only one of several possible types, and a more systematic investigation of other instabilities would be desirable. Our starting point is the observation that for systems that are holonomic except for an external force of period T, the Poincaré map $\{x(nT), \dot{x}(nT)\} \to \{x(nT+T), \dot{x}(nT+T)\}$ is sufficient to characterize the stability and indeed the entire solution. This is, of course, the principle of stroboscopic sampling that has been used successfully to reveal fine structure in numerical computations.[7] The possible mechanisms for the loss of stability of maps have recently become well cataloged[9]; these mechanisms translate into different classes of perturbations $\xi(t)$ that must be considered. In the main part of this paper (Sec. III), we show how stability may be tested with respect to a judiciously chosen $\xi(t)$ within each class, giving a fairly accurate analytic global picture. Unexpected features may also be revealed in certain regions of parameter space.

These remarks will be illustrated with the Duffing oscillator

$$\ddot{x} = -\gamma\dot{x} - x + 4x^3 + F\cos\omega t \tag{1.1}$$

and the results are presented in terms of the response curve showing

$$R = 6(\langle x^2\rangle - \langle x\rangle^2)$$

vs ω^2 for fixed γ and F, where $\langle\ \rangle$ denotes the average over a period, from $t = nT$ to $(n+1)T$, $T = 2\pi/\omega$. The following well-known properties[1,2,5,6,8] are reproduced by the stability analysis.

(a) The response curve for small F is the usual resonance type, skewing as F increases.

(b) In the nonlinear region, the response curve breaks into two disjoint parts, and if ω^2 is varied adiabatically, the response jumps from one to the other, showing hysteresis. The low-frequency part first loses parity invariance and then becomes chaotic via the period-doubling scenario.[6]

Our analysis predicts a third possibility, which is confirmed by numerical computation.

(c) In a certain small range of F, the response curve contains an extra isolated portion which occupies only a small region in the ω^2-R plane and is not reached from the other parts by adiabatic variation of the driving frequency. Thus this novel feature might not have been discovered without the analytic treatment as a guide.

In all cases, connection between the disjoint parts of the response curve is provided by unstable solutions, giving a global view which complements numerical results. It is suggested that the method outlined in this paper, based on recently gained knowledge on maps, will be used for the qualitative study of the stability of other externally driven nonlinear oscillators.

II. APPROXIMATE SOLUTION

A. Symmetric solution

Symmetry of the Duffing equation under parity inversion $x \to -x$, $t \to t + T/2$ would suggest that a solution be sought with only odd harmonics. For not too large driving forces, the third (fifth,...) harmonic is important only when $\omega^2 \approx 1/3^2$ ($1/5^2$,...), so for $\omega^2 \gtrsim 0.2$, it will be reasonable to

 0022-2488/86/020502-05$02.50 © 1986 American Institute of Physics

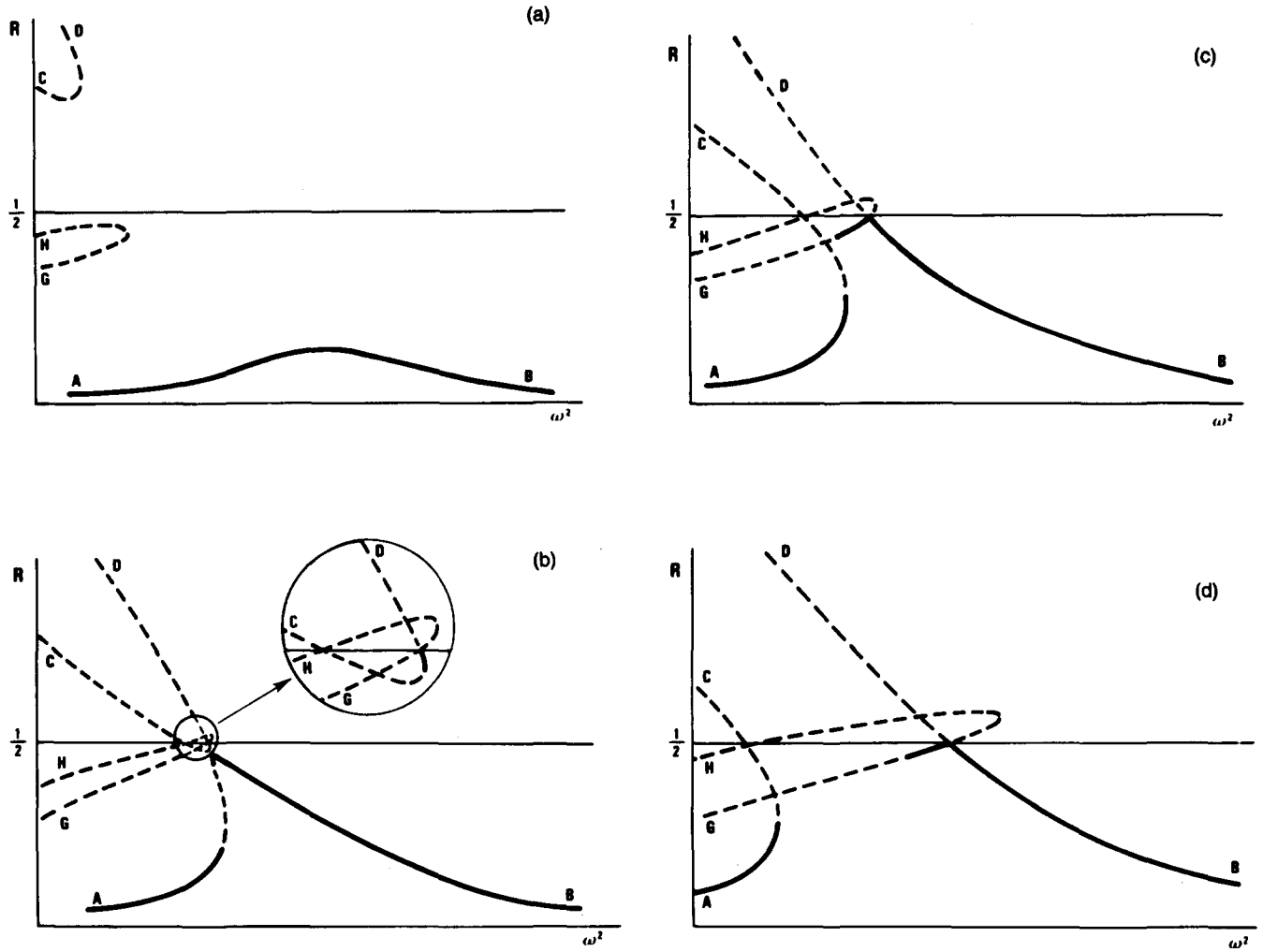


FIG. 1. Schematic relation between R and ω^2 based on approximate algebraic analysis. Solid line represents stable solution; broken line represents unphysical or unstable solution. The four figures (a)–(d) refer to increasing values of F.

substitute

$$x(t) = a\cos\omega t + b\sin\omega t \tag{2.1}$$

into the Duffing equation and drop higher harmonics, giving an algebraic equation for the response R:

$$R\,[(1-\omega^2-R)^2+\gamma^2\omega^2] = 3F^2\,, \tag{2.2}$$

where in this case $R = 3(a^2+b^2)$. The underdamped case ($\gamma<2$) is more interesting and we set $\gamma = 0.4$ everywhere below.[5,6] The graphs of R vs ω^2 are shown schematically in Fig. 1 for increasing F. (The curves GH refer to the asymmetric solution; see below. Broken curves refer to unstable solutions; for stability see Sec. III.) For small F, the four extremes are joined along AB, CD, the former being the usual resonance curve [Fig. 1(a)]. As F increases, CD intersects GH at $R = \frac{1}{2}$ and AB skews, R becoming multivalued in ω^2 [Fig. 1(b)]. Next the curves AB, CD touch and reconnect along AC, BD [Fig. 1(c)], with AC shifting to smaller ω^2 [Fig. 1(d)] and finally disappearing to negative ω^2 with increasing F. These properties are all well known, but the general impression seems to be that CD [e.g., in Fig. 1(b)] is unstable.[5,10] This is not necessarily the case, as we shall see.

B. Asymmetric solution

The parity invariance of the equation need not be shared by the solution. Ignoring second and higher harmonics, we may approximate the asymmetric solution as

$$x(t) = e + a'\cos\omega t + b'\sin\omega t, \tag{2.3}$$

which leads to

$$R\,[(2+\omega^2-5R)^2+\gamma^2\omega^2] = 3F^2\,, \tag{2.4}$$

$$e^2 = \tfrac{1}{4}(1-2R)\,. \tag{2.5}$$

The $\omega^2 - R$ graphs are shown by the curves GH in Fig. 1. First, the region $R > \frac{1}{2}$ is unphysical on account of (2.5) and is shown by dotted lines. Second, the even and odd solutions intersect at $R = \frac{1}{2}$, as is evident from (2.2) and (2.4).

III. STABILITY

A. Relation to Poincaré map

The main point of this present paper is to utilize the recently gained understanding of iterated maps. Consider the Poincaré map M:

$$M:\quad \mathbf{v} = \{x(nT),\dot{x}(nT)\} \to \mathbf{v}' = \{x(nT+T),\dot{x}(nT+T)\}\,. \tag{3.1}$$

A solution $x(t)$ of the ODE with period T corresponds to a fixed point $\mathbf{v}^*$ of M. If $x(t)$ is perturbed to $x(t)+\xi(t)$, then correspondingly $\mathbf{v}=\mathbf{v}^*+\Delta\mathbf{v}_n$, where

$$\Delta\mathbf{v}_n=\{\xi(nT),\dot{\xi}(nT)\}\,. \tag{3.2}$$

The fixed point $\mathbf{v}^*$ can lose stability in one of three ways.

(i) A tangent bifurcation is characterized by an eigenvalue $\lambda=1+\epsilon$, $0<\epsilon\ll 1$, so that $\Delta\mathbf{v}_n\sim(1+\epsilon)^n\Delta\mathbf{v}_0$, and thus $\xi(nT)\sim\xi(0)(1+\epsilon)^n$. In other words, $\xi(t)$ is nearly periodic:

$$\xi(t)=\sum c_k e^{ik\omega t}+\text{c.c.},$$

where the c_k are slowly varying coefficients. We may further divide ξ into odd and even parts, with the following representatives labeled as tangent odd (TO) and tangent even (TE):

$$\text{TO: } \xi(t)=ce^{i\omega t}+c^*e^{-i\omega t}, \tag{3.3}$$

$$\text{TE: } \xi(t)=c\,. \tag{3.4}$$

(ii) A period-doubling bifurcation (PD) is characterized by an eigenvalue $\lambda=-(1+\epsilon)$, $0<\epsilon\ll 1$, so that $\Delta\mathbf{v}_n\sim[-(1+\epsilon)]^n\Delta\mathbf{v}_0$ and $\xi(nT)\sim(-1)^n(1+\epsilon)^n\xi(0)$. It is seen that $\xi(t)$ can be written as

$$\xi(t)=\sum c_k e^{i(k+1/2)\omega t}+\text{c.c.}\,,$$

with slowly varying c_k. A simple representative is

$$\text{PD: } \xi(t)=ce^{i\omega t/2}+c^*e^{-i\omega t/2}\,. \tag{3.5}$$

(iii) A Hopf bifurcation (H) is characterized by a pair of complex eigenvalues $\lambda=(1+\epsilon)e^{\pm i\theta}$, $0<\epsilon\ll 1$, $\theta\neq 0,\pi,\ldots$. One of the eigenvalues will give

$$\Delta\mathbf{v}_n\sim(1+\epsilon)^n R(\theta)^n\Delta\mathbf{v}_0\,,$$

where $R(\theta)$ is the rotation matrix of angle θ. Thus

$$\xi(nT)\sim(1+\epsilon)^n[\cos(n\Omega t)\xi(0)+\sin(n\Omega t)\dot{\xi}(0)]\,,$$

where $\Omega=\theta/T\neq 0,\omega/2,\ldots$. Thus we may put

$$\xi(t)=\sum c_k e^{i(k\omega+\Omega)t}+\text{c.c.},$$

with c_k slowly varying, and a simple representative is

$$\text{H: } \xi(t)=ce^{i\Omega t}+c^*e^{-i\Omega t}\,. \tag{3.6}$$

The idea is to test the approximate solutions against the perturbations (3.3)–(3.6). Instead of studying the growth of a *function* $\xi(t)$, we only have to study the growth of a *number* c.

B. TO perturbation of symmetric solution

Change $x\to x+\xi$ with ξ as in (3.3). Put this into the Duffing equation, keeping only the terms that are linear in ξ and that have the same frequency as ξ itself:

$$\ddot{c}+2i\omega\dot{c}-\omega^2 c=-c-\gamma(\dot{c}+i\omega c)+6(a^2+b^2)c+3(a-bi)^2c^*\,, \tag{3.7}$$

with a, b given by (2.1) and (2.2). The growth rate $\alpha=\dot{c}/c$ satisfies

$$(\alpha^2+\gamma\alpha-\omega^2+1-2R)^2+(\gamma+2\alpha)^2\omega^2=R^2\,, \tag{3.8}$$

and the onset of instability occurs at $\alpha=0$,

$$(1-\omega^2-2R)^2+\gamma^2\omega^2=R^2\,,$$

which coincides with the condition $dR/d\omega^2=\infty$ obtained from (2.2). A little algebra shows that the three branches (e.g., A, C, D) are, in increasing R, respectively stable, unstable, and stable.

In fact under the TO perturbation, the original solution $a\cos\omega t+b\sin\omega t$ remains of the same type, but with a, b now slowly varying. Such adiabatic perturbations have always been considered in the literature[1–5,10]; however, the other perturbations described below have not been considered systematically.

C. TE perturbation of symmetric solution

Using ξ in (3.4), we find

$$\ddot{c}+\gamma\dot{c}=-(1-2R)c\,, \tag{3.9}$$

so that

$$\alpha=\tfrac{1}{2}\{-\gamma\pm[\gamma^2-4(1-2R)]^{1/2}\}\,.$$

If $R>\frac{1}{2}$, one of the roots is positive and the symmetric solution becomes unstable to TE perturbations. The important dividing line $R=\frac{1}{2}$ is shown in Fig. 1.

In fact, under a TE perturbation, the solution becomes $a\cos\omega t+b\sin\omega t+e$, i.e., it loses parity invariance. Thus it is no surprise that the TE instability occurs exactly at the point where the asymmetric solution crosses the symmetric solution.

D. H perturbation of symmetric solution

Use ξ as in (3.5) to get

$$\ddot{c}+2i\Omega\dot{c}-\Omega^2 c=-c-\gamma(\dot{c}+i\Omega c)+6(a^2+b^2)c\,, \tag{3.10}$$

leading to

$$\alpha^2+2i\alpha\Omega-\Omega^2=-1-\gamma\alpha-i\gamma\Omega+2R\,. \tag{3.11}$$

Since α is real (otherwise it amounts to a redefinition of Ω), (3.11) gives

$$\alpha^2-\Omega^2=-1-\gamma\alpha+2R\,, \tag{3.12a}$$

$$2\alpha\Omega=-\gamma\Omega\,, \tag{3.12b}$$

giving $\alpha=-\gamma/2$, which means all H perturbations are damped.

E. PD perturbation of symmetric solution

The linear equation satisfied by ξ is

$$\ddot{\xi}+\gamma\dot{\xi}+\xi-12x^2\xi=0\,, \tag{3.13}$$

and we assume ξ to be given by (3.5). In the case of the symmetric solution, x^2 has only even harmonics, so the $e^{\pm i\omega t/2}$ terms in ξ are not coupled. (Contrast the case of the asymmetric solution below.) Then the equation for the growth rate α is the same as (3.12), but with $\Omega=\omega/2$, so the symmetric solution is stable against period-doubling.

F. TO/E perturbation of asymmetric solution

Since the asymmetric solution already contains both even and odd harmonics, one would not expect the eigenvectors $\xi(t)$ to be purely even or odd, so the TO and TE perturbations should be combined. Such a perturbation in fact corresponds to (2.3) with e, a', b' allowed to vary slowly. An

algebraic equation is obtained for the eigenvalue $\alpha(=\dot{e}/e=\dot{a}'/a'=\dot{b}'/b')$ and the onset of instability ($\alpha=0$) occurs at the point where $dR/d\omega^2=\infty$. It is easily shown that the lower branch (G) is stable while the upper branch (H) is unstable.

G. H perturbation of asymmetric solution

In analogy to (3.12), we get $\alpha=-\gamma/2$, so all such perturbations are damped.

H. PD perturbation of asymmetric solution

For the asymmetric solution, x^2 in (3.13) contains $e^{\pm i\omega t}$ terms, which would couple the c and c^* terms in ξ. Thus in this case the PD perturbation is not obtained just by setting $\Omega=\omega/2$ in the equations for Hopf bifurcation. Rather

$$\ddot{c}+i\omega\dot{c}-\frac{\omega^2}{4}c=-c-\gamma\left(\dot{c}+i\frac{\omega}{2}c\right)+(3-4R)c+12e(a-bi)c^*, \tag{3.14}$$

where the last term is the coupling discussed. Putting $\dot{c}=\alpha c$ gives

$$\left[\alpha^2+1-\frac{\omega^2}{4}+\gamma\alpha-(3-4R)\right]^2+(\gamma+2\alpha)^2\frac{\omega^2}{4}=12R(1-2R). \tag{3.15}$$

We immediately see that for $R=\frac{1}{2}$ there is no real solution, so the asymmetric solution is stable. Suppose period-doubling occurs (i.e., $\alpha=0$) when $R=\frac{1}{2}-\epsilon$, then to $O(\epsilon)$, we find

$$\epsilon=(\omega^2/48)(\gamma^2+\omega^2/4)\approx 3\times 10^{-3},$$

if $\gamma=0.4$, $\omega^2\approx 0.5$. In other words, the asymmetric solution is created at the intersection with the symmetric solution at $R=\frac{1}{2}$, and loses stability due to period-doubling a very short distance away.

Once period-doubling has occurred, we should seek approximate solutions of the form[5]

$$x(t)=e+a'\cos\omega t+b'\sin\omega t+g'\cos\frac{\omega}{2}t+h'\sin\frac{\omega}{2}t$$

(where the corresponding vector v would be a fixed point of M^2) and test its stability with respect to further period-doubling, say represented by the perturbation

$$\xi(t)=ce^{i\omega t/4}+\text{c.c.}$$

However, the analogous situation in the case of maps has been extensively studied and one knows that a period-doubling cascade develops and ends in chaos along a route exhibiting a rich pattern of regularities.[9,11,12]

I. Global picture

Putting the above together yields the following picture as F increases. For small F, only the lower branch AB of the symmetric solution is stable [Fig. 1(a)]. As F increases, an island of stability appears [Fig. 1(b)], consisting of a symmetric solution (on curve D) which becomes asymmetric with decreasing ω^2 (on the curve G) and finally undergoes a period-doubling sequence to chaos. In the situation shown in Fig. 1(b), it may be expected that as ω^2 is decreased, the chaotic attractor may intersect the unstable solution C, resulting in a crisis[13] that drives the solution to the lower branch A. This state of affairs is confirmed by numerical computation at $F=0.110$ using 100 predictor–corrector steps per cycle (Fig. 2).

Upon further increase in F, the curve AB skews and develops vertical tangents, with the middle portion of AB

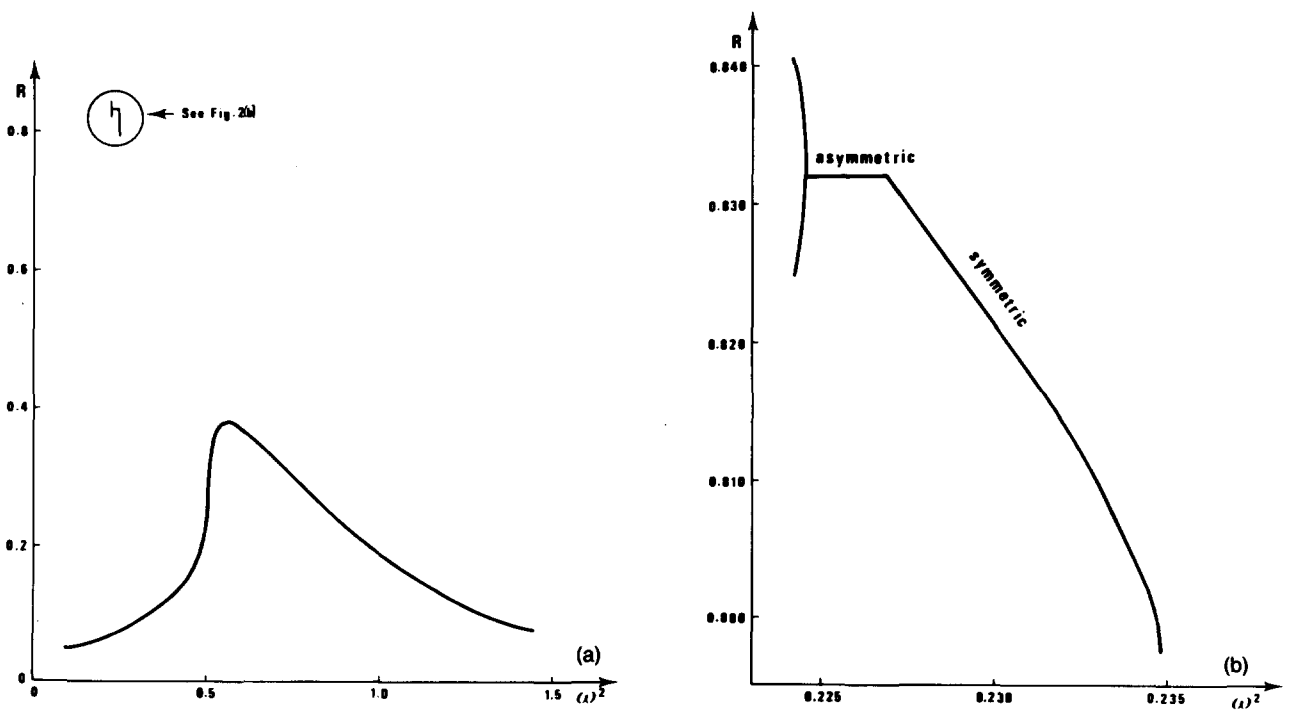


FIG. 2. (a) Actual relation between R and ω^2 obtained by numerical solution of ODE at $F=0.110$. (b) The island of stability enlarged.

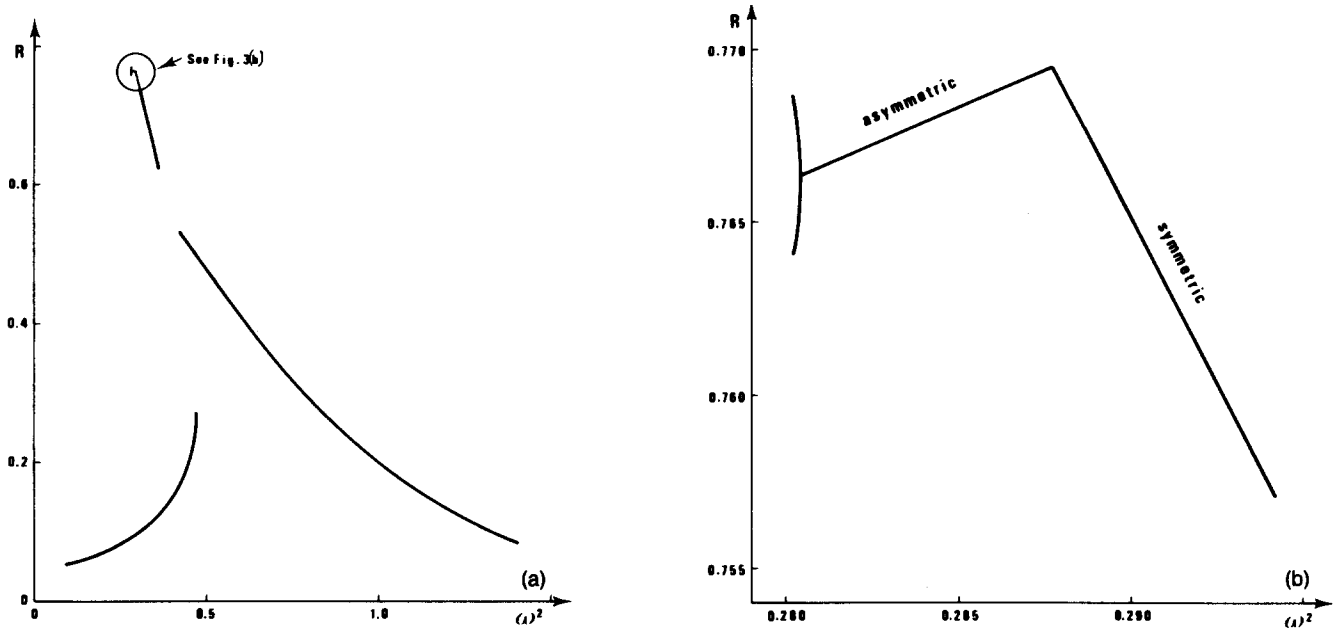


FIG. 3. (a) Actual relation between R and ω^2 obtained by numerical solution of ODE at $F = 0.115$. (b) The island of stability enlarged.

being unstable; the island of stability remains [Fig. 1(b)]. In this situation, the response curve should consist of three disjoint parts, as is confirmed by numerical computation at $F = 0.115$ (Fig. 3). The features in Fig. 2 and Fig. 3 would have been difficult to foresee or locate without the approximate stability analysis. At even larger F, CD and AB touch and rejoin. The high-frequency part of the symmetric solution (B) joins directly onto the asymmetric solution (G), which period-doubles to chaos [Fig. 1(c)].[6]

As F increases further, AC recedes to the left [Fig. 1(d)]. Then the chaotic attractor born out of the period-doubling sequence has no opportunity for a crisis going down to branch A, rather a crisis may develop when the chaotic attractor intersects the unstable solution D and becomes driven to infinity. The branch AC eventually disappears.

All these features are confirmed by numerical simulation, though the transitions do not necessarily occur at the precise values predicted.

IV. DISCUSSION

The stability analysis reveals two general properties. Consider in general

$$\ddot{x} = -\gamma\dot{x} + F(x) + G(t) . \tag{4.1}$$

When $x(t)$ is perturbed to $x(t) + \xi(t)$,

$$\ddot{\xi} = -\gamma\dot{\xi} + F'(x)\xi . \tag{4.2}$$

For a Hopf bifurcation, we use (3.6) for ξ and retain only terms of the type $e^{\pm i\Omega t}$ in (4.2). Since $\Omega \neq 0, \omega/2, \ldots$, only the time-independent part of $F'(x)$ will generate a term of the right type, so that with the same notation as before,

$$\alpha^2 + 2i\alpha\Omega - \Omega^2 = -\gamma(\alpha + i\Omega) + \langle F'(x)\rangle , \tag{4.3}$$

whose imaginary part gives $\alpha = -\gamma/2 < 0$. Thus Hopf bifurcations never occur in systems of this type.[14]

For period-doubling bifurcations, Ω is replaced by $\omega/2$. Now if x has only odd harmonics and $F(x)$ is odd in x, then $F'(x)$ does not contain the frequencies $\pm\omega$, so the $e^{i\omega t/2}$ and $e^{-i\omega t/2}$ terms are not coupled, and again (4.3) holds, giving $\alpha < 0$. So a symmetric solution must first lose parity invariance before it can undergo period-doubling.[14,15]

The essential idea of this paper may be summarized as follows. The search for limit cycles using approximate solutions of the form say $a\cos\omega t + b\sin\omega t$ with constants a, b is classic, and the study of their stability by allowing slow variation of a, b is also well known. However, such perturbations reveal only one of the possible ways to lose stability. The Poincaré map provides simple test functions for the study of other instabilities. The technique is shown to be reasonably accurate in depicting the global features in the case of the Duffing equation. Just as the approximation for the solution can be improved by retaining more terms in say (2.1), so the stability criterion can likewise be improved by retaining more terms in say (3.3), leading to a finite-dimensional eigenvalue problem for α, which is always soluble. This method should be useful for analyzing other forced vibrating systems described by ODE's (including coupled ODE's).

[1] D. Y. Hsieh, J. Math. Phys. **16**, 275 (1975).
[2] D. Y. Hsieh, J. Math. Phys. **18**, 1093 (1977).
[3] D. Y. Hsieh, J. Math. Phys. **19**, 1147 (1979).
[4] D. Y. Hsieh, J. Math. Phys. **21**, 722 (1979).
[5] J. N. Elgin, D. Forster, and S. Sarkar, Phys. Lett. A **94**, 195 (1983).
[6] For Duffing equation, see, e.g., B. H. Huberman and J. P. Crutchfield, Phys. Rev. Lett. **93**, 1743 (1979).
[7] For Brusselator, see, e.g., B-L. Hao and S-Y. Zhang, Phys. Lett. A **87**, 267 (1982).
[8] For example, D. D'Humieres, M. R. Beasley, B. A. Huberman, and A. Libchaber, Phys. Rev. A **26**, 3483 (1983).
[9] E. Ott, Rev. Mod. Phys. **53**, 655 (1981); B. Hu, Phys. Rep. **91**, 233 (1982).
[10] For example, P. Hagedorn, *Non-linear Oscillations* (Oxford U.P., Oxford, 1982); R. E. Mickens, *An Introduction to Non-linear Oscillations* (Cambridge U.P., Cambridge, 1981).
[11] M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1979); see also works cited in Ref. 9.
[12] K. L. Liu, W. S. Lo, and K. Young, Phys. Lett. A **105**, 103 (1984).
[13] C. Grebogi, E. Ott, and J.A. Yorke, Phys. Rev. Lett. **48**, 1507 (1982).
[14] J. W. Swift and K. Wiesenfeld, Phys. Rev. Lett. **52**, 705 (1984).
[15] S. Novak and R. G. Frelich, Phys. Rev. A **26**, 3660 (1982).

Relativistic brachistochrone

Harris F. Goldstein and Carl M. Bender
Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 4 September 1985; accepted for publication 16 October 1985)

The trajectory joining two points $\mathbf{a}_1$ and $\mathbf{a}_2$, which minimizes the transit time for a particle, initially at rest, to fall in a uniform gravitational field from $\mathbf{a}_1$ to $\mathbf{a}_2$, is called the brachistochrone. Johann Bernoulli was the first to find an analytical form for the brachistochrone; in 1696, he discovered that the trajectory is a cycloid. In this paper the relativistic generalization of this classic problem is presented. Four separate curves are actually identified: a particle falling in both a uniform electric and uniform gravitational field is considered. The curves that minimize the times of flight measured by an observer in a laboratory in which $\mathbf{a}_1$ and $\mathbf{a}_2$ are fixed and also the curves that minimize the proper times of flight are found.

I. CLASSICAL BRACHISTOCHRONE

The brachistochrone problem is perhaps the most famous problem in the calculus of variations. It concerns a particle in a uniform gravitational field constrained to slide without friction along a curve joining the points $\mathbf{a}_1 = (x_1, y_1)$ and $\mathbf{a}_2 = (x_2, y_2)$. The problem is to find the curve that minimizes the time of fall.

We begin this problem by setting up a line-integral representation of the time of fall T:

$$T = \int_{(x_1, y_1)}^{(x_2, y_2)} \frac{ds}{v(s)}, \tag{1}$$

where s is the path length and $v(s)$ is the velocity of the particle. It is convenient to choose the coordinates so that (x_1, y_1) is the origin and so that the force field points along the x axis, as in Fig. 1. If we designate the path as $y(x)$ then the expression for T becomes an ordinary integral

$$T[y] = \int_0^{x_2} dx \frac{[1 + (y')^2]^{1/2}}{v} + \epsilon[y(0)] + \zeta[y(x_2) - y_2]. \tag{2}$$

Here, ϵ and ζ are Lagrange multipliers that impose the constraints that $y(0) = 0$ and that $y(x_2) = y_2$. Taking the functional derivative of (2) with respect to $y(w)$ gives

$$\frac{\delta T}{\delta y(w)} = \int_0^{x_2} dx \frac{y'(d/dx)\delta(x - w)}{[1 + (y')^2]^{1/2} v(x)} + \epsilon\delta(w) + \zeta\delta(w - x_2), \tag{3}$$

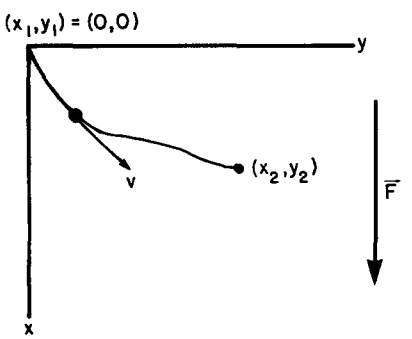


FIG. 1. The configuration and choice of coordinate system for the brachistochrone problem.

where by conservation of energy we can make the crucial assumption, also valid in the relativistic cases, that the particle velocity v is a function of x but is independent of y.

Integrating by parts we obtain

$$\frac{\delta T}{\delta y(w)} = -\int_0^{x_2} dx\, \delta(x - w) \frac{d}{dx}\left\{\frac{y'}{[1 + (y')^2]^{1/2} v(x)}\right\} + \frac{y'\delta(x - w)}{[1 + (y')^2]^{1/2} v}\Bigg|_{x=0}^{x=x_2} + \epsilon\delta(w) + \zeta\delta(w - x_2). \tag{4}$$

To minimize T we require that the functional derivative (4) vanish identically. From the interior of the interval $0 < x < x_2$ we obtain the differential equation

$$\frac{d}{dx}\left\{\frac{y'}{[1 + (y')^2]^{1/2} v(x)}\right\} = 0, \tag{5}$$

and from the end points $x = 0$ and $x = x_2$ we obtain the values of the Lagrange multipliers

$$\epsilon = \lim_{x \to 0} \frac{y'(x)}{\sqrt{1 + [y'(x)]^2}} v(x), \quad \zeta = \frac{-y'(x_2)}{\sqrt{1 + [y'(x_2)]^2}} v(x_2).$$

The solution to (5) is

$$y = \int_0^{x_2} dx \frac{k v(x)}{\sqrt{1 - k^2 [v(x)]^2}}, \tag{6}$$

where k is an integration constant.

This derivation of (6) is general; it is valid for any $v(x)$. In the classical case we take the force field to be a uniform gravitational field, and therefore, $v = (2gx)^{1/2}$. The path $y(x)$ is then a cycloid of the form

$$x = a(1 - \cos\theta), \quad y = a(\theta - \sin\theta),$$

where the parameter a is determined by the end point (x_2, y_2).

In this paper, we investigate the somewhat more difficult case of a particle falling under the influence of a uniform force field, but we do not neglect relativistic effects. In Sec. II we discuss the simplest case in which the particle falls in a vertical straight line. We discuss the physical differences between a particle falling in uniform electric field and a particle falling in a uniform gravitational field. In Sec. III we consider the general problem; we examine the curves that minimize the time of fall in the lab frame and also in the frame moving with the particle, for both a uniform electric field and a uniform gravitational field.

 0022-2488/86/020507-05$02.50 © 1986 American Institute of Physics

II. VERTICALLY FALLING RELATIVISTIC PARTICLES

To convey some intuitive understanding of the motion of a particle falling in a gravitational or an electric field we begin by studying the motion of a vertically falling particle. Since this is a study of realtivistic kinematics we will of course ignore the effects of energy lost due to radiation. To determine the motion of a relativistic particle falling in a uniform field we use Newton's Law:

$$\frac{dp}{dt} = F. \tag{7}$$

Consider a particle of charge q and rest mass m. If this particle falls in a uniform electric field of magnitude E the equation of motion (7) becomes

$$\frac{d}{dt}\left\{\frac{mv(t)}{\sqrt{1-[v(t)]^2/c^2}}\right\} = Eq\,. \tag{8a}$$

On the other hand, if this particle falls in a uniform gravitational field g then Eq. (7) becomes

$$\frac{d}{dt}\left\{\frac{mv(t)}{\sqrt{1-[v(t)]^2/c^2}}\right\} = \frac{mg}{\sqrt{1-[v(t)]^2/c^2}}\,. \tag{8b}$$

Note that (8a) and (8b) are different because the gravitational force is proportional to the relativistic (velocity dependent) mass and not the rest mass. Thus, as the particle's speed increases the gravitational force becomes stronger. No such effect occurs in (8a) because total electric charge is a relativistic scalar.

It is easy to obtain analytical solutions to both differential equations (8). If the particle is at rest at $t = 0$, we have

$$v_{\rm elec}(t) = cEqt/\sqrt{m^2c^2 + E^2q^2t}\,, \tag{9a}$$

$$v_{\rm grav}(t) = c\tanh(gt/c)\,. \tag{9b}$$

In the nonrelativistic case, $c\to\infty$, Eq. (9) reduces to the usual results $v_{\rm elec}(t) = Eqt/m$ and $v_{\rm grav}(t) = gt$. The results in (9) could have been derived just as easily from a conservation of energy argument. In fact, we will use the principle of conservation of energy in Sec. III when we derive the equation for the relativistic brachistochrone.

If we integrate (9) with respect to t and assume that at $t = 0$, the particle is situated at $x = 0$, we get the position of the particle as a function of t:

$$x_{\rm elec}(t) = (mc^2/Eq)(\sqrt{1 + E^2q^2t^2/(m^2c^2)} - 1)\,, \tag{10a}$$

$$x_{\rm grav}(t) = (c^2/g)\log[\cosh(gt/c)]\,. \tag{10b}$$

To compare the results in Eqs. (10) let us take the gravitational and electrical forces to be equal for stationary particles: $Eq = mg$. Eliminating Eq/m from (10a) and examining the large t behavior of $x_{\rm grav}(t) - x_{\rm elec}(t)$ gives

$$\lim_{t\to\infty}[x_{\rm grav}(t) - x_{\rm elec}(t)] = (c^2/g)(1 - \ln 2)\,. \tag{11}$$

The right side of (11) is positive because the gravitational force gradually increases with time while the electric force remains constant.

III. RELATIVISTIC BRACHISTOCHRONE

A. Uniform electric field

As we saw in Sec. II it is simplest to consider the case of a velocity-independent force. Consider a particle of charge q and rest mass m falling under the influence of an electric field of magnitude E. Energy conservation gives

$$mc^2 = mc^2/\sqrt{1 - v^2/c^2} - qEx\,.$$

Solving for v we get

$$v = c\sqrt{1 - [1/(1+\alpha x)]^2}\,, \tag{12}$$

where $\alpha = qE/(mc^2)$. Substituting (12) into (6) gives

$$y = \int_0^x dx' \sqrt{\frac{k^2(1+\alpha x')^2 - k^2}{(1-k^2)(1+\alpha x')^2 + k^2}}\,, \tag{13}$$

where we have absorbed c into the constant k. There are three distinct regions of values of k^2, where the function $y(x)$ has very different behaviors: $0 < k^2 < 1$, $k^2 = 1$, and $k^2 > 1$.

1. Region 1: $0 < k^2 < 1$

In this region we have

$$y = \xi\int_0^x dx' \sqrt{\frac{(1+\alpha x')^2 - 1}{(1+\alpha x')^2 + \xi^2}} = \frac{\xi}{\alpha}\int_1^{1+\alpha x} du \sqrt{\frac{u^2-1}{u^2+\xi^2}}\,, \tag{14}$$

where

$$\xi^2 \equiv k^2/(1-k^2)$$

is determined from the condition $y(x_2) = y_2$. Here, as x becomes large, the integrand approaches 1 and the graph of y approaches a straight line of slope ξ. The solutions in this region are very different from the classical case, since they increase without bound, while the classical solutions are bounded and cyclic.

2. Region 2: $k^2 = 1$

For this value of k^2, $y(x)$ reduces to

$$y = \frac{1}{\alpha}\int_1^{1+\alpha x} du\sqrt{u^2-1}\,,$$

which can be expressed in terms of simple functions as

$$y = \frac{1}{2\alpha}\left[(1+\alpha x)\sqrt{(1+\alpha x)^2 - 1} - \log\left[(1+\alpha x) + \sqrt{(1+\alpha x)^2 - 1}\right]\right]. \tag{15}$$

For $\alpha x \gg 1$,

$$y \simeq (1/2\alpha)(\alpha^2x^2) = \tfrac{1}{2}\alpha x^2\,,$$

a parabola. This function is the boundary between the region 1 solutions, which increase without bound in the x direction, and those in region 3, which turn around and return to the y axis. Note that in the classical limit $\alpha\to 0$, this boundary approaches the x axis, which is the vertically falling case described in Sec. II.

3. Region 3: $k^2 > 1$

In this region we have

$$y = \frac{\xi}{\alpha}\int_1^{1+\alpha x} du\sqrt{\frac{u^2-1}{\xi^2 - u^2}}\,, \tag{16}$$

where

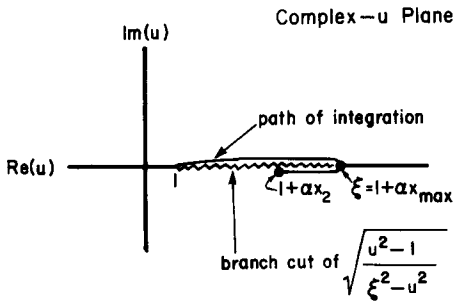


FIG. 2. The complex-u plane showing the path of integration necessary to compute the entire graph of the brachistochrone.

$$\xi^2 = k^2/(k^2-1)\,.$$

Since the integrand must be real, we have a lower bound on ξ of $1+\alpha x_2$. We must be cautious with this solution, however, because there is a singularity in the integrand at $u=\xi$. Notice now that there is a branch cut in the complex-u plane connecting $u=1$ and $u=\xi$. If we integrate around this branch cut (see Fig. 2), we get a graph of $y(x)$ that curves back to the y axis, just as the cycloid does in the classical solution.

Figure 3 shows the family of curves in all three regions for $\alpha=10$. As we approach the nonrelativistic limit the curve corresponding to region 2 moves toward the x axis, region 3 expands, and region 1 shrinks in area. In the extreme nonrelativistic limit ($c=\infty$, $\alpha=0$), region 1 disappears entirely, region 2 lies on the x axis, and the curves in region 3 become the cycloids discovered by Bernoulli.

Figure 4 shows the solutions for a given end point (x_2, y_2) and several values of α. For α small, the curve is indistinguishable from the classical cylcoid.

B. Proper-time electric-field brachistochrone

In relativistic mechanics it is possible to pose two different brachistochrone problems. We can now ask what is the curve that minimizes the elapsed proper time as measured by an observer moving with the particle. For this we must backtrack slightly and look at our original expression for T or, more precisely, the expression for dT, namely $dT=ds/v$. The element of proper time, $d\tau$, is given by

$$d\tau = dT\sqrt{1-v^2/c^2}\,,$$

so

$$\tau = \int_0^{x_2} dx\,\frac{\sqrt{(1+(y')^2}}{v}\sqrt{1-\frac{v^2}{c^2}} + \epsilon y(0) + \zeta\,[\,y(x_2)-y_2\,]\,. \tag{17}$$

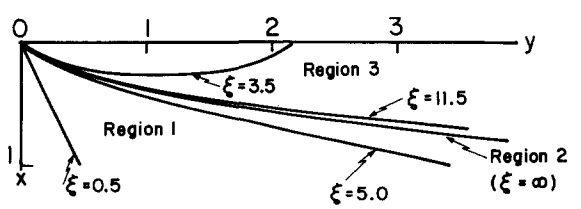


FIG. 3. Relativistic brachistochrones corresponding to $\alpha=10$. Curves in regions 1, 2, and 3 are shown.

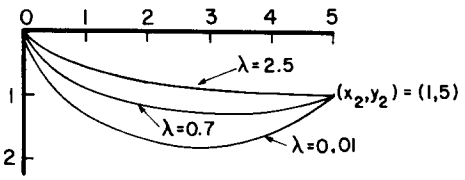


FIG. 4. The lab-time brachistochrones for the electric field with fixed endpoints and various values of α. The case $\alpha=0.01$ is indistinguishable from the cycloid.

This is the generalization of (2).

Performing the variations as before and solving for an arbitrary $v(x)$ we get

$$y = \int_0^x dx'\sqrt{\frac{k^2v^2}{(1-v^2/c^2)-k^2v^2}}\,.$$

For the electric field, we have

$$v = c\sqrt{1-[1/(1+\alpha x)]^2}$$

or

$$\sqrt{1-v^2/c^2} = 1/(1+\alpha x)\,.$$

Substituting, we find

$$y = \int_0^x dx'\sqrt{\frac{k^2[(1+\alpha x')^2-1]}{1-k^2[(1+\alpha x')^2-1]}} = \int_1^{1+\alpha x}\frac{du}{\alpha}\sqrt{\frac{u^2-1}{\xi^2-u^2}}\,, \tag{18}$$

where

$$\xi^2 = (k^2+1)/k^2\,.$$

This result is interesting because except for the missing factor of the parameter ξ, this solution is identical in form to that of region 3 in the lab-time case. We see then that we have lost the phenomenon of unbounded solutions by considering proper time!

Figure 5 shows the solutions for fixed (x_2, y_2) and various values of α.

C. Uniform gravitational field

The problem of determining the relativistic brachistochrone in a uniform gravitational field is somewhat more com-

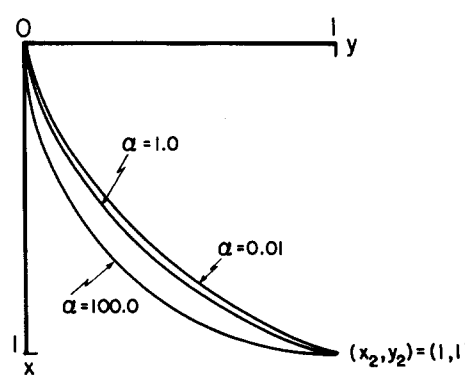


FIG. 5. Proper-time brachistochrones for the electric field. Observe that the lab-time brachistochrones become less concave as the field strength E increases while the proper-time brachistochrones in this limit become more concave.

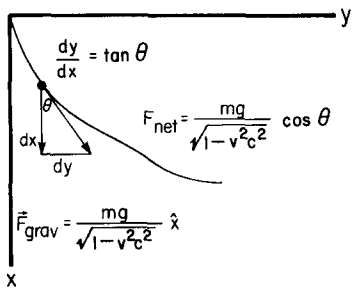


FIG. 6. Relativistic force diagram for a particle on a relativistic brachistochrone in a uniform gravitational field.

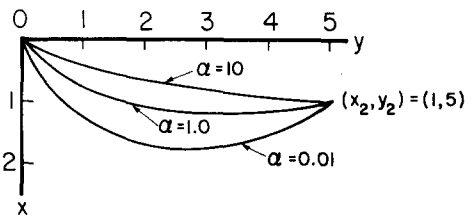


FIG. 8. The lab-time brachistochrones for the uniform gravitational field with fixed endpoints and various values of α. The case $\alpha = 0.01$ is indistinguishable from the cycloid.

plicated than the same problem for a a uniform electric field. The additional complexity occurs because, as discussed in Sec. II, the gravitational force depends on the relativistic mass of the particle, which is a function of velocity, while the electric force is only dependent on the total charge, which is invariant. Consequently, we must find the expression for the velocity of the particle.

First, by conservation of energy, we have

$$mc^2 = mc^2/\sqrt{1 - v^2/c^2} - \mathscr{E} ,$$

where we define $\mathscr{E}$ to be the energy gained by the particle from the force field, given by

$$\mathscr{E} = \int \mathbf{F} \cdot d\mathbf{s} .$$

Now, the force on the particle due to gravity and acting in the direction of motion is (see Fig. 6)

$$F = (mg/\sqrt{1 - v^2/c^2}\,) \cos\theta ,$$

where θ is the angle between the direction of motion of the particle and the x axis. But since $y' = \tan\theta$ we must have

$$\cos\theta = (1 + \tan^2\theta)^{-1/2} = [1 + (y')^2]^{-1/2} ,$$

and since $ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + (y')^2}\, dx$ we get

$$\mathscr{E} = \int dx' \frac{mg}{\sqrt{1 - v^2/c^2}} .$$

Now we have two equations in $\mathscr{E}$ and v. Solving the energy conservation equation for v and eliminating v, we get

$$\mathscr{E} = \int^x mg\left(1 + \frac{\mathscr{E}}{mc^2}\right) dx' .$$

Differentiating with respect to x, we obtain a differential equation for $\mathscr{E}$,

$$\mathscr{E}' = mg(1 + \mathscr{E}/mc^2) ,$$

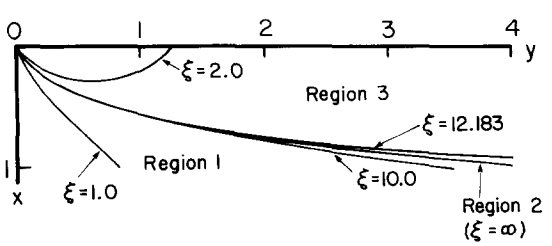


FIG. 7. Several curves in the family of brachistochrones for the gravitational field with $\alpha = 2.5$.

whose exact solution with $\mathscr{E}(0) = 0$ is

$$\mathscr{E} = mc^2(e^{gx/c^2} - 1) .$$

Substituting this back into the energy conservation equation, we finally have an expression for $v(x)$:

$$v = c\sqrt{1 - e^{-2\alpha x}} ,$$

where $\alpha = g/c^2$.

Putting this into our general expression for the brachistochrone for lab time we get

$$y = \int_0^x dx' \sqrt{\frac{k^2(1 - e^{-2\alpha x'})}{(1 - k^2) + k^2 e^{-2\alpha x'}}} . \tag{19}$$

As in the uniform electric field case, the family of solutions is divided into three regions: $0 < k^2 < 1$, $k^2 = 1$, and $k^2 > 1$.

The boundary region $k^2 = 1$ in this case gives a brachistochrone whose analytical solution is

$$y = (1/\alpha)\left[\sqrt{e^{2\alpha x} - 1} - \sec^{-1} e^{\alpha x}\right] , \tag{20}$$

which behaves like an exponential for large x, in contrast to the boundary region for the uniform electric field, which behaves like a parabola. (See Figs. 7 and 8.)

D. Proper-time gravitational-field brachistochrone

By a slight modification of the procedure for solving the uniform-electric-field proper-time brachistochrone we find the expression for the uniform-gravitational-field proper-time brachistochrone to be

$$y = \int_0^x dx' \sqrt{\frac{e^{2\alpha x'} - 1}{\xi^2 - e^{2\alpha x'}}} .$$

The behavior of this solution is unexpected. Let us examine this integral for large x. We have

$$y < \int_0^x dx' \frac{e^{\alpha x'}}{\sqrt{\xi^2 - e^{2\alpha x'}}} < \frac{1}{\alpha}\int_1^\xi \frac{du}{\sqrt{\xi^2 - u^2}}$$

$$= \frac{1}{\alpha}\left[\arcsin 1 - \arcsin\frac{1}{\xi}\right]$$

$$= \frac{1}{\alpha}\left[\frac{\pi}{2} - \arcsin\frac{1}{\xi}\right] < \frac{\pi}{2\alpha} .$$

Since we must integrate along both sides of the branch cut, the curve is bounded in the y direction by twice this quantity or π/α. At first glance, this result appears absurd; it seems that there are no brachistochrones that pass through points where $y_2 > \pi/\alpha$. However, we *do* have brachistochrones to points lying in the region $y_2 > \pi/\alpha$. We can see in Fig. 9 that

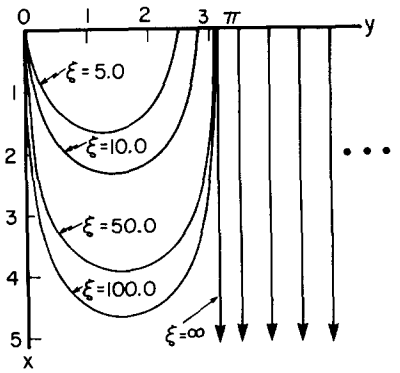


FIG. 9. The family of proper-time gravitational-field brachistochrones for $\alpha = 1.0$.

as $\xi \to \infty$, the brachistochrones become deeper. The limit of these curves is a path that falls vertically to infinity on the x axis, moves to $y = \pi/\alpha$ while at $x = \infty$, and then returns to $x = 0$ along the line $y = \pi/\alpha$. The proper time experienced by the particle along this path is

$$\tau = 2\int_0^\infty dx \frac{\sqrt{1 - v^2/c^2}}{v} = \frac{2}{c}\int_0^\infty dx \frac{e^{-\alpha x}}{\sqrt{1 - e^{-2\alpha x}}}$$

$$= \frac{2}{\alpha c}\int_1^\infty \frac{1}{u\sqrt{u^2 - 1}} = \frac{\pi}{\alpha c},$$

which is finite. Because the particle is traveling at the speed of light at $x = \infty$ it experiences zero proper time while moving in the y direction, independent of the distance traveled in that direction. Therefore, we can go to any point (x_2, y_2) for which $y_2 > \pi/\alpha$ by going to infinity along the x axis and coming back up along the line $y = y_2$.

Similar solutions do not exist for the proper-time electric-field brachistochrones. The proper time for a particle to fall to infinity in an electric field is infinite so all brachistochrones in the electric field must have finite length.

ACKNOWLEDGMENT

We thank the U.S. Department of Energy for financial support.

Front form and point form formulation of predictive relativistic mechanics. Noninteraction theorems

X. Jaen and A. Molina
Grup de Relativitat de la Secció de Física SCC (IEC), Departament de Física Teòrica, Universitat de Barcelona, Barcelona, Spain

V. Iranzo
Grup de Relativitat de la Secció de Física SCC (IEC), Universitat Politècnica de Catalunya (ETSECCP), Barcelona, Spain

(Received 14 September 1984; accepted for publication 20 September 1985)

The front form and the point form of dynamics are studied in the framework of predictive relativistic mechanics. The noninteraction theorem is proved when a Poincaré-invariant Hamiltonian formulation with canonical position coordinates is required.

I. INTRODUCTION

Instantaneous relativistic dynamics of particles with direct interaction was initiated in a celebrated paper by Dirac,[1] entitled "Forms of relativistic dynamics." The line of thought set up there was further developed by Bakanjian, Thomas, and Foldy,[2] in the framework that Dirac had called "instant form."

The subsequent development of the theory met the important drawback of the so-called "no-interaction theorem." [3] In general terms, it states that, if the position coordinates of the particles are to be canonical coordinates, and the particle worldlines must be Poincaré invariant, then the only systems that are compatible with both requirements are those consisting of free particles.

One attempt to circumvent this problem was initiated by Currie,[4] and later on it has generated a rather wide stream of literature, which is known as predictive relativistic mechanics (and maybe, it should be called an instant form of PRM). It consists, first, in giving up the Hamiltonian formalism, which was taken for granted in former approaches, and starting from a more elementary level. The fundamental assumptions in predictive relativistic mechanics are (i) the equations of motion of the particles are Newton-like, that is, the acceleration of each particle is a given function of positions and velocities of all particles; and (ii) Poincaré invariance, which is understood to mean two things: the acceleration functions must be formally the same in all inertial reference frames, and particle worldlines must be Poincaré invariant. These requirements imply that some condition (the so-called Currie–Hill equations[5]) must be fulfilled by the acceleration functions. In addition, they also ensure the possibility of setting up a realization of the Poincaré algebra[6] on the system's tangent space (the one spanned by positions and velocities). Now, the no-interaction result can be obtained again[7] if one seeks for a Hamiltonian formalism such that the aforementioned realization of the Poincaré algebra is canonical, and the position coordinates can be taken as canonical.

As far as we know, all proofs of the no-interaction theorem hitherto derived share a common feature, namely, physical variables are assumed to be simultaneous in a given inertial frame. This is a specific trait of the "instant form" of relativistic dynamics. However, in the pioneering paper by Dirac,[1] two other possibilities were considered, namely, the "front form" and the "point form" (in fact, a later paper by Leutwyler and Stern[7] increases that number by two more "forms").

One then wonders whether the no-interaction theorem, or a related result, also holds in these two alternative forms of relativistic dynamics. Although this is, indeed, an interesting point to be elucidated, it seems not to have been proven yet. Indeed, in a relatively recent paper by Leutwyler and Stern[7] we can find the following sentence: "Although this no go theorem has been established only for theories of class (i) (i.e., the "instant form" of relativistic dynamics) it likely also holds for the remaining four forms of Hamiltonian dynamics."

In the present paper we intend to give an answer to the question that is more or less implicit in the quoted sentence, and derive a no-interaction theorem in the front form as well as in the point form. The master lines of our proof are the same as those of the proof given by Hill[5] for the no-interaction theorem in the "instant form."

In a natural way, the paper is divided in two parts. The first one (Secs. II and III) is devoted to the front form, and the second one (Secs. IV and V) to the point form. Besides, each part is organized in two sections: one devoted to develop what could be called the front (resp. point) form of predictive relativistic mechanics, and the other to prove the no-interaction theorem.

II. FRONT FORM OF PREDICTIVE RELATIVISTIC MECHANICS

In the instant form of predictive relativistic mechanics[8] (which has been its only formulation up to now), the extended configuration space of N spinless particles is spanned by the $3N+1$ variables: t, x_b^i, $b=1,\ldots,N$, $i=1,2,3$; where the evolution parameter is the time coordinate as measured in a given inertial frame, and the x_b^i are the space coordinates of the event determined by the intersection of the worldline of particle b and the space hyperplane $x^4=t$.

The equations of motion are then required to be second-order differential equations, that is,

$$\frac{d^2x_b^i}{dt^2}=a_b^i(x_c^i,v_e^k,t), \quad \frac{dx_b^i}{dt}=v_b^i . \tag{2.1}$$

 0022-2488/86/020512-07$02.50 © 1986 American Institute of Physics

Thus the space of initial data is spanned by the following $6N+1$ variables:

$$t,\ x_b^i,\ v_c^k,\quad b,c=1,\ldots,N,\quad i,k=1,2,3\,.$$

If the space hyperplanes $x^4=t$ characterize the instant form, likewise the null hyperplanes $x^3+x^4=\lambda$ will play a central role in the front form of relativistic dynamics (here t and λ are two real parameters). So, the extended configuration space in the front form will be coordinated by the $3N+1$ variables: $\lambda, x_b^i, b=1,\ldots,N, i=1,2,3$; where λ is the evolution parameter and x_b^i, $i=1,2,3$, are the space coordinates of the event where the worldline of particle b meets the null hyperplane

$$x^3+x^4=\lambda\,, \tag{2.2}$$

or, using the notation introduced in (A5),

$$x^+=\lambda$$

(the same value of λ for all particles).

For convenience, our configuration space coordinates will be (see Appendix A)

$$x_a^A,\quad a=1,\ldots,N,\quad A=1,2,-\,,$$

rather than the Cartesian x_a^i, $i=1,2,3$.

We now require the motion to be governed by a second-order differential system,

$$\frac{d^2x_b^A}{d\lambda^2}=a_b^A(x_a^B,v_c^D;\lambda)\,,\quad \frac{dx^A}{d\lambda}=v_b^A\,, \tag{2.3}$$
$$a,b,c=1,\ldots,N,\quad A,B,D=1,2,-\,.$$

For every given solution of (2.3), we have a set of N worldlines describing the history of the system. Indeed, if $\varphi_b^A(x_a^B,v_c^D;\lambda)$, $A,B,D=1,2,-$, $a,b,c=1,\ldots,N$, is the solution of (2.3) corresponding to the initial data

$$\varphi_b^A(x_a^B,v_c^D;0)=x_b^A\,,$$
$$\frac{\partial\varphi_b^A}{\partial\lambda}(x_a^B,v_c^D;0)=v_b^A\,, \tag{2.4}$$

then, according to (2.2) and (A5), the worldline $x_b^\mu(\lambda)$ of particle b will be taken as

$$\begin{aligned} x_b^i(\lambda)&=\varphi_b^i(x_a^B,v_c^D;\lambda)\,,\quad i=1,2,\\ x_b^3(\lambda)&=\lambda/2+\varphi_b^-(x_a^B,v_c^D;\lambda)\,,\\ x_b^4(\lambda)&=\lambda/2-\varphi_b^-(x_a^B,v_c^D;\lambda)\,, \end{aligned} \tag{2.5}$$

which in the adapted coordinates (A4) reads

$$x_b^A(\lambda)=\varphi_b^A(x_a^B,v_c^D;\lambda)\,,\quad x_b^+(\lambda)=\lambda\,. \tag{2.6}$$

Similarly to the instant form description, the principle of relativity will be used at two different levels. First, the "acceleration" functions a_b^A on the right of Eq. (2.3) must have the same form in every inertial frame. And second, the dynamic system must be worldline invariant. The latter requirement means the same as in the instant form case, namely, that if S and S' are two inertial frames related to each other by a Poincaré transformation $(\mathscr{L}_{\overline{B}}^{\overline{A}},\mathscr{A}^{\overline{D}})$, $\overline{A},\overline{B},\overline{D}=1,2,-,+$ —see the Appendix— and $x_b^{\overline{A}}(\lambda)$, $b=1,\ldots,N$, $\overline{A}=1,2,-,+$ are the worldlines of the particles in the frame S, when the system starts from a given set of initial data $Z_0\equiv(x_a^A,v_c^D)$, $a,c=1,\ldots,N$, $A,D=1,2,-$; then the Poincaré-transformed worldlines

$$x_b'^{\overline{A}}(\lambda)=\mathscr{L}_{\overline{B}}^{\overline{A}}(x_b^{\overline{B}}(\lambda)-A^{\overline{B}}) \tag{2.7}$$

must be obtained in the frame S', when the system starts from the transformed set of initial data $z_0'\equiv(x_a'^A,v_c'^D)$.

Thus, as happens in the instant form, the mapping $z_0\to z_0'$ defines the induced action of Poincaré transformation $(\mathscr{L}_{\overline{B}}^{\overline{A}},\mathscr{A}^{\overline{D}})$ on the space of initial data.

In terms of the adapted coordinates, this mapping reads

$$\begin{aligned} x_a'^A&=f_a^A(x_b^B,v_c^D;\epsilon^{\overline{E}},\epsilon^{\overline{FG}})\,,\\ v_a'^A&=g_a^A(x_b^B,v_c^D;\epsilon^{\overline{E}},\epsilon^{\overline{FG}})\,, \end{aligned} \tag{2.8}$$

where $\epsilon^{\overline{E}},\epsilon^{\overline{FG}}$, $\overline{E},\overline{F},\overline{G}=1,2,-,+$ denote the parameters characterizing the specific Poincaré transformation—see Eq. (A11).

The infinitesimal generators are then obtained in the usual way:

$$\begin{aligned} \mathbf{P}_{\overline{E}}&=\frac{\partial f_a^A}{\partial\epsilon^{\overline{E}}}(x_b^B,v_c^D;0,0)\frac{\partial}{\partial x_a^A}+\frac{\partial g_b^B}{\partial\epsilon^{\overline{E}}}(x_b^B,v_c^D,0,0)\frac{\partial}{\partial v_b^B}\,,\\ \mathbf{J}_{\overline{EF}}&=\frac{\partial f_a^A}{\partial\epsilon^{\overline{EF}}}(x_b^B,v_c^D;0,0)\frac{\partial}{\partial x_a^A}+\frac{\partial g_b^B}{\partial\epsilon^{\overline{EF}}}(x_b^B,v_c^D,0,0)\frac{\partial}{\partial v_b^B}\,. \end{aligned} \tag{2.9}$$

To obtain the specific expressions for these generators, we shall work out the condition of worldline invariance (2.7), together with the worldline equations (2.6). Introducing the latter into both sides of (2.7), we obtain

$$\begin{aligned} \varphi_a^A(z_0',\lambda_a(z_0,\lambda))=\mathrm{TX}\mathscr{L}_E^A[\,\varphi_a^E(z_0;\lambda)-\mathscr{A}^E\,]\\ +\mathscr{L}_+^A\,[\lambda-\mathscr{A}^+]\,,\quad A=1,2,-\,. \end{aligned} \tag{2.10}$$

It should be noticed that, since the "acceleration" functions have the same form in frame S and in S', the same general solution φ_a^A has been substituted into both sides of Eq. (2.7). However, whereas in the right-hand side we take the initial data $z_0=(x_a^A,v_b^B)$, which correspond to the frame S, in the left-hand side we have to put the transformed initial data $z_0'=(x_a'^A,v_b'^B)$ which correspond to the worldlines as viewed from the frame S'. Moreover, the value of the evolution parameter in the left-hand side of Eq. (2.10), which we have written as $\lambda_a(z_0,\lambda)$, will be presumed different from the parameter λ in the right-hand side. This is due to the fact that the worldline invariance only ensures that each worldline transforms into another one as a whole, no matter how the respective parametrizations are related to each other.

In our case, the relation between parametrizations, i.e, the function $\lambda_a(z_0,\lambda)$, can be derived from the fact that the evolution parameter corresponds to the space-time coordinate x^+; so that we have

$$\lambda_a(z_0,\lambda)=\mathscr{L}_B^+(\varphi_a^B(z_0,\lambda)-\mathscr{A}^B)+\mathscr{L}^+ +(\lambda-\mathscr{A}^+)\,. \tag{2.11}$$

By taking derivatives in (2.10) and (2.11) with respect to λ we obtain the relation between velocities

$$\dot\varphi_a^A(z_0',\lambda_a(z_0,\lambda))\cdot\dot\lambda_a(z_0,\lambda)=\mathscr{L}_E^A\cdot\dot\varphi_a^E(z_0,\lambda)+\mathscr{L}_+^A\,, \tag{2.12}$$

with

$$\dot{\lambda}_a(z_0,\lambda) = \mathscr{L}_B^+ \cdot \dot{\varphi}_a^B(z_0,\lambda) + \mathscr{L}_+^+ ,$$

which, after one further differentiation, yields the relation between accelerations:

$$\ddot{\varphi}_a^A(z_0', \lambda_a(z_0,\lambda))\dot{\lambda}_a^2(z_a,\lambda) + \dot{\varphi}_a^A(z_0', \lambda_a(z_0,\lambda))\ddot{\lambda}_a(z_0,\lambda) = \mathscr{L}_B^A \ddot{\varphi}_a^B(z_0,\lambda) \tag{2.13}$$

and

$$\ddot{\lambda}_a(z_0,\lambda) = \mathscr{L}_B^+ \ddot{\varphi}_a^B(z_0,\lambda) .$$

Since φ_a^A is a solution of the differential system (2.3), Eqs. (2.13) can be written as

$$\mathscr{A}_b^A(\varphi_b^B(z_0', \lambda_a(z_0,\lambda)); \dot{\varphi}_c^D(z_0', \lambda_a(z_0,\lambda)); \lambda_a(z_0,\lambda)) \cdot \dot{\lambda}_a^2(z_0,\lambda) + \dot{\varphi}_a^A(z_0'; \lambda_a(z_0,\lambda)) \cdot \ddot{\lambda}_b(z_0,\lambda) = \mathscr{L}_B^A \mathscr{A}_a^B(\varphi_b^D(z_0,\lambda), \dot{\varphi}_c^F(z_0,\lambda), \lambda), \tag{2.14}$$

$$\ddot{\lambda}_a(z_0,\lambda) = \mathscr{L}_B^+ \mathscr{A}_a^B(\varphi_b^D(z_0,\lambda), \dot{\varphi}_c^F(z_0,\lambda), \lambda) .$$

Equations (2.10)–(2.14), which hold for every value of λ and for every Poincaré transformation $(\mathscr{L}_B^A, \mathscr{A}^D)$, actually determine the functions f_a^A, g_b^B in (2.8). Although, apart from a few trivial cases, it would be impossible to derive explicit expressions for such functions, the above equations permit us to obtain the infinitesimal generators in a rather straightforward way.

Indeed, introducing the infinitesimal expression (A12) for the Poincaré transformation $(\mathscr{L}_B^A, \mathscr{A}^D)$ into Eqs. (2.10)–(2.12) and keeping first-order terms only, we obtain, after some manipulation,

$$\mathbf{P}_+ = \sum_{A=1,2,-} \sum_{a=1}^{N} \left[v_a^A \frac{\partial}{\partial x_a^A} + a_a^A \frac{\partial}{\partial v_a^A} \right], \tag{2.15a}$$

$$\mathbf{P}_A = -\sum_{a=1}^{N} \frac{\partial}{\partial x_a^A}, \quad A = 1,2,- , \tag{2.15b}$$

$$\mathbf{J}_{+r} = -\sum_{a=1}^{N} \left\{ (x_{a+}\delta_r^A + x_{ar} v_a^A) \frac{\partial}{\partial x_a^A} - (v_{a+}\delta_r^A + x_{ar} a_a^A + v_{ar} v_a^A) \frac{\partial}{\partial v_a^A} \right\}, \tag{2.16a}$$

$$\mathbf{J}_{+-} = -\sum_{a=1}^{N} \left\{ x_{a+} \frac{\partial}{\partial x_a^-} + v_{a+} \frac{\partial}{\partial v_a^-} + \sum_{A=1,2,-} v_a^A \frac{\partial}{\partial v_a^A} \right\}, \tag{2.16b}$$

$$\mathbf{J}_{r-} = \sum_{a=1}^{N} \left\{ -x_{ar} \frac{\partial}{\partial x_a^-} - v_{ar} \frac{\partial}{\partial v_a^-} + \frac{\partial}{\partial v_a^r} \right\}, \tag{2.16c}$$

$$\mathbf{J}_{1z} = \sum_{a=1}^{N} \left\{ -x_{az} \frac{\partial}{\partial x_a^1} - x_{a1} \frac{\partial}{\partial x_a^2} + v_{az} \frac{\partial}{\partial v_a^1} - v_{a1} \frac{\partial}{\partial v_a^2} \right\}, \tag{2.16d}$$

where $r = 1,2$.

Notice that, as expected, there are seven kinematical generators $\mathbf{P}_A$, $\mathbf{J}_{12}$, $\mathbf{J}_{r-}$, $\mathbf{J}_{+-}$, $A = 1,2,-$, $r = 1, 2$; and three dynamical ones $\mathbf{P}_+$, $\mathbf{J}_{+r}$, $r = 1,2$; or Hamiltonians.

Now, by introducing the same infinitesimal Poincaré transformation (A12) into Eqs. (2.14), we obtain the set of differential equations

$$\begin{aligned} &\mathbf{P}_A a_a^B = 0, \quad A,B = 1,2,- , \\ &\mathbf{J}_{12} a_a^B = a_a^2 \delta_1^B - a_a^1 \delta_2^B , \\ &\mathbf{J}_{r-} a_a^B = -a_a^r \delta_-^B , \\ &\mathbf{J}_{+-} a_a^B = -(a_a^- \delta_-^D + z a_a^D) , \\ &\mathbf{J}_{+r} a_a^B = -z v_{ar} a_a^B - a_{ar} v_a^B - a_a + \delta_r^B - x_a^r (\mathbf{P}_+ a_a^B) , \end{aligned} \tag{2.17}$$

which play a similar role as the Currie–Hill conditions[5] in the instant form of predictive relativistic mechanics.

III. NO-INTERACTION THEOREM IN THE FRONT FORM OF DYNAMICS

Let us now assume that we have a symplectic structure in the space of initial data, such that the "position variables" x_a^A, $A = 1,2,-$, are canonical, i.e., the symplectic form is

$$\sigma = dx_a^A \wedge dp_A^a , \tag{3.1}$$

where $p_A^a = p_A^a(x_b^B, v_c^D)$, $a,b,c = 1,\ldots,N$, $A,B,D = 1,2,-$, and summation over repeated indices (A as well as a) is understood.

Let us furthermore assume that the realization of the Poincaré group that we discussed in the previous section is canonical relatively to σ. This implies that σ is Poincaré invariant or, equivalently,

$$\mathscr{L}(\mathbf{P}_{\overline{A}})\sigma = 0, \quad \mathscr{L}(\mathbf{J}_{\overline{AB}})\sigma = 0 , \tag{3.2}$$

where $\overline{A},\overline{B} = 1,2,-,+$ and $\mathscr{L}$ means "Lie derivative."

As a consequence of Eq. (3.2), there exist ten generating functions $P_{\overline{A}}(x,p)$, $J_{\overline{AB}}(X,p)$ such that

$$i(\mathbf{P}_{\overline{A}})\sigma = -dP_{\overline{A}} \quad \text{and} \quad i(\mathbf{J}_{\overline{AB}})\sigma = -dJ_{\overline{AB}} ,$$

or

$$\mathbf{P}_{\overline{A}} = \{P_{\overline{A}}, -\} \quad \text{and} \quad \mathbf{J}_{\overline{AB}} = \{J_{\overline{AB}}, -\} , \tag{3.3}$$

where i means "inner product" and $\{\ ,\ \}$ is the Poisson bracket associated to σ.

Now, using Eq. (3.3) and the expressions (2.15) and (2.16) for the Poincaré generators, we arrive at

$$\begin{aligned} &\{x_a^A, P_B\} = \delta_B^A, \quad \{x_a^A, J_{12}\} = -x_a^2 \delta_1^A + x_a^i \delta_z^A , \\ &\{x_a^A, J_{+-}\} = x_a^- \delta_-^A , \quad \{x_a^A, J_{r-}\} = x_a^r \delta_-^A , \\ &\{x_a^A, J_{+r}\} = x_a^- \delta_r^A + x_a^r v_a^A , \quad \{x_a^A, P_+\} = -v_a^A , \\ &r = 1,2, \quad A,B = 1,2,- . \end{aligned} \tag{3.4}$$

By applying the commutation relation of the Poincaré algebra and using the Jacobi identity and Eqs. (3.4), we find, after some calculations,

$$\{J_{+r}, \{x_b^B, x_a^A\}\} = \delta_r^i \{x_a^-, x_b^B\} + v_a^A \{x_a^r, x_b^B\} + x_a^r \{v_a^A, x_b^B\} - v_b^B \{x_b^r, x_a^A\} - x_b^r \{v_b^B, x_a^A\} , \tag{3.5}$$

$$\{x_a^A, v_b^B\} = \{x_b^B, v_a^A\} + \{\{x_b^B, x_a^A\}, P_+\} , \tag{3.6}$$

$$\{x_a^A, a_b^B\} = \{v_b^B, v_a^A\} + \{\{v_b^B, x_a^A\}, P_+\}, \tag{3.7}$$

$$\begin{aligned} &- \{\{x_b^B, v_a^A\}, J_{+r}\} + \{x_b^-, v_a^A\}\delta_r^B + \{x_a^r, v_a^A\}\delta_r^B \\ &\quad + \{x_b^r, v_a^A\}v_b^B + \{x_b^B, x_a^A\}x_b^r \\ &\quad = \{v_a^-, x_b^B\}\delta_r^A + v_a^r\{v_a^A, x_b^B\} + v_a^A\{v_a^r, x_b^B\} \\ &\qquad - a_a^A\{x_a^r, x_b^B\} - x_a^r\{a_a^A, x_b^B\}, \end{aligned} \tag{3.8}$$

$$\begin{aligned} &- \{\{v_b^B, v_a^A\}, J_{+r}\} + \{v_b^-, v_a^A\}\delta_r^B + \{v_b^r, v_a^A\}v_b^B \\ &\quad + \{v_b^B, v_a^A\}v_b^r - \{x_b^r, v_a^A\}a_b^B - \{a_b^B, v_a^A\}x_b^r \\ &\quad = \{v_a^-, v_b^B\}\delta_r^A + \{v_a^r, v_b^B\}v_a^A \\ &\qquad + \{v_a^A, v_b^B\}v_a^r - \{x_a^r, v_b^B\}a_a^A - \{a_a^A, v_b^B\}x_a^r. \end{aligned} \tag{3.9}$$

Since the coordinates x_a^A are assumed to be canonical, we have that $\{x_a^A, x_b^B\} = 0$, which, combined with (3.5) and (3.6), implies

$$x_a^r\{v_b^A, x_a^B\} - x_b^r\{v_b^A, x_A^B\} = 0, \tag{3.10}$$

whence

$$\{v_b^A, x_a^B\} = 0, \quad \text{for } a \neq b. \tag{3.11}$$

From (3.8), (3.10), and (3.11), we obtain

$$\{a_a^A, x_b^B\} = 0, \quad a \neq b, \tag{3.12}$$

which, introduced into (3.7), yields

$$\{v_a^A, v_b^B\} = 0, \quad a \neq b. \tag{3.13}$$

Now, using the Jacobi identity with v_a^A, v_b^B, and P_+, we can write

$$\{a_a^A, v_b^B\} = \{a_b^B, v_a^A\} + \{P_+, \{v_a^A, v_b^B\}\},$$

which, with the help of (3.13), leads to

$$\{a_a^A, v_b^B\} = \{a_b^B, v_a^A\}.$$

Upon substitution into (3.9), this finally yields

$$\{a_b^B, v_a^A\} = 0, \quad a \neq b. \tag{3.14}$$

Now taking into account the identity

$$\{f, g\} = \sum_{i,j=1}^{k} \{\eta_i, \eta_j\} \frac{\partial f}{\partial \eta_i} \frac{\partial g}{\partial \eta_j}, \tag{3.15}$$

where f and g depend on the variables $\eta_1 \cdots \eta_k$ only; we can write

$$\begin{aligned} \{x_a^A, a_b^B\} &= \{x_a^A, x_c^D\} \frac{\partial a_b^B}{\partial x_c^D} + \{x_a^A, v_c^D\} \frac{\partial a_b^B}{\partial v_c^D}, \\ \{v_a^A, a_b^B\} &= \{v_a^A, x_c^D\} \frac{\partial a_b^B}{\partial x_c^D} + \{v_a^A, v_c^D\} \frac{\partial a_b^B}{\partial v_c^D}, \end{aligned} \tag{3.16}$$

whence, by using Eqs. (3.10)–(3.14) there follows that

$$\frac{\partial a_b^B}{\partial v_c^D} = 0, \quad \frac{\partial a_b^B}{\partial x_c^D} = 0, \quad c \neq b. \tag{3.17}$$

That is, the acceleration a_b^B, $B = 1,2,-$, of each particle b does not depend on the positions and velocities of the remaining ones, but only on its own position x_b^A and velocity v_b^A. This conclusion would be enough to consider that the no-interaction result is proven, since the motion of each particle is not affected by the presence of the others. However, in the case we are considering (i.e., front form) a little bit deeper analysis reveals that the accelerations actually vanish.

Indeed, from (2.15b) and (3.17), we have

$$\frac{\partial a_a^B}{\partial x_a^A} = 0, \tag{3.18}$$

and using (2.16), (3.17), and (3.18), we arrive at

$$a_a^A = 0, \quad a = 1,\ldots, N, \quad A = 1,2,-, \tag{3.19}$$

which completes the proof.

IV. POINT FORM OF PREDICTIVE RELATIVISTIC MECHANICS

In the instant and front forms of dynamics, the construction of the configuration space was somehow linked to the choice of either the space hyperplanes $x^4 = t$ or the null ones $x^3 + x^4 = \lambda$, respectively. In the point form, the hyperboloids $x^\mu x_\mu = -\lambda^2$ will be assigned a similar role.

Each point in the extended configuration space will be characterized by $3N+1$ coordinates (x_a^i, λ), $a = 1,\ldots, N$, $i = 1, 2, 3$, where λ is taken as an evolution parameter and the x_a^i are the spacelike coordinates of the event where the worldline of the ath particle intersects the hyperboloid

$$x^\mu x_\mu = -\lambda^2. \tag{4.1}$$

As in the earlier two cases, the equations of motion are second-order differential equations

$$\frac{d^2 x_a^i}{d\lambda^2} = a_a^i(x_b^i, v_c^k, \lambda), \quad \frac{dx_a^i}{d\lambda} = v_a^i, \tag{4.2}$$

where $a, b, c = 1,\ldots, N$, $i, j, k = 1, 2, 3$.

Now, let $\varphi_a^i(x_b^j, v_c^k, \lambda_0, \lambda)$ be the general solution of (4.2) determined by the initial conditions

$$\begin{aligned} \varphi_a^i(x_b^j, v_c^k, \lambda_0; \lambda_0) &= x_a^i, \\ \dot{\varphi}_a^i(x_b^j, v_c^k, \lambda_0; \lambda_0) &= v_a^i, \end{aligned} \tag{4.3}$$

where an overdot means a partial derivative with respect to the parameter λ. As in the former two cases, the worldline of the ath particle is then defined by $\varphi_a^\mu(x_b^j, v_c^k, \lambda_0; \lambda)$, where

$$\varphi_a^0(x_b^j, v_c^k, \lambda_0; \lambda) = \left\{\lambda^2 + \sum_{i=1}^{3} \varphi_a^i(x_b^j, v_c^k, \lambda_0; \lambda)^2\right\}^{1/2}. \tag{4.4}$$

And, as before, we shall require Poincaré invariance of worldlines that reads as

$$\varphi_a^\mu(z_0', \lambda_a(z_0, \lambda)) = L^\mu{}_\nu [\varphi_a^\nu(z_0, \lambda) - A^\nu], \tag{4.5}$$

where z_0 and z_0' are abbreviations of the initial data, $(x_b^j, v_c^k, \lambda_0)$ and $(x_b'^j, v_c'^k, \lambda_0')$, respectively. These initial data correspond to two different inertial frames that are related to each other by the given Poincaré transformation $(L^\mu{}_\nu, A^\rho)$. The mapping $z_0 \to z_0'$ defines the induced transformation on the extended configuration space.

The value of the parameter $\lambda_a(z_0, \lambda)$ on the left-hand side of Eq. (4.5) can be easily derived. Indeed, taking (4.4) into account, we have

$$\lambda_a^2(z_0, \lambda) = -[\varphi_a^\nu(z_0, \lambda) - A^\mu][\varphi_{a\nu}(z_0, \lambda) - A_\nu]. \tag{4.6}$$

As is easily seen from this equation, and also from (4.4), the correspondence between λ and the time coordinate φ^0_a is not one-to-one. In order to avoid the nondifferentiability in the branch point $\lambda = 0$, we shall take hereafter λ_0 and λ positive. Moreover, the translation parameters A^μ will be assumed to be small enough for $\lambda^2_a(z_0, \lambda)$ on the left-hand side of Eq. (4.6) to remain positive.

By differentiating (4.5) with respect to λ we obtain the transformation formula for the velocities

$$L^\mu{}_\nu\, \dot\varphi^\nu_a(z_0, \lambda) = \dot\varphi^\mu_a(z'_0, \lambda_a(z_0, \lambda))\, \dot\lambda_a(z_0, \lambda), \tag{4.7}$$

where

$$\dot\lambda_a(z_0, \lambda) = [\lambda_a(z_0, \lambda)]^{-1}[\varphi^\nu_a(z_0, \lambda) - A^\nu]\, \dot\varphi_{a\nu}(z_0, \lambda). \tag{4.8}$$

And, differentiating again, we have the relationship among accelerations

$$L^\mu{}_\nu\, \ddot\varphi^\nu_a(z_0, \lambda) = \ddot\varphi^\mu_a(z'_0, \lambda_a(z_0, \lambda))\, \lambda^2_a(z_0, \lambda) + \dot\varphi^\mu_a(z'_0, \lambda_a(z_0, \lambda))\, \ddot\lambda_a(z_0, \lambda), \tag{4.9}$$

where

$$\ddot\lambda_a = -\frac{\dot\lambda^2_a}{\lambda_a} + \frac{1}{\lambda_a}\{(\varphi^\nu_a(z_0, \lambda) - A^\nu)\, \ddot\varphi_{a\nu} + \dot\varphi^\nu_a(z_0, \lambda)\, \dot\varphi_{a\nu}(z_0, \lambda)\}. \tag{4.10}$$

Now using (4.3) and (4.4), taking (4.6)and (4.8) into account, and setting $\lambda = \lambda_0$, Eqs. (4.5) and (4.7) yield a set of $6N$ implicit equations involving $x^i_a, v^j_b, x'^k_c, v'^l_d, \lambda_0, L^\mu{}_\nu, A^\rho$. Similarly to the front form case, it will be generally impossible to derive explicit expressions

$$x'^i_a = f^i_a(x, v, \lambda_0; L, A), \quad v'^i_a = g^i_a(x, v, \lambda_0; L, A)$$

for the action induced by a finite Poincaré transformation $(L^\mu{}_\nu, A^\rho)$. However, by introducing the infinitesmial expressions

$$L^\mu{}_\nu = \delta^\mu{}_\nu + \epsilon^{\alpha\beta}(\delta^\mu{}_\alpha\, \eta_{\nu\beta} - \delta^\mu{}_\beta\, \eta_{\nu\alpha}) + O(\epsilon^2), \quad A^\rho = \epsilon^\rho \tag{4.11}$$

into Eqs. (4.5) and (4.7), we can easily derive the following expressions from ten infinitesimal generators:

$$\mathbf{P}_0 = \sum_{a,i}\left\{\frac{(\lambda + \mathbf{x}^2_a)^{1/2}}{\lambda} v^i_a \frac{\partial}{\partial x^i_a} + \left[\frac{v^i_a}{(\lambda^2 + \mathbf{x}^2)^{1/2}}\left(1 + \frac{\mathbf{x}_a \cdot \mathbf{v}_a}{\lambda} - \frac{\lambda^2 + \mathbf{x}^2_a}{\lambda^2}\right) + \frac{(\lambda^2 + \mathbf{x}^2_a)^{1/2}}{\lambda} a^i_a\right]\frac{\partial}{\partial v^i_a}\right\},$$

$$\mathbf{P}_j = -\sum_{a,i}\left\{\left(\frac{x^j_a v^i_a}{\lambda} + \delta^i_j\right)\frac{\partial}{\partial x^i_a} + \frac{1}{\lambda}\left(x^j_a a^i_a + v^i_a v^j_a - \frac{v^i_a x^j_a}{\lambda}\right)\frac{\partial}{\partial v^i_a}\right\}, \tag{4.12}$$

$$\mathbf{J}_{0j} = \sum_a\left\{(\lambda^2 + \mathbf{x}^2)^{1/2}\frac{\partial}{\partial x^j_a} + \frac{\lambda + \mathbf{x}_a \cdot \mathbf{v}_a}{(\lambda^2 + \mathbf{x}^2)^{1/2}}\frac{\partial}{\partial v^j_a}\right\},$$

$$\mathbf{J}_{ij} = \sum_a\left\{x^j_a \frac{\partial}{\partial x^i_a} - x^i_a \frac{\partial}{\partial x^j_a} + v^j_a \frac{\partial}{\partial v^i_a} - v^i_a \frac{\partial}{\partial v^j_a}\right\}.$$

Finally, substituting the infinitesimal Poincaré transformation (4.11) into Eq. (4.9), and taking (4.10) and (4.12) into account, we arrive at

$$\mathbf{J}_{ij}\, a^k_a = a^j_a\, \delta^i_k - a^i_a\, \delta^j_k,$$

$$\mathbf{J}_{0l}\, a^k_a = \frac{\delta^k_l}{\sqrt{\lambda^2 + \mathbf{x}^2_a}}\left(1 + \mathbf{x}^2_a + \mathbf{x}_a\, \mathbf{a}_a - \frac{(\lambda + \mathbf{x}_a \mathbf{x}_a)^2}{\lambda^2 + \mathbf{x}^2_a}\right),$$

$$\mathbf{P}_l\, a^k_a = \frac{2}{\lambda^2}\left(x^l_a\, a^k_a + v^l_a\, v^k_a - \frac{x^l_a\, v^k_a}{\lambda}\right) - \frac{1}{\lambda}\left(a^l_a\, v^k_a + 2v^l_a\, a^k_a + x^l_a \frac{\partial a^k_a}{\partial \lambda}\right), \tag{4.13}$$

$$\mathbf{P}_0\, a^k_a = \frac{2}{\lambda(\lambda^2 + \mathbf{x}^2_a)^{1/2}}\left(\lambda + \mathbf{x}_a \mathbf{v}_a - \frac{(\lambda^2 + \mathbf{x}^2_a)}{\lambda}\right) a^k_a + \frac{1}{\lambda(\lambda^2 + \mathbf{x}^2_a)^{1/2}}\left(\mathbf{v}^2_a + \mathbf{x}_a \mathbf{a}_a - \frac{\mathbf{x}_a v_a}{\lambda} - \frac{(\lambda + \mathbf{x}_a \mathbf{v}_a)}{\lambda} + \frac{2(\lambda^2 + \mathbf{x}^2_a)}{\lambda^2} - \frac{(\lambda + \mathbf{x}_a \mathbf{v}_a)^2}{(\lambda^2 + \mathbf{x}^2_a)}\right) v^k_a + \frac{(\lambda^2 + \mathbf{x}^2_a)^{1/2}}{\lambda}\frac{\partial a^k_a}{\partial \lambda},$$

where

$$\mathbf{H} = \sum_{a,i}\left(v^i_a \frac{\partial}{\partial x^i_a} + a^i_a \frac{\partial}{\partial v^i_a}\right) + \frac{\partial}{\partial \lambda} \tag{4.14}$$

is the infinitesimal generator of λ evolution.

Equations (4.13) play a similar role as Currie-Hill conditions in the instant form of dynamics. It can be easily proven that they are equivalent to the following requirements.

(i) The generators $\mathbf{P}_\mu, \mathbf{J}_{\mu\nu}$ given by (4.12) generate a realization of the Poincaré algebra on the space of initial data (i.e., their commutation relations are the suitable ones).

(ii) This realization is invariant under λ evolution, that is,

$$[\mathbf{H}, \mathbf{P}_\mu] = [\mathbf{H}, \mathbf{J}_{\mu\nu}] = 0. \tag{4.15}$$

V. THE NO-INTERACTION THEOREM IN THE POINT FORM OF DYNAMICS

Let us now assume that by introducing some $3N$ momenta $p^i_a(\mathbf{x}_b, \mathbf{v}_c, \lambda)$, $a = 1,...,N$, $i = 1, 2, 3$, the extended configuration space can be mapped onto the extended phase space, spanned by the $6N + 1$ independent variables (x^i_a, p^j_b, λ). Let us further assume that the latter is endowed with the canonical structure defined by the elementary Poisson brackets

$$\{x^i_o, x^j_b\} = \{p^i_a, p^j_b\} = 0, \quad \{x^i_a, p^j_b\} = \delta_{ab}\, \delta^{ij} \tag{5.1}$$

and the Poincaré transformations as well as λ evolution are both canonical relative to this structure.

The latter condition implies the existence of 11 generating functions $H, P_\mu, J_{\mu\nu}, \mu, \nu = 1, 2, 3, 4$, such that

$$\mathbf{H} = \{H, \}, \quad \mathbf{P}_\mu = \{P_\mu, \}, \quad \mathbf{J}_{\mu\nu} = \{J_{\mu\nu}, \}, \tag{5.2}$$

where $\mathbf{H}$, $\mathbf{P}_\mu$, and $\mathbf{J}_{\mu\nu}$ are given by (4.14) and (4.12), respectively.

According to (4.12), (4.14), and (5.1), we have

$$\{H, x_a^i\} = v_a^i , \tag{5.3a}$$

$$\{P_0, x_a^i\} = [(\lambda^2 + \mathbf{x}_a^2)^{1/2}/\lambda\ ]\ v_a^i , \tag{5.3b}$$

$$\{P_j, x_a^i\} = -x_a^j\, v_a^i/\lambda - \delta_j^i , \tag{5.3c}$$

$$\{J_{jk}, x_a^i\} = \delta_j^i\, x_a^k - \delta_k^i\, x_a^j , \tag{5.3d}$$

$$\{J_{0j}, x_a^i\} = (\lambda^2 + \mathbf{x}_a^2)^{1/2}\, \delta_j^i . \tag{5.3e}$$

Now, writing down the Jocobi identity for P_j, x_a^i, x_b^k, taking (5.1) into account and using (5.2), we arrive at

$$x_a^j\ \{x_b^k, v_a^i\} = x_b^j\ \{x_a^i, v_b^k\} . \tag{5.4}$$

Applying the same treatment to the functions P_0, x_a^i, x_b^j, we have

$$\{x_b^j, v_a^i\} = \{x_a^i, v_b^j\} , \tag{5.5}$$

which, once introduced in (5.4), implies

$$\{x_b^j, v_a^i\} = 0, \quad a \neq b . \tag{5.6}$$

Then, starting from the Jacobi indentity for P_j, v_a^i, x_b^j and using (5.2), (5.3), and (5.6) we obtain

$$x_a^j\ \{x_b^k, a_a^i\} = x_b^j \{v_a^i, v_b^k\} , \tag{5.7}$$

which, substituted in the Jacobi identity corresponding to P_0, v_a^i, x_b^j leads us to

$$\{x_b^k, a_a^i\} = 0, \quad \{v_b^k, v_a^i\} = 0, \quad a \neq b . \tag{5.8}$$

Finally, using (5.8) in the Jacobi identity corresponding to P_j, v_a^i, v_b^k, we arrive at

$$\{v_b^k, a_a^i\} = 0, \quad a \neq b . \tag{5.9}$$

Since the Poisson bracket has rank $6N$ and the mapping $(x_a^i, v_b^j, \lambda) \to (x_a^i, P_b^j, \lambda)$ is assumed to have rank $6N+1$, there follows from (5.8) and (5.9) that

$$\frac{\partial a_b^j}{\partial x_a^i} = 0, \quad \frac{\partial a_b^j}{\partial v_a^i} = 0, \quad a \neq b , \tag{5.10}$$

or, equivalently,

$$a_b^j = a_b^j(x_b^i, v_b^k, \lambda) , \tag{5.10'}$$

which means that the acceleration of the b th particle does not depend on the state of motion of the remaining one. We can conclude that particles do not interact among themselves.

However, the no-interaction theorem we are intending to prove goes further still. Indeed, not only does it state that particles do not interact but also that their worldlines are straight.

The specific form of the acceleration a_b^i can be determined by introducing (5.10') into (4.13). This leads us to

$$a_b^i = \frac{1}{\lambda}\left[1 + \mathbf{v}_b^2 - \frac{(\lambda + \mathbf{x}_b \cdot \mathbf{v}_b)^2}{\lambda^2 + \mathbf{x}_b^2}\right] v_b^i , \tag{5.11}$$

which, since the acceleration is proportional to the velocity, implies that motions are rectilinear. A suitable reparametrization of trajectories will yield uniform motions, and therefore, the proof is complete.

In fact, the *a priori* knowledge that the above-mentioned suitable parameter will be the "physical time" φ_a^0, and its relationship (4.4) to the "unphysical" scalar parameter λ, will allow us to derive the general solution of Eq. (5.11). Indeed, we can easily prove that

$$\mathbf{c} + \mathbf{b}\, T(\lambda, \lambda_0, \mathbf{c}, \mathbf{b}), \tag{5.12}$$

with

$$T = (1 - \mathbf{b}^2)^{-1}\{\mathbf{b}\cdot\mathbf{c} - b_0 - ((\mathbf{b}\cdot\mathbf{c} - t_0)^2 + (\lambda^2 - \lambda_0^2)(1 - \mathbf{b}^2))^{1/2}\} \tag{5.13}$$

and

$$t_0 = (\lambda_0^2 - \mathbf{c}^2)^{1/2}$$

is the general solution of Eq. (5.11); the parameters $\mathbf{b}\cdot\mathbf{c}$ being related to the initial data $\underset{0}{\mathbf{x}_b}$ $\underset{0}{\mathbf{v}_b}$ according to

$$\mathbf{c} = \underset{0}{\mathbf{x}_b}, \quad \mathbf{b} = \frac{t_0}{\lambda_0 + \underset{0}{\mathbf{x}_b}\ \underset{0}{\mathbf{v}_b}} \cdot \underset{0}{\mathbf{v}_b} . \tag{5.14}$$

APPENDIX A

In this Appendix we present the most useful expressions concerning the action of the Poincaré group on the Minkowski space M_4, in terms of the set of coordinates and parameters which are most suitable for the null plane formalism (or front form). Throughout this paper, we take $c = 1$ and $\eta_{\mu\nu} = (+\ +\ +\ -)$, $\mu, \nu = 1, 2, 3, 4$.

If x_a^μ, $\mu = 1, 2, 3, 4$, are the Cartesian coordinates of an event in M_4, then the Poincaré transformation $(L^\mu{}_\nu, A^\mu)$ changes them into

$$x'^\mu = L^\mu{}_\nu\, (x^\nu - A^\nu) . \tag{A1}$$

A proper orthochronous Poincaré transformation $(L^\mu{}_\nu, A^\nu)$ is characterized by ten parameters $(\epsilon^\lambda, \epsilon^{\mu\nu})$, $\lambda, \mu, \nu = 1, 2, 3, 4$, $\mu < \nu$. In the standard parametrization, and for infinitesimal values of these parameters, we have that

$$L^\alpha{}_\beta = \delta^\alpha{}_\beta + \epsilon^{\mu\nu}\ (\delta^\alpha{}_\mu\, \eta_{\nu\beta} - \delta^\alpha{}_\nu\, \eta_{\mu\beta}) + O(\epsilon^2), \tag{A2}$$

$$A^\alpha = \epsilon^\mu\, \delta^\alpha{}_\mu .$$

Hence, in Cartesian coordinates, and in the standard parametrization, the infinitesimal generators are

$$\mathbf{P}_\mu = -\frac{\partial}{\partial x^\mu}, \quad \mathbf{J}_{\mu\nu} = x_\nu \frac{\partial}{\partial x^\mu} - x_\mu \frac{\partial}{\partial x^\nu} . \tag{A3}$$

In the front form, it is more convenient to use the new adapted coordinates

$$x^{\bar{A}} = M^{\bar{A}}{}_\mu\, x^\mu, \quad \bar{A} = 1, 2, +, - , \tag{A4}$$

where

$$x^+ = x^3 + x^4, \quad x^- = \tfrac{1}{2}(x^3 + x^4), \quad x^1 = x^1, \quad x^2 = x^2 , \tag{A5}$$

that is,

$$M^{\bar{A}}{}_\mu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1/2 & -1/2 \end{pmatrix} .$$

In these coordinates, the Minkowski metric is given by

$$\eta_{\overline{AB}} = (M^{-1})^{\mu}{}_{\overline{A}}\,(M^{-1})^{\nu}{}_{\overline{B}}\,\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{A6}$$

From this expression it is obvious that

$$M_{\overline{A}}{}^{\mu} \equiv \eta_{\overline{AB}}\, M^{\overline{B}}{}_{\nu}\, \eta^{\mu\nu} = (M^{-1})^{\mu}{}_{\overline{A}}\,. \tag{A7}$$

So, the lowering and raising of indices works as

$$M^{1} = M_{1}, \quad M^{2} = M_{2}, \quad M^{+} = M_{-}, \quad M^{-} = M_{+}\,. \tag{A8}$$

Expressed in these new coordinates, the transformation (A1) reads

$$x'^{\overline{A}} = \mathscr{L}^{\overline{A}}{}_{\overline{B}}\,(x^{\overline{B}} - \mathscr{A}^{\overline{B}}), \quad \overline{A}, \overline{B} = 1, 2, +, -\,, \tag{A9}$$

where

$$\mathscr{L}^{\overline{A}}{}_{\overline{B}} \equiv M^{\overline{A}}{}_{\mu}\, L^{\mu}{}_{\nu}\, (M^{-1})^{\nu}{}_{\overline{B}}, \quad \mathscr{A}^{\overline{B}} \equiv M^{\overline{B}}{}_{\nu}\, \overline{A}^{\nu}\,. \tag{A10}$$

According to these definitions, and performing the change of parameters,

$$\epsilon^{\overline{A}} = M^{\overline{A}}{}_{\mu}\, \epsilon^{\mu}, \quad \epsilon^{\overline{AB}} = \epsilon^{\mu\nu}\, M^{\overline{A}}{}_{\mu}\, M^{\overline{B}}{}_{\nu}\,, \tag{A11}$$

the infinitesimal expressions (A2) transform into

$$\mathscr{L}^{\overline{A}}{}_{\overline{B}} = \delta^{\overline{A}}{}_{\overline{B}} + \epsilon^{\overline{CD}}(\delta^{\overline{A}}{}_{\overline{C}}\, \eta_{\overline{DB}} - \delta^{\overline{A}}{}_{\overline{D}}\, \eta_{\overline{CB}}) + O(\epsilon^{2})\,,$$
$$\mathscr{A}^{\overline{A}} = \epsilon^{\overline{C}}\, \delta^{\overline{A}}_{\overline{C}}\,. \tag{A.12}$$

Hence, the corresponding generators are

$$\mathbf{P}_{\overline{A}} = -\frac{\partial}{\partial x^{\overline{A}}}, \quad \mathbf{J}_{\overline{AB}} = x_{\overline{B}}\,\frac{\partial}{\partial x^{\overline{A}}} - x_{\overline{A}}\,\frac{\partial}{\partial x^{\overline{B}}}\,. \tag{A13}$$

The relationship between the two sets of generators (A3) and (A13) can be easily derived from their definitions and Eq. (A11); this relationship being

$$\mathbf{P}_{\overline{A}} = M_{\overline{A}}{}^{\mu}\, \mathbf{P}_{\mu}, \quad \mathbf{J}_{\overline{AB}} = M_{\overline{A}}{}^{\mu}\, M_{\overline{B}}{}^{\nu}\, \mathbf{J}_{\mu\nu}\,. \tag{A14}$$

Notice that the coordinates $x^{\overline{A}}$, $\overline{A} = 1, 2, +, -$, are specially suitable to work in the instant form approach, since the null hyperplane equation $x^{3} = x^{4} = 0$ is written in the new coordinates $x^{+} = 0$. Moreover, the generators $\mathbf{P}_{\overline{A}}$ and $\mathbf{J}_{\overline{AB}}$ split in a natural way into kinematic ones (those preserving the null hyperplane)

$$\mathbf{P}_{-}, \mathbf{P}_{1}, \mathbf{P}_{2}, \mathbf{J}_{12}, \mathbf{J}_{1-}, \mathbf{J}_{2-}, \mathbf{J}_{+-}$$

and dynamic ones

$$\mathbf{P}_{+}, \mathbf{J}_{1+}, \mathbf{J}_{2+}\,.$$

APPENDIX B

In the future sheet of the light cone in Minkowski space defined by

$$x^{4} > 0, \quad x^{\mu} x_{\mu} < 0\,, \tag{B1}$$

we introduce the following coordinates:

$$y^{i} = x^{i}, \quad i = 1, 2, 3, \quad y^{4} = (-x^{\mu} x_{\mu})^{1/2}\,. \tag{B2}$$

In terms of these coordinates, the generators of infinitesimal Poincaré transformations are

$$\mathbf{P}_{\mu} = -\frac{((y^{4})^{2} + \sum_{i} (y^{i})^{2})^{1/2}}{y^{4}}\,\frac{\partial}{\partial y^{4}}\,,$$
$$\mathbf{P}_{i} = \frac{y^{i}}{y^{4}}\,\frac{\partial}{\partial y^{4}} - \frac{\partial}{\partial y^{i}}\,,$$
$$\mathbf{J}_{0i} = \left((y^{4})^{2} + \sum_{j} (y^{j})^{2}\right)^{1/2}\frac{\partial}{\partial y^{i}}\,, \tag{B3}$$
$$\mathbf{J}_{ij} = y^{j}\,\frac{\partial}{\partial y^{i}} - y^{i}\,\frac{\partial}{\partial y^{j}}\,.$$

[1] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[2] L. H. Thomas, Phys. Rev. **85**, 868 (1952); B. Bakamjian and L. H. Thomas, Phys. Rev. **92**, 1300 (1953); L. L. Foldy, Phys. Rev. **122**, 275 (1961).
[3] D.G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); H. Leutwyler, Nuovo Cimento **37**, 556 (1965).
[4] D. G. Currie, Phys. Rev. **142**, 817 (1966); R. N. Hill, Math. Phys. **8**, 201 (1967); E. H. Kerner, Math. Phys. **9**, 222 (1968). All the above references can be found in *The Theory of Action at a Distance in Relativistic Particle Dynamics*, edited by E. H. Kerner (Gordon and Breach, New York, 1972). Also, P. Droz-Vincent, Phys. Scr. **2**, 129 (1970); L. Bel, Ann. Inst. H. Poincaré **XII**, 307 (1970); **XIV**, 189 (1971).
[5] R. N. Hill, Math. Phys. **8**, 1756 (1967).
[6] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[7] H. Leutwyler and J. Stern, Ann. Phys. **112**, 94 (1978).
[8] L. Bel, Ann. Inst. H. Poincaré **XII**, 307 (1970); **XIV**, 189 (1971).

Realization of Poincaré group induced by a second-order ordinary differential system. Noninteraction theorem

X. Jaen and J. Llosa
Grup de Relativitat de la Secció de Física (SCC) IEC, Departament de Física Teòrica, Universitat de Barcelona, 08028 Barcelona, Spain

F. Marqués
Grup de Relativitat de la Secció de Física (SCC) IEC, Universitat Politècnica de Catalunya (ETSECCP), 08028 Barcelona, Spain

A. Montoto
Grup de Relativitat de la Secció de Física (SCC) IEC, Departament de Física Teòrica, Universitat de Barcelona, 08028 Barcelona, Spain

(Received 31 January 1985; accepted for publication 16 October 1985)

A generalization of the predictive relativistic mechanics is studied where the initial conditions are taken on a general hypersurface of M^4. The induced realizations of the Poincaré group are obtained. The same procedure is used for the Galileo group. Noninteraction theorems are derived for both groups.

I. INTRODUCTION

The generalization of the no-interaction theorem presented here is undertaken in the framework of predictive relativistic mechanics (PRM),[1] that is, the same Newtonian equations of motion remain valid for every inertial observer. Relativistic invariance is then understood as referring to worldlines, thus adopting the standpoint first stated by Minkowski[2] that "...physical laws might find their most perfect expression as reciprocal relations between these worldlines."

In the usual formulation of PRM, the configuration space for an N-point particle system is spanned by the $3N$ simultaneous position coordinates of the particles (simultaneity here presumes an inertial observer describing the dynamics of the system). In this respect, the usual PRM approach is similar to the *instant form* of Dirac[3] for Hamiltonian relativistic dynamics.

Also, in most of the various derivations of the noninteraction theorem, the *instant form* assumption plays a crucial role: the simultaneous position of particles are assumed to be either canonical coordinates in the Hamiltonian approaches[4] or the variables spanning the configuration space in the Lagrangian formulations.[5]

Other derivations of the theorem, that will not be considered here, correspond to the covariant formalism of PRM (see Ref. 6) or to the Hamiltonian relativistic systems with constraints approach.[7] An interesting review on the subject can be found in Ref. 8.

Dirac proposed[3] two other possible formulations of Hamiltonian relativistic dynamics besides the *instant form*: namely, the *front form* and the *point form*. So the question arose of whether the instant form assumption was essential to the noninteraction result, or if a similar output could be obtained in the framework of the other two Dirac forms. This point has been studied in a recent work[9] and the answer is yes.

At this point, why should we restrict ourselves to the three Dirac forms of Hamiltonian relativistic dynamics? From a historical point of view, it becomes apparent that Dirac proposed these three forms as different possible ways of simplification for a wider problem: the derivation of Poisson realizations for the Poincaré algebra. Nevertheless, Dirac himself proposed, and later developed, a technique—relativistic Hamiltonian dynamics with constraints[10]—which permits us to obtain a much wider solution to this problem, beyond the rigid restrictions of the above-mentioned three forms of dynamics.

So, as far as the noninteraction result is concerned, the following question becomes legitimate: Does it hold beyond the narrow framework of the three forms of Dirac?

A first answer to this question has been given partially[7] in the framework of Hamiltonian relativistic systems with constraints. Under some assumptions it has been proved by an example that some mass-shell constraints and fixations could be chosen such that, albeit positions of particles are taken as canonical coordinates, the noninteraction implications are circumvented. However, that model has not any interest beyond the mathematical one: not only is it unphysical, but also the procedure to reconstruct the particles' worldlines from the phase space trajectories is rather sophisticated, owing to the fact that the fixations are chosen not to have a clear kinematical meaning, but to yield some wanted specific Dirac brackets.

We are going to undertake another generalization of the noninteraction theorem, now always keeping in mind the kinematic aspects of the problem, that is, the way any given inertial observer will have to recover the particles' worldlines from the configuration space trajectories. To this end, let us analyze how it is done in the *instant form* approach. For every given value λ of a certain parameter, an inertial observer takes the space coordinates $x_a^i(\lambda)$, $a = 1,...,N$, of each particle when

$$x_1^0 = x_2^0 = \cdots = x_N^0 = \lambda, \tag{1.1}$$

and the configuration space curve $(x_1^i(\lambda),...,x_N^i(\lambda))$ describes the evolution of the system. Conversely, the N worldlines are recovered from a given configuration space trajectory

 0022-2488/86/020519-05$02.50 © 1986 American Institute of Physics

$(x_1^i(\lambda),\ldots,x_N^i(\lambda))$, by merely taking $(\lambda,x_a^i(\lambda))$, $a=1,\ldots,N$.

Other requirements in PRM are (i) that the trajectories of the system are the integrals of a second-order differential system on the configuration space and (ii) the relativistic invariance in terms of worldlines. A further development of the latter two conditions yields the so-called *induced realization* of the Poincaré group on the cophase space and the Currie–Hill equations.[1]

The *front form* (resp. *point form*) approach operates in a quite similar way.[9] For every given value λ of a certain parameter, an inertial observer takes the space coordinates $x_a^i(\lambda)$, $a=1,\ldots,N$ when

$$x_a^0+x_a^3=\lambda \tag{1.2}$$

$$(\text{resp. } x_a^{02}-\mathbf{x}_a^2=\lambda^2). \tag{1.3}$$

The evolution of the system in the configuration space is then given by $(x_1^i(\lambda),\ldots,x_N^i(\lambda))$. Conversely, for a certain configuration space curve, the worldline of the ath particle is given by

$$(\lambda-x_a^3(\lambda),x_a^i(\lambda)),\quad a=1,\ldots,N, \tag{1.4}$$

$$\left[\text{resp. } (\sqrt{\lambda^2+\mathbf{x}_a^2(\lambda)},x_a^i(\lambda)),\ a=1,\ldots,N\right]. \tag{1.5}$$

Similarly, the requirements of relativistic invariance and that configuration space trajectories fulfill the second-order differential system

$$\frac{d^2x_a^i}{d\lambda^2}=a_a^i(\mathbf{x}_b,\mathbf{v}_c,\lambda),$$

also lead to an induced realization of Poincaré group on the extended cophase space and to some restrictions on the accelerations that play the same role as the Currie–Hill conditions do in the *instant form* approach.

In Sec. II, we shall start from an analysis of the common features of these three approaches, in order to generalize the predictive relativistic mechanics framework. Then, in Sec. III, we prove a generalization of the noninteraction theorem. Finally, in Sec. IV, we extend the previous study to Newtonian dynamics, analyzing the noninteraction theorem in this case.

II. THE GENERALIZED PREDICTIVE RELATIVISTIC MECHANICS FRAMEWORK

The three approaches we commented on at the end of last section (resp. instant, front, and point forms) share the following common features.

(i) *Newtonian equations of motion:* The configuration space of the N-point particle system is spanned by the $3N$ position coordinates of the particles x_a^i, $a=1,\ldots,N$, $i=1,2,3$, and the evolution is governed by a second-order differential system

$$\frac{dx_a^i}{d\lambda}=v_a^i,\quad \frac{dv_a^i}{d\lambda}=a_a^i(x_b^j,v_c^k,\lambda), \tag{2.1}$$

whose functional form does not depend on the inertial observer describing the dynamics.

(ii) *A specific rule to construct the worldlines:* For each particular solution

$$x_a^i(\lambda)=\varphi_a^i(\overset{*}{x}{}_b^j,\overset{*}{v}{}_c^k,\lambda_0;\lambda),\quad a=1,\ldots,N, \tag{2.2}$$

of (2.1) with the initial condition

$$\varphi_a^i(\overset{*}{x},\overset{*}{v},\lambda_0;\lambda_0)=\overset{*}{x}{}_a^i,\quad \frac{\partial\varphi_a^i}{\partial\lambda}(\overset{*}{x},\overset{*}{v},\lambda_0;\lambda_0)=\overset{*}{v}{}_a^i, \tag{2.3}$$

the worldline $x_a^\mu(\lambda)$ for the ath particle is obtained by taking the space coordinates (2.2) and finding the time coordinate $x_a^0(\lambda)\equiv\varphi_a^0(\overset{*}{x}{}_b^j,\overset{*}{v}{}_c^k,\lambda_0;\lambda)$ from

$$h(x_a^\mu(\lambda))=\lambda, \tag{2.4}$$

where $h(x^\mu)$ is a fixed function on space-time M_4—the same function for every inertial observer and every particle [remember Eqs. (1.1), (1.2), and (1.3) for the three forms commented on above].

Intuitively, this means that each inertial observer constructs his configuration space by taking the space coordinates of all particles when their worldlines cross a previously chosen parameterized set of space-time hypersurfaces: $h(x^\mu)=\lambda$.

In order that $x_a^0(\lambda)$ can be obtained from the implicit equation (2.4), the partial derivative $\partial h/\partial x^0$ must not vanish.

(iii) *Relativistic invariance of worldlines:* Let $\mathcal{S}$ and $\mathcal{S}'$ be two inertial observers connected by the Poincaré transformation

$$x'^\mu=L^\mu{}_\nu\cdot(x^\nu-A^\nu). \tag{2.5}$$

Let us assume that the worldlines $\varphi_a^\mu(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\lambda)$ are obtained by $\mathcal{S}$ from some initial data $(\mathbf{x}_b,\mathbf{v}_c,\lambda_0)$. Then, the transformed space-time curves

$$L^\mu{}_\nu\cdot[\varphi_a^\nu(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\lambda)-A^\nu] \tag{2.6}$$

must be such that are obtained by $\mathcal{S}'$ starting from another set of initial data $(\mathbf{x}'_a,\mathbf{v}'_b,\lambda_0)$. That is,

$$L^\mu{}_\nu\cdot[\varphi_a^\nu(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\lambda)-A^\nu]=\varphi_a^\mu(\mathbf{x}'_b,\mathbf{v}'_c,\lambda_0;\lambda'_a), \tag{2.7}$$

for every λ. The parameter λ'_a of the right-hand side is determined by the condition

$$\lambda'_a(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;L^\mu{}_\nu,A^\mu;\lambda)=h(L^\mu{}_\nu\cdot[\varphi_a^\nu(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\lambda)-A^\nu]), \tag{2.8}$$

which stems from requiring (2.4) to hold also for $\mathcal{S}'$.

It can be easily obtained from (2.4) that

$$\lambda'_a(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\delta^\mu_\nu,0;\lambda)=\lambda. \tag{2.9}$$

The new $\mathcal{S}'$ initial data $(\mathbf{x}'_a,\mathbf{v}'_b,\lambda_0)$ will depend on the former $\mathcal{S}$ ones $(\mathbf{x}_b,\mathbf{v}_c,\lambda_0)$ and on the Poincaré transformation, $(L^\mu{}_\nu,A^\mu)\in\mathcal{P}$, which relates $\mathcal{S}'$ to $\mathcal{S}$. That is,

$$x_a^{i\prime}=f_a^i(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;L^\mu{}_\nu,A^\mu),$$
$$v_a^{i\prime}=g_a^i(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;L^\mu{}_\nu,A^\mu). \tag{2.10}$$

We are not going to derive explicit expressions for these functions, f_a^i and g_b^j, which define the Poincaré transformation induced by the given $(L^\mu{}_\nu,A^\mu)\in\mathcal{P}$ on the extended cophase space $\Gamma(6N+1)$, nor are we going to prove by a direct manipulation that they form an actual group realization. Instead, a close examination of commutation relations will ultimately prove this point.[11]

Note that induced Poincaré transformations act as

$$(\mathbf{x}_a,\mathbf{v}_b,\lambda_0)\to(\mathbf{x}'_a,\mathbf{v}'_b,\lambda_0)$$

thus leaving invariant, *by prescription*, the sheets $\lambda=$ const,

of $\Gamma(6N+1)$. In any parametrization $(L^{\mu}{}_{\nu}(\epsilon_I), A^{\mu}(\epsilon_J))$, $I,J=1,\ldots,10$, the infinitesimal generators for induced Poincaré transformations are given by

$$\Lambda_I^* = \sum_a \left[\left(\frac{\partial f_a^i}{\partial \epsilon_I} \right)_{(\epsilon)=0} \cdot \frac{\partial}{\partial x_a^i} + \left(\frac{\partial g_a^i}{\partial \epsilon_I} \right)_{(\epsilon)=0} \cdot \frac{\partial}{\partial v_a^i} \right] \tag{2.11}$$

[summation over repeated space (Roman) or space-time (Greek) indices will be hereafter understood].

The coefficients on the right-hand side of (2.11) can be obtained by taking partial derivatives with respect to ϵ_I and then making $(\epsilon_J)=(0)$ in the expression

$$\begin{aligned} &\varphi_a^{\mu}(f_b^i(\mathbf{x},\mathbf{v},\lambda_0;\epsilon_I), g_c^j(\mathbf{x},\mathbf{v},\lambda_0;\epsilon_I),\lambda_0; \\ &\quad h(L^{\rho}{}_{\sigma}\cdot[\varphi_a^{\sigma}(\mathbf{x},\mathbf{v},\lambda_0;\lambda) - A^{\sigma}])) \\ &\quad = L^{\mu}{}_{\nu}\cdot[\varphi_a^{\nu}(\mathbf{x}_b,\mathbf{v}_c,\lambda_0;\lambda) - A^{\nu}], \end{aligned} \tag{2.12}$$

which results from considering (2.7), (2.8), and (2.10) together.

Since Eq. (2.12) holds for every value of λ—at least in an open neighborhood—it can be proved easily that any induced Poincaré transformation commutes with dynamical evolution. That is, the diagram

$$\begin{array}{ccc} (\mathbf{x}_{a0},\mathbf{v}_{b0},\lambda_0) & \xrightarrow{\text{induced Poincaré } (L^{\mu}{}_{\nu}, A^{\mu})} & (x'_{a0},v'_{b0},\lambda_0) \\ \Big\downarrow \text{dynamical evolution} & & \Big\downarrow \text{dynamical evolution} \\ (\mathbf{x}_{a1},\mathbf{v}_{b1},\lambda_1) & \xrightarrow{\text{induced Poincaré } (L^{\mu}{}_{\nu}, A^{\mu})} & (x'_{a1},v'_{b1},\lambda_1) \end{array}$$

is commutative.

In terms of infinitesimal generators, this condition is equivalent[11] to the vanishing of the Lie brackets

$$[\Lambda_I^*, \mathbf{D}] = 0, \tag{2.13}$$

where

$$\mathbf{D} = \sum_a \left[v_a^i \cdot \frac{\partial}{\partial x_a^i} + a_a^i(\mathbf{x},\mathbf{v},\lambda_0)\cdot\frac{\partial}{\partial v_a^j} \right] + \frac{\partial}{\partial \lambda_0} \tag{2.14}$$

is the infinitesimal generator of dynamical evolution on $\Gamma(6N+1)$.

In order to find out the coefficients of the generators Λ_I^*, we infer from (2.12) that

$$\Lambda_I^* x_a^i = C_{I0}^i \cdot \varphi_a^0 + C_{Ij}^i \cdot x_a^j - C_I^i - v_a^i \cdot (\Lambda_I h)(\mathbf{x}_a, \varphi_a^0), \tag{2.15}$$

where

$$C_{I\nu}^{\mu} \equiv \left(\frac{\partial L^{\mu}{}_{\nu}}{\partial \epsilon_I} \right)_{(\epsilon)=0}, \quad C_I^{\mu} \equiv \left(\frac{\partial A^{\mu}}{\partial \epsilon_I} \right)_{(\epsilon)=0}, \tag{2.16}$$

$\varphi_a^0 \equiv \varphi_a^0(\mathbf{x}_a,\lambda_0)$ is the solution of

$$h(\mathbf{x}_a, \varphi_a^0) = \lambda_0, \tag{2.17}$$

and the meaning of $(\Lambda_I h)$ is explained in detail in the Appendix.

Second, from (2.13) and (2.14) we have that

$$\Lambda_I^* v_a^i = \Lambda_I^*(\mathbf{D}x_a^i) = \mathbf{D}(\Lambda_I^* x_a^i),$$

and therefore

$$\begin{aligned} \Lambda_I^* v_a^i &= C_{I0}^i \cdot (\mathbf{D}\varphi_a^0) + C_{Ij}^i \cdot v_a^j - v_a^i \cdot \mathbf{D}((\Lambda_I h)(\mathbf{x}_a, v_a^0)) \\ &\quad - a_a^i(\mathbf{x}_b,\mathbf{v}_c,\lambda_0)\cdot(\Lambda_I h)(\mathbf{x}_a,\varphi_a^0). \end{aligned} \tag{2.18}$$

Finally, by applying the commutator (2.13) to v_a^i, we obtain

$$\begin{aligned} \Lambda_I^* a_a^i &= \mathbf{D}\Lambda_I^* v_a^i \\ &= C_{I0}^i \cdot \mathbf{D}^2\varphi_a^0 + C_{Ij}^i \cdot \mathbf{a}_a^j - (\mathbf{D}a_a^i)\cdot(\Lambda_I h) \\ &\quad - 2\mathbf{D}(\Lambda_I h)\cdot a_a^i - v_a^i \cdot \mathbf{D}^2(\Lambda_I h), \end{aligned} \tag{2.19}$$

which must be understood as necessary conditions on the accelerations $a_a^i(\mathbf{x}_b,\mathbf{v}_c,\lambda)$ for the relativistic worldline invariance (2.7) to be accomplished. These conditions will play the same role as Currie–Hill equations[1] in the instant form of PRM.

At this point, we can prove that the commutation relations among the generators Λ_I^*, $I=1,\ldots,10$, are precisely those of the Poincaré group. First, after a simple but rather tedious calculation, we arrive at

$$[\Lambda_I^*, \Lambda_J^*] x_a^i = C_{IJ}^K \cdot \Lambda_K^* x_a^i. \tag{2.20}$$

Second, taking (2.13) and (2.20) into account, we have

$$[\Lambda_I^*, \Lambda_J^*] v_a^i = C_{IJ}^K \cdot \Lambda_K^* v_a^i, \tag{2.21}$$

and, since $\Lambda_I^* \lambda_0 = 0$, we finally obtain

$$[\Lambda_I^*, \Lambda_J^*] = C_{IJ}^K \cdot \Lambda_K^*, \tag{2.22}$$

where C_{IJ}^K, $I,J,K=1,\ldots,10$, stand for the structure constants of the Lie algebra of Poincaré. (See the Appendix.)

III. NONINTERACTION THEOREM

Let us now assume that there is a Poisson bracket structure[12] of maximum rank on $\Gamma(6N+1)$ such that we have the following.

(i) The coordinate λ_0 of $\Gamma(6N+1)$ is neutral relative to this Poisson bracket, i.e.,

$$\{\lambda_0, f\} = 0,$$

for every function on $\Gamma(6N+1)$.

(ii) Induced Poincaré transformations are canonical. That is, there exist ten generating functions $\Lambda_I^*(\mathbf{x},\mathbf{v},\lambda_0)$, $I=1,\ldots,10$, such that

$$\Lambda_I^* f = \{\Lambda_I^*, f\}, \tag{3.1}$$

for every function f.

(iii) The $3N$ position coordinates x_a^i can be complemented with $3N$ conjugated momenta $p_j^b(\mathbf{x},\mathbf{v},\lambda_0)$ thereby obtaining a set of canonical variables whose elementary Poisson brackets are

$$\{x_a^i, x_b^j\} = 0, \tag{3.2}$$

$$\{x_a^i, p_j^b\} = \delta_a^b \cdot \delta_j^i, \quad \{p_i^b, p_j^a\} = 0. \tag{3.3}$$

As is well known—the proof can be found in any treatise on advanced analytical mechanics[13]—Eq. (3.2) is the necessary and sufficient condition for the differential system (3.3) to have a solution.

Substituting f in Eq. (3.1) by either x_a^j or v_b^i, and taking (2.14)–(2.16) into account, we have

$$\{\Lambda_I^*, x_a^i\} = C_{I0}^i \cdot \varphi_a^0 + C_{Ij}^i x^j{}_a - C_I^i - v_a^i \cdot (\Lambda_I h)(\mathbf{x}_a, \varphi_a^0), \tag{3.4}$$

$$\{\Lambda_I^*, v_a^i\} = C_{I0}^i \cdot (\mathbf{D}\varphi_a^0) + C_{Ij}^i \cdot v_a^j - a_a^i \cdot (\Lambda_I h)(\mathbf{x}_a, \varphi_a^0) - v_a^i \cdot \mathbf{D}((\Lambda_I h)(\mathbf{x}_a, \varphi_a^0)). \tag{3.5}$$

Since λ_0 is a neutral function relatively to the Poisson bracket, taking (3.2) into account and using a known property of Poisson brackets, we can write

$$\{x_a^i, a_b^j\} = \{x_a^i, v_c^k\} \frac{\partial a_b^j}{\partial v_c^k}, \tag{3.6}$$

$$\{v_a^i, a_b^j\} = \{v_a^i, x_c^k\} \frac{\partial a_b^j}{\partial x_c^k} + \{v_a^i, v_c^k\} \frac{\partial a_b^j}{\partial v_c^k}. \tag{3.7}$$

Then, by applying the Jacobi identity and taking (3.2) into account, we obtain

$$\{x_a^i, \{\Lambda_I^*, x_b^j\}\} + \{x_b^j, \{x_a^i, \Lambda_I^*\}\} = 0,$$

which, using (3.4), yields

$$[(\Lambda_I h)(\mathbf{x}_a, \varphi_a^0) - (\Lambda_I h)(\mathbf{x}_b, \varphi_b^0)] \cdot \{x_b^j, v_a^i\} = 0. \tag{3.8}$$

This finally implies

$$\{x_b^j, v_a^i\} = 0, \quad \forall a \neq b. \tag{3.9}$$

By repeating the same treatment with (3.9) instead of (3.2), we obtain

$$\{x_a^i, a_b^j\} + \{v_a^i, v_b^j\} \cdot (\Lambda_I h)(\mathbf{x}_a, \varphi_a^0) = 0.$$

Then, by considering together this expression and the one that results from interchanging the indices b and a, and by the same reasoning, that permitted us to pass from (3.8) to (3.9), we can write

$$\{x_a^i, a_b^j\} = \{v_a^i, v_b^j\} = 0, \quad \forall a \neq b, \tag{3.10}$$

unless $(\Lambda_I h)(\mathbf{x}_a, \varphi_a^0) = (\Lambda_I h)(\mathbf{x}_b, \varphi_b^0)$, for every $I = 1,...,10$.

Again, by repeating exactly the same procedure to Eq. (3.10), we obtain

$$\{v_a^i, a_b^j\} = 0, \quad \forall a \neq b. \tag{3.11}$$

Then, by substituting (3.9)–(3.11) into (3.6) and (3.7) and taking into account the fact that $(x_a^i, v_b^j, \lambda_0)$ is a complete set of independent variables (λ_0 being a neutral function), after some manipulation we have that

$$\frac{\partial a_b^j}{\partial x_a^i} = 0, \quad \frac{\partial a_b^j}{\partial v_a^i} = 0, \quad \forall a \neq b. \tag{3.12}$$

Therefore, the acceleration of a particle can only depend on the variables of the particle itself,

$$\mathbf{a}_b = \mathbf{a}_b(\mathbf{x}_b, \mathbf{v}_b, \lambda_0), \quad b = 1,...,N,$$

and, consequently, there is no interaction between particles. Moreover, these accelerations must be required to satisfy Eq. (2.19)—analogous to the Currie–Hill equation—which will imply further restrictions on them. Since the function $h(x^\mu)$ is unspecified, it is rather cumbersome to analyze in detail what these restrictions are like. However, once $h(x^\mu)$ is made explicit, the analysis is easier in the well-known three forms of Dirac—instant form[4] ($x^0 = \lambda$), front form[9] ($x_0 + x^3 = \lambda$), or point form[9] ($\lambda^2 = -x^\mu x_\mu$). Introducing (3.12) into (2.19), we obtain that accelerations must be parallel to velocity, that is, motions of particles are rectilinear and uniform.

The clue of what has been proved hitherto lies in the fact that for some generator of Poincaré group we have

$$(\Lambda_I h)(\mathbf{x}_a, \varphi_a^0) \neq (\Lambda_I h)(\mathbf{x}_b, \varphi_a^0), \quad a \neq b. \tag{3.13}$$

Indeed, let us assume that there is a function $h(x^\mu)$ such that

$$(\Lambda_I h)(\mathbf{x}_a, \varphi_a^0(\mathbf{x}_a, \lambda_0)) = (\Lambda_I h)(\mathbf{x}_b, \varphi_b^0(\mathbf{x}_b, \lambda_0)), \quad \forall I = 1,...,10,$$

or, according to Eq. (A2), that

$$h_\mu(\mathbf{x}_a, \varphi_a^0) \cdot [C_{Ij}^\mu x_a^j + C_{I0}^\mu \cdot \varphi_a^0 - C_I^\mu] = h_\mu(\mathbf{x}_b, \varphi_b^0) \cdot [C_{Ij}^\mu \cdot x_b^j + C_{I0}^\mu \cdot \varphi_b^0 - C_I^\mu], \tag{3.14}$$

which, in the case of translations, yields

$$h_\mu(\mathbf{x}_a, \varphi_a^0) = h_\mu(\mathbf{x}_b, \varphi_b^0),$$

the most general solution of which is

$$h(x^\mu) = a_\nu \cdot x^\nu + b. \tag{3.15}$$

The requirement that (3.44) is to be fulfilled in case of rotations, $I = (ij)$, restricts a little bit more the form of $h(x)$, specifically

$$h(x^\mu) = a_0 x^0 + b. \tag{3.16}$$

Finally, Eq. (31.4) specialized for boosts leads to

$$a_0 \cdot x_a^i = a_0 \cdot x_b^i, \quad a \neq b,$$

which implies that $a_0 = 0$.

We have arrive at $h(x^\mu) = b$, constant, which contradicts the previous assumption that

$$\frac{\partial h}{\partial x^0}(x^\mu) \neq 0.$$

Consequently, we have proved that is not possible in any way to choose a hypersurface-defining function $h(x^\mu)$ such that the noninteraction result can be avoided, provided that a canonical realization of the Poincaré algebra and the canonical character of position coordinates are simultaneously assumed.

IV. A "NONINTERACTION THEOREM" IN GALILEAN DYNAMICS

The three forms of Dirac for relativistic dynamics coincide when the Galilean limit ($c \to \infty$) is taken, thus leading to the natural form used by physicists for centuries. In spite of this, we shall devote this section to the seemingly academic task of extending the results formally obtained to Galilean dynamics. The outcome will be pleasantly surprising.

Since the results obtained in Secs. II and III have been derived in a rather generic way, similar results will hold for any transformation group of space-time, e.g., the Galilei group.

So, we can also speak of induced Galilean transformations on the extended cophase space and all that has been presented in Sec. II holds by changing "Poincaré" to "Galileo"—this change affects the coefficients $C_{I\nu}^\mu$ and C_I^μ and the structure constants C_{IJ}^K.

Also, as has been remarked at the end of Sec. III, in order to avoid the no-interaction result, we must find a function $h(\mathbf{x},t)$ on space-time fulfilling Eq. (3.14).

As we did in the last section, specializing (3.14) to space rotations and space-time translations, we obtain that h must be

$$h(\mathbf{x},t) = a \cdot t + b, \tag{4.1}$$

if (3.14) is to be fulfilled.

And, since the generators of Galilean boosts in space-time are $-t(\partial/\partial x^i)$, (3.14) is identically satisfied in this case.

We have therefore arrived at the interesting result that, even in Galilean dynamics, we could wind up with a noninteraction result if we had not chosen the suitable form—i.e., the function $h(\mathbf{x},t)$. However, contrary to the relativistic case, Galilean dynamics permits only one way—instant form—to escape from noninteraction, and this is precisely the one that has been naively used from the beginning of Hamiltonian classical mechanics.

V. CONCLUDING REMARKS

The starting point of the present paper has been that the equations of motion for an N-point particle system are second-order differential equations and that the configuration space of each inertial observer is constructed by taking particle positions when their worldlines cross a given space-time hypersurface $h(x^\mu) = \text{const}$. Relativistic invariance then has been imposed by requiring the equations of motion and the space-time hypersurface to have the same functional form for every inertial observer.

The outcome is a generalization of predictive relativistic mechanics,[1] which is recovered as a particular case of the framework here developed, by taking $h(x^\mu) = x_0$. And in this generalized framework we have obtained the conditions that accelerations must fulfill if worldline relativistic invariance is required. These conditions appear as the counterpart of the well-known Currie–Hill equations of predictive relativistic mechanics.

We have proved then that the requirement of having a canonical formulation for induced Poincaré transformations, where position coordinates can be taken as canonical ones, unavoidably implies noninteraction.

As a consequence of the general manner as the problem has been dealt with; the results obtained in the first part of the paper for the Poincaré group can be easily translated to the case of any other group of space-time transformations. We thus have written them for the case of the Galilei group, thus concluding, in Newtonian dynamics, the only way of avoiding noninteraction theorems in the usual "instant form," which always has been used as the natural one in classical Hamiltonian mechanics.

APPENDIX

For a given function $h(x^\mu)$ on the Minkowski space and for the standard realization of Poincaré group, we have

$$(\Lambda_I h)(x) \equiv (\partial_\mu h)(x)[C^\mu_{I\nu} x^\nu - C^\mu_I]. \tag{A1}$$

According to this, we define the shortened notation

$$(\Lambda_I h)_a \equiv (\Lambda_I h)(\mathbf{x}_a, \varphi^0_a)$$
$$\equiv (\partial_\mu h)(\mathbf{x}_a, \varphi^0_a)[C^\mu_{Ij} x^j_a + C^\mu_{I0}\varphi^0_a - C^\mu_I], \tag{A2}$$

where $\varphi^0_a(\mathbf{x}_a, \lambda_0)$ is obtained by solving (2.17).

On the other hand, in the usual parametrization of the Poincaré group

$$L^\mu{}_\nu = \delta^\mu{}_\nu + \omega^{\alpha\beta}\delta^\mu_\alpha \eta_{\nu\beta} + O(\omega^2), \quad \omega^{\alpha\beta} = -\omega^{\beta\alpha}, \tag{A3}$$
$$A^\mu = \epsilon^\rho \delta^\mu_\rho,$$
$$A^\mu = \epsilon^\rho \delta^\mu_\rho,$$

we have that

$$C^\mu_{(\alpha\beta)\nu} = \delta^\mu_\alpha \eta_{\nu\beta} - \delta^\mu_\beta \eta_{\nu\alpha}, \quad C^\mu_{(\rho)\nu} = 0, \tag{A4}$$
$$C^\mu_{(\alpha\beta)} = 0, \quad C^\mu_{(\rho)} = \delta^\mu_\rho.$$

[1] D. G. Currie, Phys. Rev. **142**, 817 (1966); R. N. Hill, J. Math. Phys. **8**, 201 (1967); L. Bel, Ann. Inst. H. Poincaré **12**, 307 (1970).
[2] H. Minkowski, "Space and time," in *The Principle of Relativity* (Dover, New York, 1923).
[3] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[4] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); H. Leutwyler, Nuovo Cimento **37**, 556 (1965); R. N. Hill, J. Math. Phys. **8**, 1756 (1967).
[5] E. H. Kerner, J. Math. Phys. **9**, 222 (1968); J. Martin and J. L. Sanz, J. Math. Phys. **19**, 780 (1978); J. Martin and J. L. Sanz, J. Math. Phys. **20**, 25 (1979).
[6] P. Droz-Vincent, Nuovo Cimento **128**, 1 (1972); L. Bel (private communication).
[7] A. Komar, Phys. Rev. D **18**, 3617 (1978); V. Iranzo, J. Llosa, F. Marques, A. Molina, and A. Montoto, Phys. Rev. D **31**, 314 (1985).
[8] S. Chelkowski, J. Nietendel, and R. Suchanek, Acta Phys. Pol. B **11**, 809 (1980).
[9] X. Jaen, A. Molina, and V. Iranzo, "Front form and point form fomulation of predictive relativistic mechanics. Non interaction theorems," preprint (UB-FT-12.84).
[10] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, New York, 1964).
[11] L. S. Pontrjagin, *Grupos Contínuos* (MIR, Moscow, 1978).
[12] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974), Chap. 9.
[13] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), Chap. VI.

Boson realization from quantum constraints

Mircea Iosifescu and Horia Scutaru
Department of Theoretical Physics, Central Institute of Physics, Bucharest, P.O. Box MG-6, Romania

(Received 21 January 1985; accepted for publication 16 October 1985)

Generalized Holstein–Primakoff realizations deduced by Deenen, Quesne, and Papanicolaou are obtained directly from the algebraic identities satisfied in collective subspaces by the infinitesimal generators of the corresponding dynamical groups.

I. INTRODUCTION

During the last few years it was possible to observe an increasing role played by dynamical groups in such fields of physics as the theory of collective motion in nuclei[1–12] or the study of $1/N$ expansions in quantum mechanics and field theory.[13,14]

In all these works, the construction of boson realizations of the various irreducible representations of dynamical Lie algebras proved to be very useful. Among all boson realizations, the Holstein–Primakoff realizations are the most interesting, because they are Hermitian and minimal, i.e., they utilize the minimum number of bosonic operators. Holstein–Primakoff realizations have been obtained by Deenen and Quesne[3] for the collective representations of the real-symplectic Lie algebras sp($2d,R$) and by Papanicolaou[14] for the orthogonal algebras so($2d;R$); for the Lie algebra sp($4;R$), the Holstein–Primakoff realization has been obtained by Mlodinow and Papanicolaou.[13]

It is the aim of the present paper to prove that Holstein–Primakoff realizations can be derived in a purely algebraic way. This fact had already been observed by Okubo,[15] who deduced a Holstein–Primakoff realization for the completely symmetric irreducible representations of the unitary Lie algebras u(d), utilizing the second-degree polynomial identities satisfied by their infinitesimal generators. In the present paper, we point out a method by which this approach can be extended to all dynamical Lie algebras and to all their finite- or infinite-dimensional representations with maximal weight vector, the generators of which satisfy second-degree polynomial identities.

The classical analog of the Holstein–Primakoff realization gives a method for an algebraic determination of the canonical coordinates on the symplectic manifold in the dual space of the dynamical Lie algebra, defined by the classical polynomial identities, which correspond to the quantum polynomial identities satisfied by the representations. This problem has been considered by Mukunda.[16]

In previous papers,[17,20] we obtained a classification of all second-degree polynomials on the dual space of a Lie algebra (or in the universal enveloping algebra of a Lie algebra), which can lead to identities in the classical case (in the quantum case). Using this result, we identified the second-degree polynomial identities among the generators of the collective subrepresentations considered by Deenen and Quesne[3] or by Papanicolaou.[14] Taking these identities as our starting point, we derived the fundamental equations obtained in Refs. 3 and 14 in a purely algebraic way.

II. DEFINITIONS AND NOTATIONS

We shall treat both algebras—the symplectic and the orthogonal—on an equal footing and study their boson realizations simultaneously. To do that, we shall use a unifying notation, by introducing a parameter ϵ defined by

$$\epsilon = \begin{cases} -1, & \text{for sp}(2d;R), \\ +1, & \text{for so}(2d;R). \end{cases} \tag{2.1}$$

Thus, let F_ϵ be the Fock space for a system with Nd degrees of freedom, on which representations of the canonical commutation ($\epsilon = -1$) or anticommutation ($\epsilon = +1$) relations are defined

$$\begin{aligned} &[b_{is}, b_{jt}]_\epsilon = [b_{is}^*, b_{jt}^*]_\epsilon = 0, \\ &[b_{is}, b_{jt}^*]_\epsilon = \delta_{ij}\delta_{st} I \quad (i,j = 1,\ldots,d;\ s,t = 1,\ldots,N). \end{aligned} \tag{2.2}$$

On F_ϵ, reducible skew-adjoint representations of sp($2d;R$) and so($2d;R$) are generated by

$$\begin{aligned} A_{ij} &= \sum_{s=1}^{N} b_{is}^* b_{js} - \epsilon(N/2)\delta_{ij} I, \\ B_{ij} &= \sum_{s=1}^{N} b_{is}^* b_{js}^*, \\ C_{ij} &= -\epsilon \sum_{s=1}^{N} b_{is} b_{js}. \end{aligned} \tag{2.3}$$

To simplify notations, we write $b_{is}, A_{ij}, B_{ij}, C_{ij}$, and I instead of $b_{is}^\epsilon, A_{ij}^\epsilon, B_{ij}^\epsilon, C_{ij}^\epsilon$, and I^ϵ, as a consequent use of convention (2.1) would require.

The structure relations for the two algebras will be written in the unified form

$$\begin{aligned} &[A_{ij}, A_{kl}]_{-1} = \delta_{jk} A_{il} - \delta_{il} A_{kj}, \\ &[A_{ij}, B_{kl}]_{-1} = \delta_{jk} B_{il} - \epsilon\delta_{jl} B_{ik}, \\ &[A_{ij}, C_{kl}]_{-1} = \epsilon\delta_{il} C_{jk} - \delta_{ik} C_{jl}, \\ &[B_{ij}, C_{kl}]_{-1} = -\delta_{jk} A_{il} - \delta_{il} A_{jk} \\ &\qquad\qquad + \epsilon\delta_{ik} A_{jl} + \epsilon\delta_{jl} A_{ik}. \end{aligned} \tag{2.4}$$

We shall use, for both algebras, the matrix notations

$$A = (A_{ij}), \quad B = (B_{ij}), \quad C = (C_{ij}) \quad (i,j = 1,\ldots,d), \tag{2.5}$$

i.e., A, B, and C are $d \times d$ matrices, the elements of which are the operators (2.3). We shall denote by M^t the transpose of M, and by M^* the matrix, the elements of which are the adjoints of the corresponding elements of M. With these notations, we have

 0022-2488/86/020524-05$02.50 © 1986 American Institute of Physics

$$B^t = -\epsilon B, \quad C^t = -\epsilon C, \quad A^t = A^*,$$
$$B^* = C, \quad C^* = B. \tag{2.6}$$

III. THE RESULTS OF DEENEN, QUESNE, AND PAPANICOLAOU

The representations (2.3) of sp(2d;R) and so(2d;R) are reducible on F_ϵ. The Fock space F_ϵ can be decomposed into subspaces in which the algebras act in an irreducible way. Such an irreducible submodule F_ϵ^c has been considered in Refs. 3, 13, and 14 (and called collective subspace in Ref. 3); F_ϵ^c is spanned by vectors of the form

$$B_{ij} B_{kl} \cdots B_{pq} v_\epsilon, \tag{3.1}$$

the vector $v_\epsilon \in F_\epsilon$ (called the vacuum state) being defined by

$$b_{ir} v_\epsilon = 0 \quad (i = 1,...,d; \quad r = 1,...,N). \tag{3.2}$$

The irreducible representation on F_ϵ^c, labeled by the positive integer N, is an infinite-dimensional representation [a spinorial representation with Dynkin indices $(0,...,0,N)$] of sp(2d;R) [of so(2d;R)], which belongs to the holomorphic discrete series representations associated with the Siegel half-plane.[21]

In Refs. 2, 3, 13, and 14, a bosonic Fock space $\mathscr{B}_\epsilon$ has been introduced, on which the following representation of the canonical commutation relations acts:

$$[a_{ij}, a_{kl}]_{-1} = [a_{ij}^*, a_{kl}^*]_{-1} = 0,$$
$$[a_{ij}\, a_{kl}^*]_{-1} = (\delta_{ik}\delta_{jl} - \epsilon\delta_{il}\,\delta_{jk})I, \tag{3.3}$$

where $a_{ij} = -\epsilon a_{ji}$, $a_{ij}^* = -\epsilon a_{ji}^*$ [i.e., we have $d(d-\epsilon)/2$ independent pairs of bosonic creation and annihilation operators]; the operators

$$A_{ij} = -\epsilon\left(\sum_{s=1}^{N} a_{is}^* a_{sj} + \frac{N}{2}\delta_{ij}\right),$$
$$B_{ij} = \sum_{s=1}^{N} X_{is} a_{sj}^* = \sum_{s=1}^{N} a_{is}^* Y_{sj}, \tag{3.4}$$
$$C_{ij} = \sum_{s=1}^{N} a_{is} X_{sj} = \sum_{s=1}^{N} Y_{is} a_{sj}$$
$$(i,j = 1,...,d),$$

defined on $\mathscr{B}_\epsilon$, generate a representation equivalent with the action of (2.3) on the collective subspace F_ϵ^c; $X_{ij}(Y_{ij})$ are $u(d)$-vector [$u(d)$-covariant vector] operators; the subalgebra $u(d)$ being generated by $A_{ij}(i,j = 1,...,d)$. Their expression results from the following Louck–Biedenharn theorem[22]: Any $u(d)$-vector or $u(d)$-covector operator is of the form

$$X_{ij} = \sum_{k=0}^{d-1} x_k(\mathrm{Tr}(a^*a),...,\mathrm{Tr}(a^*a)^d)(a^*a)_{ij}^k,$$

or (3.5)

$$Y_{ij} = \sum_{k=0}^{d-1} y_k(\mathrm{Tr}(a^*a)^t,...,\mathrm{Tr}((a^*a)^t)^d) \times((a^*a)^t)_{ij}^k,$$

respectively, where x_k (y_k) are functions having as arguments the $u(d)$-invariants $\mathrm{Tr}(a^*a)^k$ [$\mathrm{Tr}((a^*a)^t)^k$] [cf. (2.5)].

On an irreducible $u(d)$-submodule of $\mathscr{B}_{-1}$, x_k and y_k are scalar multiples of the unity operator. If we denote by x_k and y_k the corresponding scalars, the following equations hold:

$$\sum_{k=0}^{d-1} l_i^k x_k = \sqrt{l_i + N - d - 1}, \tag{3.6}$$

on the irreducible $u(d)$-submodules labeled by the Young indices $(f_1,...,f_d)$, where $l_i = f_i + d - i$ $(i = 1,...,d)$. The fundamental equations satisfied by the matrix $X = (X_{ij})(i,j = 1,...,d)$ has been obtained in Ref. 3 as a consequence of the explicit description summarized above; in matrix notation this equation is

$$X^2 = a^*a + (N - d - 1)I. \tag{3.7}$$

IV. THE ALGEBRAIC DETERMINATION OF THE HOLSTEIN–PRIMAKOFF REALIZATION

We shall first characterize the collective subspaces F_ϵ^c by the algebraic relations satisfied, on these spaces, by the generators of the representations of the algebras sp(2d;R) and so(2d;R).

Theorem 1: In the collective subspace F_ϵ^c, the generators of the representation of the algebras sp(2d;R) and so(2d;R) satisfy the second-degree polynomial identities

$$AB - BA^t = -\epsilon(AB - BA^t)^t, \tag{4.1}$$
$$CA - A^tC = -\epsilon(CA - A^tC)^t, \tag{4.2}$$
$$A^2 + ((A^t)^2)^t - BC - (CB)^t = (1/d)\mathrm{Tr}(A^2 + ((A^t)^2)^t - BC - (CB)^t)I = N((N/2) + \epsilon d - 1)I. \tag{4.3}$$

Proof: Let us denote

$$\mathscr{D} = AB - BA^t + \epsilon(AB - BA^t)^t, \tag{4.4}$$
$$\mathscr{E} = CA - A^tC + \epsilon(CA - A^tC)^t, \tag{4.5}$$
$$\mathscr{F} = A^2 + ((A^t)^2)^t - BC - (CB)^t - (1/d)\mathrm{Tr}(A^2 + ((A^t)^2)^t - BC - (CB)^t)I. \tag{4.6}$$

We have to prove that $\mathscr{D} = 0$, $\mathscr{E} = 0$, and $\mathscr{F} = 0$ on all vectors (3.1).

Let us observe that the commutators between the generators B_{ij} and the polynomials $\mathscr{D}_{kl}$, $\mathscr{E}_{kl}$, and $\mathscr{F}_{kl}$ are linear combinations of the polynomials $\mathscr{D}_{kl}$, $\mathscr{E}_{kl}$, and $\mathscr{F}_{kl}$:

$$[\mathscr{D}_{ij}, B_{kl}]_{-1} = 0,$$
$$[\mathscr{E}_{ij}, B_{kl}]_{-1} = -\epsilon\delta_{ki}\mathscr{F}_{lj} + \epsilon\delta_{lj}\mathscr{F}_{ki} + \delta_{li}\mathscr{F}_{kj} - \delta_{kj}\mathscr{F}_{li}, \tag{4.7}$$
$$[\mathscr{F}_{ij}, B_{kl}]_{-1} = \delta_{jk}\mathscr{D}_{il} - \epsilon\delta_{jl}\mathscr{D}_{ik}.$$

Hence, in order to prove that $\mathscr{D}_{kl} = \mathscr{E}_{kl} = \mathscr{F}_{kl} = 0$ on the vectors (3.1) it is sufficient to prove that $\mathscr{D}_{kl} v_\epsilon = \mathscr{E}_{kl} v_\epsilon = \mathscr{F}_{kl} v_\epsilon = 0$. This will be proved by direct computation. We have

$$A_{ij} v_\epsilon = -\epsilon(N/2)\delta_{ij} v_\epsilon, \tag{4.8}$$
$$C_{ij} v_\epsilon = 0, \tag{4.9}$$

whence

$$(AB - BA^t)v_\epsilon = (d - \epsilon)Bv_\epsilon, \tag{4.10}$$
$$(CA - A^tC)v_\epsilon = (d - \epsilon)Cv_\epsilon, \tag{4.11}$$
$$(A^2 + ((A^t)^2)^t - BC - (CB)^t)v_\epsilon = ((N/2) + \epsilon d - 1)Iv_\epsilon. \tag{4.12}$$

The operator $\mathrm{Tr}(A^2 + ((A^t)^2)^t - BC - (CB)^t)$ is the Casimir operator; hence

$$\mathrm{Tr}(A^2 + ((A^t)^2)^t - BC - (CB)^t) = dN((N/2) + \epsilon d - 1). \tag{4.13}$$

It follows that $\mathscr{D}v_\epsilon = \mathscr{E}v_\epsilon = \mathscr{F}v_\epsilon = 0$ and the proof is complete.

Remark 1: For the algebra $\mathrm{sp}(2d;R)$, another proof of Theorem 1 can be given, using a concrete realization obtained in Ref. 3 for the Hilbert subspace F_ϵ^c as a space of square integrable holomorphic functions of $d(d+1)/2$ collective variables $w_{ij}, i,j = 1,...,d$. In this concrete realization, the operators A_{ij}, B_{ij}, and C_{ij} are differential operators:

$$A = wD + (N/2)I, \quad B = w,$$

$$C = D(wD + (N - d - 1)I), \tag{4.14}$$

where $w = (w_{ij})$, $D = (1 + \delta_{ij})(\partial/\partial w_{ij})$ $(i,j = 1,...,d)$. By direct computation we obtain

$$AB - BA^t = (d+1)B, \quad CA - A^tC = (d+1)C,$$
$$A^2 + ((A^t)^2)^t - BC - (CB)^t = N((N/2) - d - 1)I, \tag{4.15}$$

whence the identities (4.1)–(4.3) follow immediately.

Remark 2: The equations between matrix elements obtained from the matrix equations (4.1)–(4.3) provide the polynomial identities associated with the subrepresentation Λ_2 of $(\mathrm{ad} \otimes \mathrm{ad})_s$ of the algebra $\mathrm{sp}(2d;R)$ (see Ref. 20) and to the subrepresentation $(2\Lambda_1)$ of $(\mathrm{ad} \otimes \mathrm{ad})_s$ of $\mathrm{so}(2d;R)$. (The Λ_i's denote fundamental weights.)

Our main result is that Eq. (3.7) and all properties of the operators X and Y can be obtained directly by imposing on the boson realization that the algebraic identities be satisfied by the collective representation.

Theorem 2: The generators of the boson realizations (3.4) satisfy the algebraic identities (4.1)–(4.3), which characterize collective representations, if and only if

$$X^2 = a^*a + (N + \epsilon d - 1)I, \tag{4.16}$$

and

$$Y^2 = (a^*a)^t + NI. \tag{4.17}$$

Proof: From the relations (3.4), which impose a form on the boson realization, and from

$$A^t = -\epsilon(aa^* + ((N/2) + \epsilon d - 1)I), \tag{4.18}$$

it follows that

$$Aa^* - a^*A^t = (d - \epsilon)a^*, \tag{4.19}$$

$$aA - A^ta = (d - \epsilon)a, \tag{4.20}$$

$$XA - AX = 0, \tag{4.21}$$

$$YA^t - A^tY = 0. \tag{4.22}$$

Hence

$$AB - BA^t = (d - \epsilon)B, \tag{4.23}$$

$$CA - A^tC = (d - \epsilon)C. \tag{4.24}$$

From $B^t = -\epsilon B$, $C^t = -\epsilon C$, (4.23) and (4.24) we get (4.1) and (4.2). Finally, from the identities (4.3) and the structure relations (2.4) we obtain the equations (4.16) and (4.17). Indeed, from the commutation relations for A_{ij} it follows that

$$((A^t)^2)^t = A^2 - dA + (\mathrm{Tr}\,A)I, \tag{4.25}$$

and from the commutation relations between B_{ij} and C_{ij} it follows that

$$(CB)^t = BC + (d - 2\epsilon)A + (\mathrm{Tr}\,A)I. \tag{4.26}$$

From relations (3.4) we obtain

$$BC = Xa^*aX = -X^2(\epsilon A + (N/2)I). \tag{4.27}$$

The identities (4.3) are now equivalent to

$$(-\epsilon A + ((N/2) + \epsilon d - 1)I)(\epsilon A + (N/2)I) = X^2(\epsilon A + (N/2)I), \tag{4.28}$$

whence

$$X^2 = -\epsilon A + ((N/2) + \epsilon d - 1)I = a^*a + (N + \epsilon d - 1)I. \tag{4.29}$$

Similarly

$$Y^2 = -\epsilon A^t + (N/2)I = (a^*a)^t + NI, \tag{4.30}$$

and the proof is complete.

To obtain the Holstein–Primakoff realizations, we must solve Eqs. (4.16) and (4.17). The operators X_{ij}, $i,j = 1,...,d$ (Y_{ij}, $i,j = 1,...,d$) form a $u(d)$-vector operator [a $u(d)$-covariant vector operator][3,23]; the theorem of Louck and Biedenharn[22] therefore can be applied in a form due to Okubo[15]: The vector (covariant vector) operators I, a^*a, $(a^*a)^2,...,(a^*a)^{d-1}$ [I, $(a^*a)^t$, $((a^*a)^t)^2,...,((a^*a)^t)^{d-1}$] are linearly independent and $(a^*a)^d$ [$((a^*a)^t)^d$] is the linear combination of these operators, given by

$$(a^*a - l_1I)(a^*a - l_2I)\cdots(a^*a - l_dI) = 0, \tag{4.31}$$

$$[((a^*a)^t - \tilde{l}_1I)((a^*a)^t - \tilde{l}_2I)\cdots((a^*a)^t - \tilde{l}_dI) = 0], \tag{4.32}$$

where, as previously, $l_i = f_i + d - i$ ($\tilde{l}_i = -f_{d-i+1} + d - i$), and $(f_1,f_2,...,f_d)$ are Young indices for irreducible representations of $u(d)$. In fact, we must consider that $l_1,l_2,...,l_d$ are functions of the $u(d)$-invariant operators $\mathrm{Tr}(a^*a),...,\mathrm{Tr}((a^*a)^d)$; in each irreducible representation of $u(d)$ these functions are scalar multiples of the unity operators, the scalars being equal to $l_1,l_2,...,l_d$, respectively.

Assuming for the vector operator X an expression of the type (3.5), we have

$$X^2 = \sum_{k=0}^{2(d-1)} u_k(\mathrm{Tr}(a^*a),\mathrm{Tr}(a^*a)^2,...,\mathrm{Tr}(a^*a)^d)(a^*a)^k, \tag{4.33}$$

where $u_k = \Sigma_{k_1+k_2=k} x_{k_1}x_{k_2}$. But, from the Louck–Biedenharn theorem

$$(a^*a)^m = \sum_{q=0}^{d-1} c_{m,q}(\mathrm{Tr}(a^*a),...,\mathrm{Tr}(a^*a)^d)(a^*a)^q, \tag{4.34}$$

where

$$\begin{aligned} c_{d,d-1} &= l_1 + l_2 + \cdots + l_d, \\ c_{d,d-2} &= -(l_1l_2 + l_1l_3 + \cdots + l_{d-1}l_d), \\ &\vdots \\ c_{d,0} &= (-1)^d l_1l_2\cdots l_{d-1}l_d, \end{aligned} \tag{4.35}$$

and

$$c_{m+1,k} = c_{m,d-1}c_{d,k} + c_{m,k-1} \quad (k = 1,...,d-1), \tag{4.36}$$

$$c_{m+1,0} = c_{m,d-1}c_{d,0}.$$

if we use (4.35) in X^2, we obtain

$$X^2 = \sum_{k=0}^{d-1} g_k(\mathrm{Tr}(a^*a),...,\mathrm{Tr}(a^*a)^d)(a^*a)^k, \tag{4.37}$$

where

$$g_k = u_k + \sum_{m=d}^{2(d-1)} u_m c_{m,k}, \quad k = 0,1,...,d-1. \tag{4.38}$$

Then, from Eq. (4.16) it follows that $g_k = 0$ for $k = 2,3,...,d-1$, $g_1 = 1$, and $g_0 = N + d - 1$.

Theorem 3: Equations (4.16) and (4.17) admit the solutions

$$X = \sum_{k=0}^{d-1} x_k(\mathrm{Tr}(a^*a),...,\mathrm{Tr}(a^*a)^d)(a^*a)^k, \tag{4.39}$$

$$Y = \sum_{k=0}^{d-1} y_k(\mathrm{Tr}(a^*a)^t,...,\mathrm{Tr}((a^*a)^t)^d)((a^*a)^t)^k, \tag{4.40}$$

where, in each irreducible representation of $u(d)$, the scalars x_k, y_k, $k = 1,...,d-1$ are the solutions of the equations

$$\left(\sum_{k=0}^{d-1} l_j^k x_k\right)^2 = l_j + N + \epsilon d - 1, \tag{4.41}$$

and

$$\left(\sum_{k=0}^{d-1} \tilde{l}_j^k y_k\right)^2 = \tilde{l}_j + N, \tag{4.42}$$

respectively.

Proof: We have

$$\sum_{k=0}^{d-1} l_j^k g_k = \sum_{k=0}^{d-1} l_j^k u_k + \sum_{m=d}^{2(d-1)} u_m \sum_{k=0}^{d-1} l_j^k c_{m,k}$$

$$= l_j + N + \epsilon d - 1, \tag{4.43}$$

and we can prove by induction (for $m > d$) that

$$\sum_{k=0}^{d-1} l_j^k c_{m,k} = l_j^m \quad (j = 1,...,d). \tag{4.44}$$

For $m = d$, relation (4.44) reduces to an evident algebraic identity. Let us assume relation (4.44) valid for $m = n$. We have

$$\sum_{k=0}^{d-1} l_j^k c_{n+1,k} = c_{n,d-1} \sum_{k=0}^{d-1} l_j^k c_{d,k} + \sum_{k=1}^{d-1} l_j^k c_{n,k-1}. \tag{4.45}$$

Taking into account the identity [(4.44); $m = d$] we obtain from (4.45) the relation (4.44) for $m = n + 1$. Then

$$\sum_{k=0}^{d-1} l_j^k g_k = \sum_{k=0}^{2(d-1)} l_j^k u_k$$

$$= \left(\sum_{k=0}^{d-1} l_j^k x_k\right)^2 = l_j + N + \epsilon d - 1. \tag{4.46}$$

The proof of relation (4.42) is analogous.

ACKNOWLEDGMENTS

It is a pleasure for the authors to thank Professor Marcos Moshinsky for calling their attention to the works of Deenen and Quesne on boson realizations of symplectic algebras. The authors are greatly indebted to Dr. Quesne and to Professor Moshinsky for sending them information previous to publication.

APPENDIX: ELEMENTARY TENSORS OF SECOND DEGREE

The elementary tensors form irreducible subspaces under the action of the adjoint representation on the enveloping algebra. The spectral analysis of this extended adjoint representation has been obtained in an abstract setting by Kostant.[24] Partial concrete results have been obtained in Refs. 15, 22, 25, and 26.

A systematic description for elementary tensors of second degree has been given in Ref. 19, using a simple method, which exploits the equivalence between the adjoint representation on the enveloping algebra and the adjoint representation on the space of all polynomials on the dual of the Lie algebra. This equivalence allows a commutative calculus of the elementary tensors in this last representation. Using a projection technique (followed by a symmetrization, in the quantum case) the elementary tensors in the enveloping algebra have been obtained.

For the Lie algebras of type D_n, to which the present paper refers, the calculation has been performed in the basis M_{ij}, $i,j = 1,...,N = 2n$, with $M_{ij} = -M_{ji}$, the generators M_{ij} satisfying the structure equations

$$[M_{ij},M_{kl}] = \delta_{il}M_{jk} + \delta_{jk}M_{il} - \delta_{ik}M_{jl} - \delta_{jl}M_{ik}. \tag{A1}$$

In the space of second-degree polynomials in the enveloping algebra there are four invariant subspaces on which the adjoint representation acts by irreducible subrepresentations of types (0), (Λ_4), $(2\Lambda_1)$, and $(2\Lambda_2)$, where $\Lambda_1,...,\Lambda_n$ denote the maximal weights of the fundamental representations.

The elementary tensors of second degree corresponding to all these representations have been deduced in Ref. 19. That corresponding to the subrepresentation $(2\Lambda_1)$ is given by the polynomials

$$T_{ps} = \sum_{i=1}^{N} M_{pi}M_{is} - \frac{1}{N}\delta_{ps}C, \tag{A2}$$

where $C = \Sigma_{i,j=1}^{N} M_{ij}M_{ji}$ is the Casimir invariant.

These expressions can be transformed to a Cartan–Weyl basis by means of the formulas

$$\begin{aligned} M_{2i,2j-1} &= (\sqrt{-1}/2)(B_{ij} + C_{ij} - A_{ij} - A_{ji}), \\ M_{2i-1,2j} &= (\sqrt{-1}/2)(B_{ij} + C_{ij} + A_{ij} + A_{ji}), \\ M_{2i-1,2j-1} &= \tfrac{1}{2}(B_{ij} - C_{ij} - A_{ij} + A_{ji}), \\ M_{2i,2j} &= \tfrac{1}{2}(B_{ij} - C_{ij} + A_{ij} - A_{ji}). \end{aligned} \tag{A3}$$

By performing the transformation (A3), we obtain from the polynomials (A2), after symmetrization, the matrix equations (4.1)–(4.3) with $\epsilon = 1$, if this tensor vanishes in a given representation.

A similar calculation gives the tensor with respect to C_n algebras, which corresponds to the subrepresentation (Λ_2) and from which the matrix equations (4.1)–(4.3) with $\epsilon = -1$ follow, when the tensor vanishes in a given representation.

[1]J. Deenen and C. Quesne, Physica A **114**, 318 (1982).
[2]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878 (1982).
[3]J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982).
[4]J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638 (1984).
[5]J. Deenen and C. Quesne, "Boson representations of the real symplectic group and their applications to the nuclear collective model," Bruxelles University, Preprint No. PTM 84-07.
[6]M. Moshinsky, Nucl. Phys. A **421**, 81c (1984).
[7]M. Moshinsky, J. Phys. A: Math. Gen. **18**, L1 (1985).
[8]O. Castaños, E. Chacon, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985).
[9]D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
[10]J. Dobaczewski, Nucl. Phys. A **369**, 213, 237 (1981); **380**, 1 (1982).
[11]A. Klein, T. D. Cohen, and C-T. Li, Ann. Phys. **141**, 382 (1982).
[12]A. Klein and C-T. Li, Prog. Theor. Phys. Suppl. **74–75**, 237 (1983).
[13]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. NY **131**, 1 (1981).
[14]N. Papanicolaou, Ann. Phys. NY **136**, 210 (1981).
[15]S. Okubo, J. Math. Phys. **16**, 528 (1975).
[16]N. Mukunda, J. Math. Phys. **8**, 1069 (1967).
[17]M. Iosifescu and H. Scutaru, J. Math. Phys. **21**, 2033 (1980).
[18]M. Iosifescu and H. Scutaru, J. Math. Phys. **25**, 2856 (1984).
[19]M. Iosifescu and H. Scutaru, "Polynomial constraints associated to dynamical symmetries," Bucharest, preprint No. FT-253-84 (1984).
[20]M. Iosifescu and H. Scutaru, "A moment map-like mapping," Bucharest, preprint No. FT-254-84, 1984.
[21]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[22]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[23]S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[24]B. Kostant, Am. J. Math. **85**, 327 (1963).
[25]S. Okubo, J. Math. Phys. **20**, 586 (1979).
[26]P. D. Jarvis, J. Phys. A: Math. Gen. **12**, 1 (1979).

Analysis of the Moyal product in a flat space

J-B. Kammerer
Ecole Centrale des Arts et Manufactures, 92290 Chatenay-Malabry, France and Ecole Nationale Supérieure des Télécommunications, 75634 Paris Cedex 13, France

(Received 23 July 1985; accepted for publication 16 October 1985)

This paper studies the mathematical properties of Moyal product defined in the space $\mathbf{R}^{2n}$. They correspond to the properties of quantum mechanics and permit us to consider classical mechanics as a limit of quantum mechanics when Planck's constant vanishes. A table of formulas and two algebras give in particular the necessary results for the approximation of the Moyal product and for the spectral resolution.

I. INTRODUCTION

First let us remind the reader of the motivations of the definition of the Moyal product and state the considered quantum problem before showing the obtained results and the plan of this paper. For a long time, the Schrödinger picture of quantum mechanics has looked extraneous to classical mechanics. "The usual probabilistic interpretation of quantum mechanics contrasts with the deterministic character of classical mechanics.[1]" On the other hand, the Heisenberg picture brings to light a correspondence between the operators, the quantum observables, and the measurable quantities. This principle of correspondence is not satisfactory.[2] In fact, the algebra of operators is not commutative; on the contrary, the set N of functions defined on the phase space $\mathbf{R}^{2n}$ has a structure of commutative algebra defined by the usual product.

Let us consider a dynamic system with constraints independent of time and n degrees of freedom. The quantization suggested by the authors of the paper[1] consists of deforming the algebra N by means of a star-product or Moyal product $*_\nu$, where ν ($i\hbar/2$) is the parameter of deformation, such as $(*_\nu, N)$ is a noncommutative algebra and N is a Lie algebra with respect to the Moyal bracket:

$$[u,v] = (2\nu)^{-1}(u*_\nu v - v*_\nu u).$$

Thus the observables $t \mapsto u(t)$ are functions of $\mathbf{R}^+$ into $\mathbf{R}^{2n}$, such that

$$\frac{du}{dt} = [u(t),H], \quad u(0) = u_0, \tag{1.1}$$

where H is the Hamiltonian function. This relation is similar to the equation of classical dynamics[3]; it corresponds to the fundamental equation of quantum mechanics in the Heisenberg picture by the Weyl map.[4]

The aim of this paper is the mathematical analysis of Moyal product in the space $\mathbf{R}^{2n}$, in order to prove the hoped results: the relation (1) generally admits stationary solutions because of the spectral resolution of the real elements (see Sec. VI); thus there is really quantization. Classical mechanics appears as a limit of quantum mechanics, when the parameter ν vanishes; for in some cases the Moyal product is an analytic function of ν at 0 (Sec. II D), or more generally a Taylor's formula is verified (Sec. III B). For example, in the case of the harmonic oscillator, the limit of the stationary solutions[1] is the classical solution.

The Moyal product is defined here with the twisted convolution[5] and studied by a "kernels composition law." This method gives a table of formulas for the functions belonging to the spaces $L^2(\mathbf{R}^{2n})$ and $\mathscr{S}(\mathbf{R}^{2n})$ (Sec. II C). The inhomogeneous symplectic group leaves the Moyal product invariant.[1,6] The extension of the Moyal product to the Fourier transforms of distributions with compact support (in particular the polynomials) gives the well-known formal definition[1] (Sec. II D).

The spaces $\mathscr{S}(\mathbf{R}^{2n})$ and $L^2(\mathbf{R}^{2n})$ are algebras; this property is important for the following: the above formulas are verified by the pairs of tempered distributions, at least one of which belongs to the space $\mathscr{O}_M^\nu$ of star-multiplication operators in $\mathscr{S}(\mathbf{R}^{2n})$. The study of this space proves, under suitable assumptions, that the formal definition gives a pth approximation of the Moyal product ($p \in \mathbf{N}$) (Sec. III B).

Some of these formulas are applicable to the distributions belonging to the space $\mathscr{B}^\nu$ of the star-multiplication operators in $L^2(\mathbf{R}^{2n})$ (Sec. IV). The projections and the resolutions of identity belong to this normed linear space equipped with an order relation (Sec. IV C).

The necessary results are collected to define resolvent, spectrum, eigenvalue, and eigenelement and to show in particular the existence of a greater eigenelement (Sec. V). Thus, it is easy to prove the spectral resolution of the real or unitary elements of $\mathscr{B}^\nu$ and of the real elements of $\mathscr{O}_M^\nu$ (Sec. VI).

II. THE STAR-PRODUCT

Notations: The greek letter ν indicates a complex number ($\nu = -\bar{\nu} \neq 0$); n an integer. If x is in $\mathbf{R}^{2n}$, α in $\mathbf{N}^{2n}$, let us set $x = (x_1,x_2)$, $\alpha = (\alpha_1,\alpha_2)$. Besides the ordinary symbols, $\overline{f}$, $\tau_a f$, and f_a, let us set $f^\dagger(x) = f(-x)$, $\overleftrightarrow{f}(x) = f(x_2,x_1)$, $\widetilde{f}(x) = \overline{f(x_2,x_1)}$, $f_1 : u \mapsto f(x_1,u)$.

A. The operators $\mathscr{T}_\nu$, $\mathscr{T}_\nu^*$

If f is a function of $\mathbf{R}^{2n}$ in $\mathbf{C}$, let us set

$$\mathscr{T}_\nu f(x) = |2\nu|^{n/2} f(x_1 + i\nu x_2, x_1 - i\nu x_2),$$
$$\mathscr{T}_\nu^* f(x) = |2\nu|^{-n/2} f((x_1 + x_2)/2, (x_1 - x_2)/2i\nu). \tag{2.1}$$

If f is a tempered distribution, $\mathscr{T}_\nu f$ and $\mathscr{T}_\nu^* f$ are defined by the relations

 0022-2488/86/020529-07$02.50 © 1986 American Institute of Physics

$\forall \phi \in \mathcal{S}(\mathbb{R}^{2n})$,

$$\langle \mathcal{T}_\nu f,\phi\rangle = \langle f,\mathcal{T}_\nu^* \phi\rangle, \quad \langle \mathcal{T}_\nu^* f,\phi\rangle = \langle f,\mathcal{T}_\nu \phi\rangle. \tag{2.2}$$

The maps $\mathcal{T}_\nu$ and $\mathcal{T}_\nu^*$ are continuous and inverse each other of $\mathcal{S}(\mathbb{R}^{2n})$ [resp. $\mathcal{S}'(\mathbb{R}^{2n})$] into itself. They are unitary operators of $L^2(\mathbb{R}^{2n})$ into itself.[7]

The operators $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}$, and F: Let $f \in \mathcal{S}'(\mathbb{R}^{2n})$. We denote by $\mathcal{F}_1 f$ (resp. $\mathcal{F}_2 f$) the image of the tempered distribution f_2 (resp. f_1) by the Fourier transform [$f_1, f_2 \in \mathcal{S}'(\mathbb{R}^n)$]. Let us write

$$\mathcal{F} = \mathcal{F}_1 \mathcal{F}_2, \quad F = \mathcal{F}_1 \mathcal{F}_2^*, \quad F^* = \mathcal{F}_1^* \mathcal{F}_2. \tag{2.3}$$

Many relations exist between these operators; an example is

$$\mathcal{F}_2^* \mathcal{T}_\nu = \mathcal{F}_2 \mathcal{T}_{-\nu}, \quad \mathcal{T}_\nu^* \mathcal{F}_2^* = \mathcal{T}_{-\nu}^* \mathcal{F}_2. \tag{2.4}$$

The image of the real $\beta_n = |4\pi\nu|^{-n/2}$ under the map $\mathcal{T}_\nu^* \mathcal{F}_2$ is the distribution δ_Δ:

$$\forall \phi \in \mathcal{S}(\mathbb{R}^{2n}), \quad \langle \delta_\Delta,\phi\rangle = \int_{\mathbb{R}^n} \phi(u,u)\, du. \tag{2.5}$$

B. The laws of composition $\circ$, $\hat{*}_\nu$, and $*_\nu$

Let $f \in L^2(\mathbb{R}^{2n})$. Let us denote by $f \circ g$ the kernel of the integral operator, product of the operators the kernels of which are f and g:

$$\forall x \in \mathbb{R}^{2n}, \quad f \circ g(x) = \int_{\mathbb{R}^n} f(x_1,u)\, g(u,x_2)\, du. \tag{2.6}$$

Let $f, g \in L^1(\mathbb{R}^{2n})$; $f \hat{*}_\nu g$ is the twisted convolution of these functions:

$$\forall x \in \mathbb{R}^{2n}, \quad f \hat{*}_\nu g(x) = \int_{\mathbb{R}^{2n}} f(t)\, g(x-t) \exp(\nu t \wedge x)\, dt, \tag{2.7}$$

$t \wedge x = t_1 x_2 - t_2 x_1$. The star-product $f *_\nu g$ of two functions f, g that belong to $\mathcal{F}(L^1(\mathbb{R}^{2n}))$ is the function[5,8]

$$f *_\nu g = (2\pi)^n F^*(Ff \hat{*}_\nu Fg). \tag{2.8}$$

Let $f, g \in \mathcal{S}(\mathbb{R}^{2n})$. We have the relations

$$\begin{aligned} f \hat{*}_\nu g &= |\pi/\nu|^{n/2} \mathcal{F}_1^* \mathcal{T}_\nu (\mathcal{T}_\nu^* \mathcal{F}_1 f \circ \mathcal{T}_\nu^* \mathcal{F}_1 g), \\ f *_\nu g &= |4\pi\nu|^{-n/2} \mathcal{F}_2^* \mathcal{T}_\nu (\mathcal{T}_\nu^* \mathcal{F}_2 f \circ \mathcal{T}_\nu^* \mathcal{F}_2 g). \end{aligned} \tag{2.9}$$

C. The algebras $\mathcal{S}(\mathbb{R}^{2n})$, $L^2(\mathbb{R}^{2n})$

Products $\circ$, $\hat{*}_\nu$, and $*_\nu$ give a structure of algebra to the spaces $\mathcal{S}(\mathbb{R}^{2n})$ and $L^2(\mathbb{R}^{2n})$ (see Refs. 9–11); these algebras are isomorphic to each other [(2.9)]. Each of these laws is a linear continuous map of $\mathcal{S}(\mathbb{R}^{2n}) \times \mathcal{S}(\mathbb{R}^{2n})$ into $\mathcal{S}(\mathbb{R}^{2n})$ and of $L^2(\mathbb{R}^{2n}) \times L^2(\mathbb{R}^{2n})$ into $L^2(\mathbb{R}^{2n})$. The star-product has the properties

(a) $\forall f, g \in L^2(\mathbb{R}^{2n})$,

$$\begin{aligned} &(f *_\nu g)^\dagger = f^\dagger *_\nu g^\dagger, \quad \tau_a(f *_\nu g) = (\tau_a f) *_\nu (\tau_a g), \\ &f *_\nu g(x) e^{iax} = f(x_1 - i\nu a_2, x_2 + i\nu a_1) \\ &\qquad \times *_\nu (g(x) e^{iax}), \quad a \in \mathbb{R}^{2n}, \\ &\overrightarrow{F^*(f *_\nu g)}_{i/\nu} = f^\dagger *_\nu (F\overleftrightarrow{g})_{i/\nu} = (F\overleftrightarrow{f})_{i/\nu} *_\nu g; \\ &f *_{ia^2} g = (f_a *_i g_a)_{1/a}; \quad a > 0; \\ &f *_\nu g = g *_{\bar\nu} f, \end{aligned} \tag{2.10}$$

$$\overrightarrow{f *_\nu g} = \bar f *_{\bar\nu} \bar g, \quad \overrightarrow{f *_\nu g} = \overleftrightarrow{f} *_{\bar\nu} \overleftrightarrow{g}, \tag{2.11}$$

$$\forall \phi \in \mathcal{S}(\mathbb{R}^{2n}), \quad \langle f, g *_\nu \phi\rangle = \langle f *_\nu g, \phi\rangle, \tag{2.12}$$

(b) $\forall f, g \in \mathcal{S}(\mathbb{R}^{2n})$,

$$\begin{aligned} &D^\alpha(f *_\nu g) = \sum_{\beta<\alpha} \frac{\alpha!}{\beta!(\alpha-\beta)!} D^\beta f *_\nu D^{\alpha-\beta} g, \\ &x^\alpha \cdot (f *_\nu g)(x) = \sum_{\beta<\alpha} \frac{\alpha! \nu^{|\beta|}}{\beta!(\alpha-\beta)!} \\ &\qquad \times (-1)^{|\beta_2|} D^{\overleftrightarrow{\beta}} f(x) *_\nu (x^{\alpha-\beta} g(x)); \end{aligned} \tag{2.13}$$

(c) $\forall f, g \in \mathcal{F}L^1(\mathbb{R}^{2n})$,

$$\overrightarrow{f *_\nu g} = (2\pi)^{-n} (\overleftrightarrow{f}_{-i\nu} \hat{*}_\nu F^* g)_{i/\nu}. \tag{2.14}$$

D. First properties

Let $\mathrm{Sp}(n,\mathbb{R})$ be the group of linear maps s to $\mathbb{R}^{2n}$ into itself, such as $\forall x, y \in \mathbb{R}^{2n}$, $s(x) \wedge s(y) = x \wedge y$ (see Ref. 12). For all $f, g \in L^2(\mathbb{R}^{2n})$, for all s such as ${}^t s \in \mathrm{Sp}(n,\mathbb{R})$, we have[1,6]

$$\forall x \in \mathbb{R}^{2n}, \quad f *_\nu g(sx) = f(sx) *_\nu g(sx). \tag{2.15}$$

Then, the star-product of two functions, which have the spherical symmetry, has this spherical symmetry. It is easy to extend the twisted convolution to two distributions, at least one of which has a compact support.[8] In particular, $\mathcal{F}\mathcal{E}'(\mathbb{R}^{2n})$ is a $*_\nu$-algebra. We have $\forall f, g \in \mathcal{F}\mathcal{E}'(\mathbb{R}^{2n})$ or $f \in \mathbb{C}[X]^{2n}$, $g \in \mathcal{S}'(\mathbb{R}^{2n})$,

$$f *_\nu g = f \cdot g + \sum_{q=1}^{\infty} \frac{\nu^q}{q!} P^q(f,g), \quad \text{in } \mathcal{S}'(\mathbb{R}^{2n}); \tag{2.16}$$

$$P^q(f,g) = \Lambda^{i_1 j_1} \cdots \Lambda^{i_q j_q} \partial_{i_1 \cdots i_q} f \cdot \partial_{j_1 \cdots j_q} g,$$

$\Lambda^{ij} = 1$ if $j = i+n$, -1 if $i = j+n$, 0 in the other cases. This relation has been defined for a long time.[13] Let $f \in \mathcal{F}\mathcal{E}'(\mathbb{R}^{2n})$; the map $\phi \to f *_\nu \phi$ is continuous of $\mathcal{S}(\mathbb{R}^{2n})$ into itself. For example,

$$\forall p, \quad 1 \leqslant p \leqslant n, \quad x_p *_\nu g = x_p g + \nu \partial_{p+n} g.$$

III. ALGEBRA OF STAR-MULTIPLICATION OPERATORS

Let $g \in \mathcal{S}(\mathbb{R}^{2n})$; the map $\phi \mapsto g *_\nu \phi$ of $\mathcal{S}(\mathbb{R}^{2n})$ into itself is continuous. The image of a tempered distribution f under the transpose of this map is the distribution $f *_\nu g$ such that the relation (2.13) holds. An example is

$$\delta *_\nu \phi(x) = |2\pi\nu^2|^{-n} (F\overleftrightarrow{\phi})(ix/\nu). \tag{3.1}$$

Likewise it is easy to define $f \circ \phi$ and $f \hat{*}_\nu \phi$; then, $\forall f \in \mathcal{S}'(\mathbb{R}^{2n})$, $\forall g \in \mathcal{S}(\mathbb{R}^{2n})$, the relations (2.9), (2.10), (2.14), and (2.15) hold.

Proposition: Let $S \in \mathcal{S}'(\mathbb{R}^{2n})$; the following properties are equivalent:

(i) $S = 0$, (ii) $\forall \phi \in \mathcal{S}(\mathbb{R}^{2n})$, $S *_\nu \phi = 0$. (3.2)

It is natural to consider the set $\mathcal{O}_M^\nu$ of distributions S such as the map $\phi \mapsto S *_\nu \phi$ is continuous of $\mathcal{S}(\mathbb{R}^{2n})$ into itself. This set is different from the space $\mathcal{O}_M(\mathbb{R}^{2n})$ of multiplication operators[14,15]; in fact, the sets $\mathcal{S}(\mathbb{R}^{2n})$, $\mathcal{F}(\mathcal{E}'(\mathbb{R}^{2n}))$, $\mathcal{E}'(\mathbb{R}^{2n})$, and the dual of $\mathcal{O}_M(\mathbb{R}^{2n})$ are subsets of $\mathcal{O}_M^\nu$ (see Ref. 16); but the function $x \mapsto \exp(x_1 x_2/\nu)$ does

not belong to $\mathcal{O}_M^\nu$. Likewise it is easy to define $\mathcal{N}^0$ (resp. $\mathcal{O}_C^{\prime\nu}$) as the set of distributions S such that the map $\phi \to S \circ \phi$ (resp. $S\hat{*}_\nu\phi$) is continuous of $\mathcal{S}(\mathbf{R}^{2n})$ into itself.[8] These spaces are isomorphic to each other (2.9). In particular, the linear one-to-one map J, of $\mathcal{N}^0$ onto $\mathcal{O}_M^\nu$ is defined by

$$T \mapsto \beta_n^{-1} \mathcal{F}_2^* \mathcal{T}_\nu T, \quad \beta_n \in \mathbf{R} \quad \text{(Sec. II A)}. \tag{3.3}$$

Let $f \in \mathcal{S}'(\mathbf{R}^{2n})$, $g \in \mathcal{O}_M^\nu$; only one distribution $f *_\nu g$ is defined by the relation (2.13). The relations (2.9), (2.10), (2.14), (2.15) hold in the same conditions. The linear space $\mathcal{O}_M^\nu$ is a $*_\nu$-algebra; the space $\mathcal{S}'(\mathbf{R}^{2n})$ is a right-module on $\mathcal{O}_M^\nu$. This algebra is nondegenerated[3]:

$$\text{if} \quad S \in \mathcal{O}_M^\nu, \quad S = 0 \Leftrightarrow \overline{S} *_\nu S = 0. \tag{3.4}$$

Finally, let f be in $\mathcal{O}_{\bar{M}}^{\nu}$ and g in $\mathcal{O}_M^\nu$; the relation (2.11) holds. Thus, let f be in $\mathcal{O}_{\bar{M}}^{\nu}$ and g be a tempered distribution. Let us define $f *_\nu g$ by this relation (2.11). Therefore, the space $\mathcal{S}'(\mathbf{R}^{2n})$ is a left-module on $\mathcal{O}_{\bar{M}}^{\nu}$. The relations (2.12) prove that the maps $f \mapsto \overline{f}$, $\overrightarrow{f}$ are isomorphisms of $\mathcal{O}_M^\nu$ onto $\mathcal{O}_{\bar{M}}^{\nu}$; these relations hold when f belongs to $\mathcal{O}_{\bar{M}}^{\nu}$ and g to $\mathcal{S}'(\mathbf{R}^{2n})$.

A. The linear space $\mathcal{O}_M^\nu$

Let us consider the family of seminorms on the space $\mathcal{O}_M^\nu$:

$$p_{\alpha;B}(T) = \operatorname*{Sup}_{\phi \in B} \operatorname*{Sup}_{x \in \mathbf{R}^{2n}} |(D^\alpha T) *_\nu \phi(x)|, \tag{3.5}$$

where $\alpha \in \mathbf{R}^{2n}$, B is a bounded set of $\mathcal{S}(\mathbf{R}^{2n})$, and $T \in \mathcal{O}_M^\nu$.

Proposition: The set $\mathcal{O}_M^\nu$ is a complete Hausdorff topological vector space. It is isomorphic to the space $L_b(\mathcal{S}(\mathbf{R}^n))$ (see Ref. 14 for notation).[17]

Proof: The space $\mathcal{O}_M^\nu$ is isomorphic to the space $\mathcal{N}^0$, which is a complete Hausdorff TVS by a family of suitable seminorms. The restriction to $\mathcal{N}^0$ of the map $T \mapsto u_T$ of $\mathcal{S}'(\mathbf{R}^{2n})$ into $L(\mathcal{S}(\mathbf{R}^n), \mathcal{S}'(\mathbf{R}^n))$, defined by

$$\forall \phi, \psi \in \mathcal{S}(\mathbf{R}^n), \quad \langle u_T(\phi), \psi \rangle = \langle T, \psi \otimes \phi \rangle, \tag{3.6}$$

is an isomorphism of $\mathcal{N}^0$ onto $L_b(\mathcal{S}(\mathbf{R}^n))$. The map $(f, g) \mapsto f *_\nu g$ of $\mathcal{S}'(\mathbf{R}^{2n}) \times \mathcal{O}_M^\nu$ (resp. $\mathcal{O}_M^\nu \times \mathcal{O}_M^\nu$) into $\mathcal{S}'(\mathbf{R}^{2n})$ (resp. $\mathcal{O}_M^\nu$) is separately continuous. The maps

$$T \mapsto \tilde{T}, \ T^\dagger, \ \tau_a T, \ e^{iax} T, \ D^\alpha T, \ x^\alpha T, \ (\overrightarrow{FT})_{i/\nu},$$

are continuous of $\mathcal{O}_M^\nu$ into itself. It is easy to obtain a characteristic property of the distributions that belong to this space; for example, $1 \otimes f$, $\delta \otimes S$, and their symmetrical when $f \in \mathcal{O}_M(\mathbf{R}^{2n})$, $S \in \mathcal{O}_C'(\mathbf{R}^{2n})$.

Let us denote by $\mathcal{N}$ the space $\mathcal{S}(\mathbf{R}^n) \hat{\otimes} \mathcal{S}'(\mathbf{R}^n)$; it is a subset of $\mathcal{N}^0$ and its topology is the topology induced by $\mathcal{N}^0$ as a subset of $\mathcal{N}^0$. The map $(\phi, S) \mapsto \phi \circ S$ of $\mathcal{S}(\mathbf{R}^{2n}) \times \mathcal{S}'(\mathbf{R}^{2n})$ into $\mathcal{N}$ is separately continuous; because of the map J (resp. F^*J) (3.3), the application $(\phi, S) \mapsto \phi *_\nu S$ (resp. $\phi \hat{*}_\nu S$) of $\mathcal{S}(\mathbf{R}^{2n}) \times \mathcal{S}'(\mathbf{R}^{2n})$ into $\mathcal{F}_2^* \mathcal{T}_\nu \mathcal{N}$ (resp. $\mathcal{F}_1^* \mathcal{T}_\nu \mathcal{N}$) is separately continuous. Finally, let ϕ_p, $p = 1,2,...$, be a regularizing sequence[14]; this sequence converges to δ in $\mathcal{O}_C'^\nu$. Consequently the space $\mathcal{O}_M^\nu$ is $\mathcal{F}_2^* \mathcal{T}_\nu (\mathcal{S}(\mathbf{R}^n) \hat{\otimes} \mathcal{S}'(\mathbf{R}^n))$.

B. Taylor's formula

Since there is no associative local noncommutative algebra,[18] it is not surprising that the support of $f *_\nu g$ is different from the support of $f \cdot g$. For example, if f, g have compact supports, the support of $f *_\nu g$ is $\mathbf{R}^{2n}$. In this case, the map $\nu \mapsto f *_\nu g$ is not an analytic function of the variable ν. Generally there is a Taylor formula.[17]

Proposition: Let f be a tempered distribution such that (i) $\exists \nu_0, \forall \nu \in]0, \nu_0]$, $f \in \mathcal{O}_M^\nu$; and (ii) $\forall \alpha \in \mathbf{N}^{2n}$, $\forall B$ bounded set of $\mathcal{S}(\mathbf{R}^{2n})$,

$$\exists C, \quad \operatorname*{Sup}_{\nu \in]0, \nu_0]} p_{\alpha,B}(f) < C.$$

We have

$$f \in \mathcal{O}_M(\mathbf{R}^{2n}), \quad \forall S \in \mathcal{S}'(\mathbf{R}^{2n}), \quad \forall p \in \mathbf{N}, \quad \exists h_\nu^p(f, S), \tag{3.7}$$

$$f *_\nu S = f \cdot S + \sum_{q=1}^{p-1} \frac{\nu^q}{q!} P^q(f, S) + \frac{\nu^p}{p!} h_\nu^p(f, S),$$

If f and S are fixed, the remainder $h_\nu^p(f, S)$ lies in a bounded set of $\mathcal{S}'(\mathbf{R}^{2n})$, when ν belongs to $]0, \nu_0]$.

For an example, let f be a tempered distribution:

$$f(x) *_\nu \exp(iax) = \exp(iax) \cdot f(x_1 + iva_2, x_2 - iva_1),$$
$$a \in \mathbf{R}^{2n}.$$

Proof: Let ϕ and ψ be some functions belonging to $\mathcal{S}(\mathbf{R}^{2n})$, x an element of $\mathbf{R}^{2n}$; the map of $i\mathbf{R}$ into $\mathbf{C}$,

$$\nu \mapsto \int_{\mathbf{R}^{2n}} \phi(t) \psi(x - t) \exp(\nu t \wedge x) \, dt,$$

is infinitely differentiable. Let p be a natural number, the Taylor formula gives the pth approximation of this function in a neighborhood of 0. Then the twisted convolution of ϕ and ψ is the sum of a polynomial in the variable ν whose degree is $p - 1$ and of a remainder that lies in a bounded set of $\mathcal{S}(\mathbf{R}^{2n})$ when ν belongs to $]0, \nu_0]$ and ϕ and ψ are in bounded sets of $\mathcal{S}(\mathbf{R}^{2n})$. Obviously an analogous result holds for the Moyal product of the functions ϕ and ψ. Then it is easy to prove the above proposition by means of the duality between $\mathcal{S}(\mathbf{R}^{2n})$ and $\mathcal{S}'(\mathbf{R}^{2n})$.

IV. THE SPACES $\mathcal{B}^\nu$ AND Π^ν

Definition: Let us denote by $\mathcal{B}^\nu$ the set of the distributions f such that the map $\phi \mapsto f *_\nu \phi$ is continuous of $\mathcal{S}(\mathbf{R}^{2n})$ equipped with the topology induced by $L^2(\mathbf{R}^{2n})$, into $L^2(\mathbf{R}^{2n})$. The natural norm is

$$\|f\|_\epsilon^\nu = \operatorname{Sup}\{\|f *_\nu \phi\| \mid \phi \in \mathcal{S}(\mathbf{R}^{2n}), \quad \|\phi\| = 1\}. \tag{4.1}$$

It is quite obvious, because of the relations (2.10) and (2.12), that the maps

$$f \mapsto f^\dagger, \ e^{iax} \cdot f, \tau_a f, \ \overrightarrow{F(f_{i\nu})}, \ \tilde{f} \quad (\text{resp. } \overline{f}, \overleftrightarrow{f}),$$

are isometries of $\mathcal{B}^\nu$ into itself (resp. $\mathcal{B}^{-\nu}$).

A. Characterization of $\mathcal{B}^\nu$

It is well known[14] that the completion of the space $L^2(\mathbf{R}^n) \otimes L^2(\mathbf{R}^n)$, equipped with the norm $\|\cdot\|_\pi$, $\|\cdot\|$, $\|\cdot\|_\epsilon$, is, respectively, the space Π, $L^2(\mathbf{R}^{2n})$, or E; the space B_ϵ is the subset of distributions that is isomorphic, by the relation

(3.6), to the set of the bilinear continuous mappings on $L^2(\mathbb{R}^n)$. Let us remember that the space Π is isomorphic to the dual of E and B_ϵ to the dual of Π. The usual bracket $\langle\cdot,\cdot\rangle$ will indicate, by extension, this duality.

It is easy to show that the linear space $L^2(\mathbb{R}^n) \otimes L^2(\mathbb{R}^n)$ is a $\circ$-algebra; because of the continuity of this map $\circ$, the spaces Π and E are $\circ$-algebras, respectively, isomorphic to space of nuclear and of compact operators. According to a classical result,[19]

$$\Pi = L^2(\mathbb{R}^{2n}) \circ L^2(\mathbb{R}^{2n}). \tag{4.2}$$

This product $\circ$ is a continuous map of $\Pi \times E$ into Π, of $E \times L^2(\mathbb{R}^{2n})$ into $L^2(\mathbb{R}^{2n})$; we have

$$\forall f,g \in E, \quad \forall h \in \Pi, \quad \langle f \circ g, h \rangle = \langle \overrightarrow{f}, g \circ \overleftarrow{h} \rangle. \tag{4.3}$$

Let $f \in B_\epsilon$, $g \in \Pi$. Since B_ϵ is isomorphic to the dual of Π, $f \circ g$ is the distribution defined by the relation (4.3); it follows that the Banach space B_ϵ is also a C^*-algebra,[20] isomorphic to the space $L(L^2(\mathbb{R}^n))$ by the relation (3.6).

Proposition: $(\mathscr{B}^\nu, *_\nu)$ is a C^*-algebra.[21]

Proof: The following properties are equivalent:

$$\mathscr{F}_2^* \mathscr{T}_\nu f \in \mathscr{B}^\nu \Leftrightarrow f \in B_\epsilon \Leftrightarrow \phi \mapsto f \circ \phi \in L(L^2(\mathbb{R}^n)).$$

Consequently the sets $\mathscr{B}^\nu$ and $\mathscr{B}^{-\nu}$ are equal; the sets $L^2(\mathbb{R}^{2n})$, $\mathscr{F}L^1(\mathbb{R}^{2n})$, $\{1\} \otimes L^\infty(\mathbb{R}^n)$, $\{\delta\} \otimes L^1(\mathbb{R}^n)$, and $\mathscr{E}'^0(\mathbb{R}^{2n})$ are subsets of $\mathscr{B}^\nu$.

B. The topology of $\mathscr{B}^\nu$

Let us denote by Π^ν, E^ν the image spaces of Π and E by the map J, by $\|\cdot\|_\pi^\nu$, $\|\cdot\|^\nu$, $\|\cdot\|_\epsilon^\nu$ the image norms of the norms of the spaces Π, $L^2(\mathbb{R}^{2n})$, E. Since Π^ν is the image of Π, we have a more precise property than the relation (4.2) (see Ref. 19):

$$\begin{array}{ccccccccc}
\mathscr{S}(\mathbb{R}^{2n}) & \hookrightarrow & \Pi^\nu & \hookrightarrow & L^2(\mathbb{R}^{2n}) & \hookrightarrow & E^\nu & \to & \mathscr{B}^\nu & \hookrightarrow \mathscr{S}'(\mathbb{R}^{2n}) \\
J\uparrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & \\
 & & \Pi & \hookrightarrow & L^2(\mathbb{R}^{2n}) & \hookrightarrow & E & \to & B_\epsilon & \\
\text{set of} & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & \\
\text{operators} & & \{\text{nuclear}\} & & \left\{\begin{matrix}\text{Hilbert–}\\ \text{Schmidt}\end{matrix}\right\} & & \{\text{compact}\} & & \{\text{bounded}\} &
\end{array}$$

C. An order relation in $\mathscr{B}^\nu$

Let us denote by $\mathscr{P}^\nu$ the subset of Π^ν

$$\mathscr{P}^\nu = \{h \mid h \in \Pi^\nu, \quad \exists k \in L^2(\mathbb{R}^{2n}): \quad h = k *_\nu \bar{k}\}.$$

Let us not define an order relation on $\mathscr{B}^\nu$, compatible with the linear structure. Let $f, g \in \mathscr{B}^\nu$,

$$f < g \Leftrightarrow \forall h \in \mathscr{P}^\nu, \quad \langle f,h \rangle^\nu \leqslant \langle g,h \rangle^\nu. \tag{4.8}$$

The elements of $\mathscr{B}^\nu$, 1, $f *_\nu \bar{f}$ $(f \in \mathscr{B}^\nu)$, $J(\rho \otimes \bar{\rho})$ $[\rho \in L^2(\mathbb{R}^n)]$, are "positive." The well-known properties of the self-adjoint operators hold again; in particular, these are the theorems on the bounded increasing sequences and the following result.

Proposition: Let f be a real element of $\mathscr{B}^\nu$; there are two reals m and M such that

$$m < f < M, \quad \|f\|_\epsilon^\nu = \mathrm{Sup}(|m|, M). \tag{4.9}$$

The reals m and M are the lower and upper bounds of the set of the reals $\langle f,h \rangle^\nu$, in which the element h belongs to the unit sphere of Π^ν.

An example—the projections: A projection is a real element π of $\mathscr{B}^\nu$ such as $\pi *_\nu \pi$ is equal to π. The four classical properties are always true; in particular the result on the inequalities between projections. A total sequence of projections is a sequence of projections such that

$$\forall p,q \in \mathbb{N}, \quad p \neq q,$$

$$\pi_p *_\nu \pi_q = 0, \quad \sum_{p=1}^{\infty} \pi_p = 1, \quad \text{in } \mathscr{B}^\nu \text{ weakly.}$$

We have

$$\forall h \in \Pi^\nu, \quad \exists k_1, k_2 \in L^2(\mathbb{R}^{2n}),$$

$$h = k_1 *_\nu k_2 \quad \text{and} \quad \|k_1\|^\nu = \|k_2\|^\nu = (\|h\|_\pi^\nu)^{1/2}. \tag{4.4}$$

The map $(f,g) \mapsto f *_\nu g$ is continuous of $\Pi^\nu \times \mathscr{B}^\nu$ into Π^ν and of $\mathscr{B}^\nu \times L^2(\mathbb{R}^{2n})$ into $L^2(\mathbb{R}^{2n})$. The space $\mathscr{B}^\nu$ is isomorphic to the dual of Π^ν; let us denote by $\langle\cdot,\cdot\rangle^\nu$ the bracket of this duality. It corresponds to element volume $\tilde{\eta} = (2\pi\hbar)^{-n} \cdot \eta$ (see Refs. 3 and 10):

$$\forall f \in \mathscr{B}^\nu, \quad \forall h \in \Pi^\nu,$$

$$\langle f,h \rangle^\nu = \beta_n^2 \langle \mathscr{T}_\nu^* \mathscr{F}_2 f, \mathscr{T}_\nu^* \mathscr{F}_2^* h \rangle, \tag{4.5}$$

$$\langle f,\phi \rangle^\nu = \beta_n^2 \langle f,\phi \rangle, \quad \langle k_1, \bar{k}_2 \rangle = \beta_n^2 (k_1 | k_2), \tag{4.6}$$

$$\|f\|_\epsilon^\nu = \mathrm{Sup}(|\langle f,h \rangle^\nu| / \|h\|_\pi^\nu) = \beta_n \| \mathscr{T}_\nu^* \mathscr{F}_2 f \|_\epsilon.$$

Because of the relation (4.4), the elements belonging to Π^ν are continuous, bounded functions. This relation allows us to prove the following results.

Proposition: Let G be a linear bounded map of $L^2(\mathbb{R}^{2n})$ into itself; then the following two properties are equivalent:

$$\text{(i)} \quad \exists g \in \mathscr{B}^\nu, \quad \forall k \in L^2(\mathbb{R}^{2n}), \quad G(k) = g *_\nu k, \tag{4.7}$$

$$\text{(ii)} \quad \forall k_1, k_2 \in L^2(\mathbb{R}^{2n}), \quad G(k_1 *_\nu k_2) = G(k_1) *_\nu k_2.$$

The Weyl map[7,8,22]: Let us call Ω the compose of the map J^{-1} of $\mathscr{B}^\nu$ onto B_ϵ (3.3) and of the canonical map of B_ϵ onto $L(L^2(\mathbb{R}^n))$ (3.6); this map is an isometry. Let f_p, $p = 1,2,\ldots$, be a sequence in $\mathscr{B}^\nu$; the convergence of the sequence of the operators $\Omega(f_p)$ to $\Omega(f)$ is strong (resp. weak) iff the sequence $(\bar{f}_p - \bar{f}) *_\nu (f_p - f)$ (resp. $f_p - f$) converges weakly on 0. We present a summary statement:

$$\forall f \in \mathscr{B}^{\nu},\quad f = \sum_{p=1}^{\infty} f *_{\nu} \pi_p, \quad \text{in } \mathscr{B}^{\nu} \text{ weakly},$$

$$\forall k \in L^2(\mathbb{R}^{2n}),\quad (\|k\|^{\nu})^2 = \sum_{p=1}^{\infty} (\|k *_{\nu} \pi_p\|^{\nu})^2.$$

Resolution of the identity: A family of projections $[t \mapsto e(t)$ is a map of $\mathbb{R}$ into $\mathscr{B}^{\nu}]$ is called a resolution of the identity iff

$$\forall t, t' \in \mathbb{R},\quad t \leqslant t',\quad e(t) *_{\nu} e(t') = e(t);$$

$$\forall t \in \mathbb{R},\quad e(t+0) = e(t),\quad \text{in } \mathscr{B}^{\nu} \text{ weakly};$$

$$\lim_{t \to -\infty} e(t) = 0,\quad \lim_{t \to \infty} e(t) = 1,\quad \text{in } \mathscr{B}^{\nu} \text{ weakly}.$$

Here, $\pi(a) = e(a) - e(a-0)$ is a projection; for any $h \in \Pi^{\nu}$, $t \mapsto \langle e(t), h \rangle^{\nu}$ is a function of bounded variation. Let us call the spectrum of the resolution of the identity $t \mapsto e(t)$ the subset $\sigma(e)$ of the reals at which the family of projections is strictly increasing. Let us denote by $e(a,b]$, $e(b) - e(a)$, $a < b$.

V. RESOLVENT AND SPECTRUM

A. A preliminary result

Let us denote by $\mathscr{O}^{\nu}$ the set $\mathscr{O}^{\nu}_{M} \cap \mathscr{O}^{\nu}_{\bar{M}}$.

Definition: Let $f \in \mathscr{O}^{\nu}_{\bar{M}}$ and let us denote by O_f and D_f the operator and its domain, defined by the relations

$$D_f = \{k \mid k \in L^2(\mathbb{R}^{2n}) \quad \text{and} \quad f *_{\nu} k \in L^2(\mathbb{R}^{2n})\},$$
$$O_f(k) = f *_{\nu} k. \tag{5.1}$$

The subset D_f of $L^2(\mathbb{R}^{2n})$ is a right ideal; a similar property exists in the algebra $\mathscr{B}^{\nu}$. The operator O_f is closed.

Proposition: Let $f \in \mathscr{O}^{\nu}$. There is only one element b in $\mathscr{B}^{\nu}$ such that

$$(1 + \bar{f} *_{\nu} f) *_{\nu} b = b *_{\nu} (1 + \bar{f} *_{\nu} f) = 1. \tag{5.2}$$

Proof: The domain D_f is dense in $L^2(\mathbb{R}^{2n})$. It is well known that the operator $I + O_f^* O_f$ admits a continuous inverse B, whose domain is $L^2(\mathbb{R}^{2n})$ (see Ref. 23). The operator $O_{\bar{f}}$ is an extension of the operator O_f^*; the operator B has the property (4.7) (i). Therefore there is a b in $\mathscr{B}^{\nu}$ such that the relation (5.2) holds.

Corollary: Let $f \in O^{\nu}$. The adjoint operator O_f^* is $O_{\bar{f}}$.

B. Eigenvalue; Eigenelement

Let $f \in \mathscr{B}^{\nu} \cup \mathscr{O}^{\nu}$; the complex number a is an eigenvalue of f iff there is an element g ($\neq 0$), belonging to $\mathscr{B}^{\nu}$, such that

$$f *_{\nu} g = g *_{\nu} f = ag. \tag{5.3}$$

The corresponding element g is an eigenelement; obviously, it is not the only eigenelement.

Proposition: Let us consider a normal element $f \in \mathscr{B}^{\nu} \cup \mathscr{O}^{\nu}$ ($f *_{\nu} \bar{f} = \bar{f} *_{\nu} f$) and suppose that the complex a is an eigenvalue; then there is a greater projection π which is an eigenelement belonging to this eigenvalue a. For all eigenelements g (belonging to a),

$$\pi *_{\nu} g = g *_{\nu} \pi = g. \tag{5.4}$$

Proof: If the complex a is an eigenvalue of the operator O_f the projection operator on the corresponding eigenspace is defined by a projection p (Sec. IV B); it is the same for the operator $k \to k *_{\nu} f$. Let q be the associated projection; it remains to show the identity between p and q. For example, the duality between E^{ν} and Π^{ν} allows us to show that if f is a real element belonging to E^{ν}, one of the reals m and M [(4.9)], which has the largest absolute value, is an eigenvalue of f.

Resolvent and spectrum: Let $f \in \mathscr{B}^{\nu} \cup \mathscr{O}^{\nu}$. A complex number z belongs to the resolvent set $\rho(f)$ iff there is an element $r(z)$ in $\mathscr{B}^{\nu}$ such that

$$r(z) *_{\nu} (z - f) = (z - f) *_{\nu} r(z) = 1. \tag{5.5}$$

The spectrum $\sigma(f)$ of f is the complement of the resolvent set. Among the classical properties, we have, if f is a real element,

$$\rho(f) \supset \mathbb{C} \backslash \mathbb{R},\quad \forall z \in \mathbb{C} \backslash \mathbb{R},\quad \|r(z)\|^{\nu}_{\epsilon} \leqslant |\operatorname{Im} z|^{-1}. \tag{5.6}$$

VI. SPECTRAL RESOLUTION OF REAL OR UNITARY ELEMENTS

First, let us define the set $C^0(e)$. Let $t \mapsto e(t)$ be a resolution of identity; $C^0(e)$ is the set of the continuous functions from some neighborhood Ω of $\sigma(e)$ into $\mathbb{C}$.

A. Elements of $\mathscr{B}^{\nu}$ defined by a resolution of identity

Proposition: Let $t \mapsto e(t)$ be a resolution of identity and $\alpha \in C^0(e)$ be a bounded function on $\sigma(e)$. There is only one element f in $\mathscr{B}^{\nu}$ such that

$$\forall h \in \Pi^{\nu},\quad \langle f, h \rangle^{\nu} = \int_{\mathbb{R}} \alpha(t)\, d\, \langle e(t), h \rangle^{\nu}. \tag{6.1}$$

We have the following properties:

$$\|f\|^{\nu}_{\epsilon} = \operatorname*{Sup}_{t \in \sigma(e)} |\alpha(t)|,\quad \forall t,\quad f *_{\nu} e(t) = e(t) *_{\nu} f; \tag{6.2}$$

$$\forall k \in L^2(\mathbb{R}^{2n}),\quad (\|f *_{\nu} k\|^{\nu})^2 = \int_{\mathbb{R}} |\alpha(t)|^2\, d\, \langle e(t), k *_{\nu} \bar{k} \rangle^{\nu}; \tag{6.3}$$

let $g \in \mathscr{B}^{\nu}$ such that $\forall t,\quad e(t) *_{\nu} g = g *_{\nu} e(t)$,

$$\text{then } f *_{\nu} g = g *_{\nu} f; \tag{6.4}$$

and

$$\text{if } \pi(a) \neq 0,\quad \text{then } a \text{ is an eigenvalue.} \tag{6.5}$$

This comes about mainly because of the duality between $\mathscr{B}^{\nu}$ and Π^{ν}. If the element f (resp. g) is defined by the function α (resp. β), $f *_{\nu} g$ is defined by the function $\alpha \cdot \beta$; therefore f is normal. In particular, if the modulus of $\alpha(t)$ is the unity, f is unitary. If the function α is real (resp. positive), f is real (resp. "positive," see Sec. IV C). If α is the map $t \mapsto t$, the reals m and M [(4.9)] are the lower and upper bounds of $\sigma(e)$.

B. Elements of $\mathscr{O}^{\nu}$ defined by a resolution of identity

Fundamental result: Let $t \mapsto e(t)$ be a resolution of identity: $\alpha \in C^0(e)$, $k \in L^2(\mathbb{R}^{2n})$. The two following properties are equivalent:

(i) $\int_{\mathbb{R}} |\alpha(t)|^2\, d\, \langle e(t), k *_{\nu} \bar{k} \rangle^{\nu} < \infty$;

and (ii) there is a real C such that for all $k' \in L^2(\mathbb{R}^{2n})$, the integral

$$\int_{-\infty}^{\infty} \alpha(t)\, d\,\langle e(t) *_\nu k, k' \rangle^\nu$$

is convergent and has a smaller modulus than $C\,\|k'\|^\nu$. Let us denote by $D_e(\alpha)$ the set of functions for which the assertion (i) is true. This set $D_e(\alpha)$ is a linear dense subset of $L^2(\mathbb{R}^{2n})$.

Proposition: Let $t \mapsto e(t)$ be a resolution of identity; $\alpha \in C^0(e)$. Let us suppose that an element f, belonging to $\mathscr{O}^\nu$, is such that $\forall k \in D_e(\alpha)$, $\forall k' \in \mathscr{S}(\mathbb{R}^{2n})$,

$$\langle f *_\nu k, k' \rangle^\nu = \int_{-\infty}^{\infty} \alpha(t)\, d\,\langle e(t), k *_\nu k' \rangle^\nu. \qquad (6.6)$$

We have the following properties.

$\forall k \in D_e(\alpha)$, $f *_\nu k \in L^2(\mathbb{R}^{2n})$. Therefore, for all $k' \in L^2(\mathbb{R}^{2n})$, the property (6.6) is true. The relation (6.3) is applicable to the elements belonging to $D_e(\alpha)$.

$\mathscr{S}(\mathbb{R}^{2n}) \subset D_e(\alpha)$. Only one element belonging to $\mathscr{O}^\nu$, satisfies the relation (6.6).

$$\forall a,\quad b \in \mathbb{R},\quad f *_\nu e[a,b] = e[a,b] *_\nu f \in \mathscr{B}^\nu,$$

$$D_f = D_e(\alpha).$$

For all $g \in \mathscr{B}^\nu$, such that $f *_\nu g$ and $g *_\nu f$ belong to $\mathscr{B}^\nu$ and such that

$$\forall t,\ e(t) *_\nu g = e(t) *_\nu g, \quad \text{then } f *_\nu g = g *_\nu f.$$

Let f' be an element of $\mathscr{B}^\nu$, defined by the relation (6.1) with the function α'. We have $f *_\nu f' = f' *_\nu f$. For each element k belonging to $D_e(\alpha)$, $f *_\nu (f' *_\nu k)$ is defined by the relation (6.6) with the function $\alpha \cdot \alpha'$.

C. Spectral resolution of the real or unitary elements of $\mathscr{B}^\nu \cup \mathscr{O}^\nu$

Proposition: Let f a real (resp. unitary) element of $\mathscr{B}^\nu$; there is only one resolution of identity such that the relation (6.1) is true for the function α: $t \mapsto t$ [resp. $t \mapsto \exp(it)$].[21]

Proof: The method used by Riesz and Nagy[23] for the spectral resolution of the self-adjoint operators applies here. This method is based on the order relation that we have in the space $\mathscr{B}^\nu$.

Corollary: Let f a real (resp. unitary) element of $\mathscr{B}^\nu$; then we have the following.

The converse properties of (6.4) and (6.5) are true; or to be more precise, if a [resp. $\exp(ia)$] is an eigenvalue, $\pi(a)$ is the greatest eigenelement.

If f is unitary, $\sigma(e) \subset [0, 2\pi]$.

Proposition: Let f be a real element of $\mathscr{O}^\nu$. There is only one resolution of identity $t \mapsto e(t)$ such that the relation (6.6) is true; the function α is $t \mapsto t$.

Proof: Let us use the Cayley transform.[24] The element $1 + 2i\,r(-i)$ is unitary and does not admit the eigenvalue 1; therefore, its spectral resolution $\theta \mapsto \epsilon(\theta)$ is continuous at 2π. The spectral resolution of f is the map

$$t \mapsto e(t) = \epsilon(\theta), \quad t = -\cot(\theta/2).$$

Then, it is easy to prove the relation (6.6).

Corollary: Let f be a real element of $\mathscr{O}^\nu$. The converse property of (6.4) is true. The real a is an eigenvalue iff $\pi(a) \neq 0$; $\pi(a)$ is the greatest projection eigenelement of f.

Spectrum: Let f be a real element belonging to $\mathscr{B}^\nu \cup \mathscr{O}^\nu$; then the sets $\sigma(e)$ and $\sigma(f)$ are equal.

Proof: If the real $a \in \rho(e)$, the function $t \mapsto 1/(a-t)$ belongs to the set $C^0(e)$ and is bounded on $\sigma(e)$; therefore $r(a)$ exists. Then $\sigma(f) \subset \sigma(e)$. Conversely, if $a \in \sigma(e)$, there is a sequence of projections $p_j, j = 1,2,...$, such that

$$\forall j \in \mathbb{N},\quad \|p_j\|_\epsilon^\nu = 1,\quad \lim_{j \to \infty} (f-a) *_\nu p_j = 0, \quad \text{in } \mathscr{B}^\nu.$$

Therefore, this real a belongs to $\sigma(f)$.

These results on the eigenvalues and the eigenprojections have earlier been foreseen using the star-exponentials;[1,3] the calculations have been made in the case of the harmonic oscillator and in the case of any element belonging to the inhomogeneous symplectic Lie algebra.[25]

VII. CONCLUSIONS

Therefore it is possible to characterize the star-algebras of functions or distributions and real elements admitting a spectral resolution. The two algebras $\mathscr{O}_M^\nu$ and $\mathscr{B}^\nu$ have different properties towards, respectively, the spaces $\mathscr{S}(\mathbb{R}^{2n})$ and $L^2(\mathbb{R}^{2n})$. One of these spaces is nuclear, the other is not. The duality between $\mathscr{B}^\nu$ and Π^ν has a physical meaning: the states are, up to a factor, projections belonging to Π^ν (see Refs. 3 and 10). The observables are in $\mathscr{B}^\nu$ or more generally in $\mathscr{O}_M^\nu$. It is easy to prove the uncertainty relation of Heisenberg.

ACKNOWLEDGMENTS

I am most grateful to A. Lichnerowicz, who suggested this work to me and guided me in my research, to M. Flato who encouraged me, and to D. Sternheimer who gave me judicious advice.

[1] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. NY **111**, 61 (1978).
[2] A. Messiah, *Quantum Mechanics* (Interscience, New York, 1962), Vol. I.
[3] A. Lichnerowicz, Lecture Notes in Mathematics, Vol. 775 (Springer, New York, 1979), p. 105.
[4] S. De Groot, *La Transformation de Weyl et la fonction de Wigner: une forme alternative de la mécanique quantique* (Les Presses de l'Université de Montréal, Montreal, 1975).
[5] K. Chi Liu, J. Math. Phys. **17**, 859 (1976).
[6] M. Cahen and S. Gutt, Lett. Math. Phys. **5**, 219 (1981).
[7] J. Pool, J. Math. Phys. **7**, 66 (1966).
[8] J-M. Maillard, C. R. Acad. Sci. Paris **298**, 35 (1984).
[9] G. Loupias and S. Miracle-Sole, Commun. Math. Phys. **2**, 31 (1966).
[10] C. Moreno, *Produits * et analyse spectrale* (Actes des Journées Relativistes, Grenoble, 1981).
[11] D. Arnal and J. C. Cortet, *Produits et représentations des groupes nilpotents* (Journées de Géométrie Différentielle, Bruxelles, 1982).
[12] C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, NJ, 1946).
[13] M. Flato, A. Lichnerowicz, and D. Sternheimer, C. R. Acad. Sci. Ser. A **283**, 19 (1976).
[14] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).

[15]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1957).
[16]M. A. Antonets, Lett. Math. Phys. **2**, 241 (1978).
[17]J-B. Kammerer, C. R. Acad. Sci. Paris **298**, 59 (1984).
[18]R. Rubio, C. R. Acad. Sci. Paris, **299**, 699 (1984).
[19]L. Schwartz, *Analyse Hilbertienne* (Ecole Polytechnique, Palaiseau, 1978).
[20]A. Guichardet, *Algèbres d'observables associées aux relations de commutation* (Armand Colin, Paris, 1968).
[21]J-B. Kammerer, C. R. Acad. Sci. Paris **295**, 317 (1982).
[22]I. E. Segal, Math. Scand. **13**, 31 (1963).
[23]F. Riescz and B. Sz. Nagy, *Leçons d'analyse fonctionnelle* (Gauthier–Villars, Paris, 1968).
[24]K. Yosida, *Functional Analysis* (Springer, New York, 1974).
[25]F. Bayen and J-M. Maillard, Lett. Math. Phys. **6**, 491 (1982).

Propagation coefficients for fixed-isospin (T,T_z) average and related spectroscopic sum rules

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Tokyo, 153, Japan

(Received 3 June 1985; accepted for publication 30 September 1985)

The reduction relation for the fixed-isospin (T,T_z) average of a general operator in the model space of many fermions is described in two forms with and without recourse to factorization of isospin z components. Algebraic treatment is developed to deduce various types of expressions for each propagation coefficient that plays the role of the Green's function in each form of the reduction relation. Propagation coefficients are described also in relation to sum rules as to fixed-isospin spectroscopic factors. These results lead to novel identities among n-j symbols and factorials.

I. INTRODUCTION

A many-fermion system is often described by the model space generated from several interacting particles being distributed over a definite number of single-particle orbits or lattice sites.[1-9] Typical cases are seen in the study of atomic and nuclear structures[1-8] and statistical mechanical treatment of various phenomena.[9] A striking feature of the model space is the existence of the reduction relation for the operator average,[1-8] implying that a type of operator average in a many-body space could be expressed in terms of the same average in a few-body space. It, called propagation of operator average, underlies the statistical treatment of nuclear spectroscopy.[1-8]

The present work is devoted to the operator average in the model space with the total isospin T and its z component Tz being fixed, i.e., free from the average. The system of isospin-$\frac{1}{2}$ fermions is treated. The discussion can be extended to the average with intrinsic spin being fixed, if the terms (T,T_z) are only replaced by (S,S_z), respectively. The orbit or the site in the model space need not be specified explicitly. We can keep in mind the Fermi-gas model, or the spherical or deformed shell model with the Pauli principle being rigorously taken into account.

The fixed-isospin average of a general k-body operator $O(k)$ in the n-body space is defined by

$$\langle O(k)\rangle^{nTT_z} \equiv \frac{\Sigma_\mu \langle nTT_z\,\mu|O(k)|nTT_z\,\mu\rangle}{d(nTT_z)}, \tag{1}$$

where the index μ stands for a set of quantum numbers that specify, together with T and T_z, the complete orthonormalized states of the n-body space. The dimensionality of the n-body states with a given T and T_z is denoted as $d(nTT_z)$, a number independent of T_z. The propagation of the average (1) is expressed as[1,7,8]

$$\langle O(k)\rangle^{nTT_z} = \sum_{tt_z} Z(nTT_z,ktt_z)\langle O(k)\rangle^{ktt_z}, \tag{2}$$

where Z stands for the proportional coefficient, called the propagation coefficient.[1-7] The k-body state is specified by $|ktt_z\omega\rangle$ as the n-body state is by $|nTT_z\,\mu\rangle$. Some other types of averages, such as the fixed-seniority average,[1,3,4] lead to the same type of reduction relations. The propagation coefficient Z fulfills an analog of the Chapman–Kolmogorov equation and plays the role akin to the Green's function.[8] It is possible to express Z as

$$Z(nTT_z,ktt_z) = \sum_{t'} S'(nTT_z,ktt_z + k't't_z'), \tag{3}$$

with $k' = n - k$, $t_z' = t_z - T_z$, and

$$S'(nTT_z,ktt_z + k't't_z') = \binom{n}{k}(tt't_zt_z'|TT_z)^2 \sum_{\omega\omega'} \langle kt\omega + k't'\omega'|nT\mu\rangle^2 \tag{4}$$

$$\equiv (tt't_zt_z'|TT_z)^2 S(nT,kt + k't'), \tag{5}$$

where the symbol $\langle\ +\ |\ \rangle$ stands for the coefficient of fractional parentage (cfp).[10] The quantity S' or S, in which the z components of isospins are reduced, represents the sum of spectroscopic factors with all the isospins being fixed.

Previous works[2,4,6] concerned, in place of (2), the alternative form

$$\frac{\Sigma_\mu \langle nT\mu\|O^{(r)}(k)\|nT\mu\rangle}{d(nTT)} = \sum_t \frac{R(nT,kt,r)\,\Sigma_\omega \langle kt\omega\|O^{(r)}(k)\|kt\omega\rangle}{d(ktt)}, \tag{6}$$

where the double-barred matrix element is reduced with respect to isospins, and $O^{(r)}(k)$ is the k-body irreducible isotensor of rank r. The coefficient R relies on r, while Z in (2) relies on T_z and t_z. The relations (2) and (6) are associated with each other by virtue of the Wigner–Eckart theorem.[10]

French[2] deduced the explicit form of R. Another formalism[4] led to the same result. However, the result cited later [see Eq. (57)], is much involved and hardly manifests inherent properties.

The purpose of the present work is to deduce various types of expressions separately for Z and R that characterize (2) and (6), respectively. Each of the new expressions manifests itself in a transparent form and has the practical advantage of ease of manipulation. The quantity S, defined by (5), is shown to be expressed in terms of dimensionality of representations described by the same Young tableau,[11] which gives a sum rule as to fixed-isospin spectroscopic factors. Another spectroscopic sum rule is given to the sum of Z's. Apparently different expressions for the same quantity

 0022-2488/86/020536-08$02.50 © 1986 American Institute of Physics

$Z(R)$, as well as the sum rules, lead to types of novel identities among n-j symbols and factorials.

The search for the possibly simplest form of Z is intended for the use of (2) to the average of a product of several operators. Using (2) in place of (6), we can escape from complexity to decompose the operator product as a sum of irreducible isotensors.

We interpret Z and S (or S') as spectroscopic factors for the nuclear fragmentation in the high energy nuclear reaction. Simple expressions of Z and S will promote analyses of mass and charge distributions of the target or projectile residues.

A survey on quantities S, Z, and R is given in Sec. II as a preparation for later discussions. In particular, these quantities are shown to be equivalent to each other. Sections III, IV, and V are devoted to getting new expressions for S, Z, and R, respectively. In Sec. III, a recurrence relation for S is presented. Solving it, we get the compact expression of S so that Z and R can be expressed in terms of a Clebsch–Gordan (CG) coefficient squared and a Racah coefficient, respectively. In Sec. IV, we deduce a remarkably simple form of Z in the case of $|T_z| = T$ or $|t_z| = t$, the most familiar case in nuclear physics. In Sec. V, this result is translated into an expression of R that involves a single CG coefficient. Subsequently, another form of R is deduced, which is particularly suitable in the case where the isotensor rank of the operator is small. The result is elucidated in terms of information propagation that obeys the difference equation for R. Section VI concerns the prescription to deal with the average of a product of several operators. In Sec. VII, the quantities Z and S are described as spectroscopic factors for the target (or projectile) fragmentation.

New expressions for Z are given in (36) and (50). Those for R are in (37), (55), and (59), each of which is to be algebraically identical to the previous result (57). Spectroscopic sum rules are presented in (34) and (69).

II. A SURVEY ON PROPAGATION COEFFICIENTS

A general k-body operator $O(k)$ is written in the form

$$O(k) = \sum_{\gamma\gamma'} \langle k\gamma | O(k) | k\gamma' \rangle A^{+}(k\gamma) A(k\gamma') , \tag{7}$$

where $\gamma = (tt_z\omega)$, i.e., all the quantum numbers specifying the k-body state. The symbol $A^{+}(k\gamma)$ indicates the state operator[12,13] that creates the state $|k\gamma\rangle$ in case it acts on vacuum. The conjugate of $A^{+}(k\gamma)$ is denoted as $A(k\gamma)$. It is postulated that $A(k=0) = 1$. For the tensor analysis in the isospin space, it is convenient to define the operator $\tilde{A}$ by

$$\tilde{A}(ktt_z\omega) = (-1)^{t-t_z} A(kt(-t_z)\omega) . \tag{8}$$

The expression (2) in the case of $O(k) = A^{+}A$ reduces to

$$\sum_{\mu} \langle n\lambda\mu | A^{+}(k\nu\omega) A(k\nu'\omega') | n\lambda\mu \rangle = \delta(\nu,\nu')\delta(\omega,\omega') Z(n\lambda,k\nu) \frac{d(n\lambda)}{d(k\nu)} , \tag{9}$$

where $\lambda = (TT_z)$ and $\nu = (tt_z)$. An inherent feature of (9) is ω independence of Z. The dimensionality $d(ktt_z)$ is given by[1]

$$d(nTT_z) = (2T+1)(N/2+1)^{-1} \times \binom{N/2+1}{n/2+T+1} \binom{N/2+1}{n/2-T} \Delta(t-|t_z|) , \tag{10}$$

where N indicates the number of the single-particle states and

$$\Delta(a) = 1 , \quad \text{if } a \geq 0 , \quad \text{and 0 otherwise.} \tag{11}$$

In the fermion system, there exists a particle–hole symmetry of the matrix element[1]

$$\langle n\lambda\mu | A^{+}(k\nu\omega) A(k'\nu'\omega') | n'\lambda'\mu' \rangle = \langle N-n'\lambda'\mu' | A(k'\nu'\omega') A^{+}(k\nu\omega) | N-n\,\lambda\mu \rangle . \tag{12}$$

By virtue of (9) and (12) we get

$$Z(n\lambda,k\nu) = Z(N-k\nu, N-n\lambda) \frac{d(k\nu)}{d(n\lambda)} . \tag{13}$$

The expression (9) is rewritten by virtue of (12) as

$$\left\langle n\lambda\mu \left| \sum_{\omega} A^{+}(k\nu\omega) A(k\nu\omega) \right| n\lambda'\mu' \right\rangle = \delta(\lambda,\lambda')\delta(\mu,\mu') Z(n\lambda,k\nu) . \tag{14}$$

The matrix element of the state operator is proportional to cfp as[12,13]

$$\langle n\lambda\mu | A^{+}(k\nu\omega) | n'\lambda'\mu' \rangle = \sqrt{\binom{n}{k}} \langle k\nu\omega + n'\lambda'\mu' | n\lambda\mu \rangle , \tag{15}$$

where $k + n' = n$. The relation (12) with $k' = 0$ reduces to a particle–hole symmetry of cfp:

$$\sqrt{\binom{n}{k}} \langle k\nu\omega + n'\lambda'\mu' | n\lambda\mu \rangle = \sqrt{\binom{N-n'}{k}} \langle k\nu\omega + (N-n)\,\lambda\mu | (N-n')\,\lambda'\mu' \rangle . \tag{16}$$

Expressing the left-hand side (lhs) (14) in terms of cfp yields

$$\binom{n}{k} \sum_{\omega\omega'\nu} \langle k\nu\omega + k'\nu'\omega' | n\lambda\mu \rangle \langle k\nu\omega + k'\nu'\omega' | n\lambda'\mu' \rangle = \delta(\lambda,\lambda')\delta(\mu,\mu') Z(n\lambda,k\nu) , \tag{17}$$

which leads to (3) followed by (4). The factorization of cfp[14] gives

$$\binom{n}{k}\sum_{\omega\omega'}\langle k\nu\omega + k'\nu'\omega'|n\lambda\mu\rangle\langle k\nu\omega + k'\nu'\omega'|n\lambda\mu'\rangle$$
$$= \delta(\mu,\mu')\times\text{a term independent of }\mu$$
$$= \delta(\mu,\mu')S'(n\lambda,k\nu + k'\nu'), \tag{18}$$

where (4) is used in the last step. We see that the quantity S' and, therefore, S are independent of μ. Substituting $O^{(r)}(k) = [A^{+}\tilde{A}]^{(r)}$ into (6) yields

$$\sum_{\mu}\langle nT\mu\|[A^{+}(kt\omega)A(kt'\omega')]^{(r)}\|nT\mu\rangle$$
$$= \delta(t,t')\delta(\omega,\omega')(-1)^{2t}\sqrt{2r+1}\,R(nT,kt,r)\frac{d(nTT)}{d(ktt)}. \tag{19}$$

Applying the Wigner–Eckart theorem to the lhs of (9) to link (9) with (19), we get

$$Z(nTT_z,ktt_z) = \frac{\Sigma_r(trt_z0|tt_z)(TrT_z0|TT_z)(2r+1)R(nT,kt,r)}{\sqrt{(2t+1)(2T+1)}}. \tag{20}$$

The inverse is given by

$$R(nT,kt,r) = \sqrt{\frac{(2T+1)}{(2t+1)}}\frac{\Sigma_{t_z}(trt_z0|tt_z)Z(nTT,ktt_z)}{(TrT0|TT)}. \tag{21}$$

Combining (21) with (20) yields

$$Z(nTT_z,ktt_z)$$
$$= \sum_r(2r+1)(trt_z0|tt_z)(TrT_z0|TT_z)$$
$$\times\frac{\Sigma_{t_z'}(trt_z'0|tt_z')Z(nTT,ktt_z')}{(2t+1)(TrT0|TT)}. \tag{22}$$

It indicates that the set $\{Z(nTT_z,ktt_z);\ T_z = T\}$ is sufficient to determine all the set of Z's. Expanding the lhs of (19) in terms of cfp's, we get

$$R(nT,kt,r) = (2T+1)\sum_{t'}W(Trt't;Tt)S(nT,kt+k't'), \tag{23}$$

which is inverted as

$$S(nT,kt+k't') = (2t'+1)$$
$$\times\frac{\Sigma_r(2r+1)W(Trt't;Tt)R(nT,kt,r)}{(2T+1)}. \tag{24}$$

We see from (20)–(24) that the quantities S, R, and Z with $T_z = T$ are transformed into each other.

The propagation coefficient Z fulfills[8]

$$\sum_{\lambda'}Z(n\lambda,n'\lambda')Z(n'\lambda',k\nu) = \binom{n-k}{n'-k}Z(n\lambda,k\nu), \tag{25}$$

a relation akin to the Chapman–Kolmogorov equation. The arguments n and λ play the roles of time and space, respectively. The relation (25) with $n' = n-1$ reads

$$\sum_{\lambda'}Z(n\lambda,n-1\lambda')Z(n-1\lambda',k\nu) = (n-k)Z(n\lambda,k\nu). \tag{26}$$

The quantity $Z(n\lambda,n-1\lambda')$ here is written as

$$Z(n\lambda,n-1\lambda')$$
$$= (T'1/2T_z't_z'|TT_z)^2S(nT,(n-1)T'+1), \tag{27}$$

with

$$S(nT,(n-1)T'+1)$$
$$= \frac{2T'+1}{2T+1}\times\begin{cases} n/2+T+1, & \text{for } T' = T-\frac{1}{2},\\ n/2-T, & \text{for } T' = T+\frac{1}{2},\end{cases} \tag{28}$$

where in (27) we have used (3) and (5), and in (28) the fixed-isospin sum rule on one-particle cfp's.[4] Substituting the explicit form of $Z(n\lambda,n-1\lambda')$ into (26), we get the recurrence relation,[8] i.e., the difference equation for Z to be solved under the condition

$$Z(k\nu,k\nu') = \delta(\nu,\nu'), \tag{29}$$

a property of Z seen from (9). The difference equation for Z is translated into that for R as

$$(n-k)(2T+1)R(nT,kt,r)$$
$$= (n/2+T+1)\sqrt{(2T-r)(2T+r+1)}$$
$$\times R(n-1\,T-\tfrac{1}{2},kt,r)$$
$$+ (n/2-T)\sqrt{(2T-r+1)(2T+r+2)}$$
$$\times R(n-1\,T+\tfrac{1}{2},kt,r). \tag{30}$$

III. THE SUM OF FIXED-ISOSPIN SPECTROSCOPIC FACTORS

Here we express S, defined by (5), in a compact form that leads to a new expression for each of Z and R.

Let us show the recurrence relation for S

$$S(nT,kt+(n-k)t')$$
$$= S((n-k)t',(n-k-1)t''+1)^{-1}\sum_{T''}(2t'+1)$$
$$\times(2T'+1)W(tTt''1/2;t'T')^2$$
$$\times S(nT,(n-1)T'+1)$$
$$\times S((n-1)T',kt+(n-k-1)t''). \tag{31}$$

Proof: We start from the identity

$$\langle i|C_\gamma A(ktt_z\omega)|j\rangle = (-1)^k\langle i|A(ktt_z\omega)C_\gamma|j\rangle, \tag{32}$$

where $\langle i| = \langle(n-k-1)t''t_z''\omega''|$ and $|j\rangle = |nTT_z\mu\rangle$. Expressing both sides of (32) in terms of cfp's, we factorize isospin z components in cfp's. Then, we get

$$\sum_{\omega'}\langle kt\omega + (n-k)t'\omega'|nT\mu\rangle\langle (n-k-1)t''\omega'' + 1\omega_0|(n-k)t'\omega'\rangle$$

$$= \sum_{T'}(-1)^{t-t''+T'}\sqrt{(2t'+1)(2T'+1)}W(tTt''\tfrac{1}{2};t'T')$$

$$\times \sum_{\mu'}\langle kt\omega + (n-k-1)t''\omega'|(n-1)T'\mu'\rangle\langle (n-1)T'\mu' + 1\omega_0|nT\mu\rangle\,. \tag{33}$$

Let us square the expression on each side and subsequently impose the sum over ω, ω'', and ω_0 on each side. After carrying out the sum by virtue of (18), we obtain (31).

The recurrence relation (31) leads to

$$S(nT,kt+k't') \equiv \binom{n}{k}\sum_{\omega\omega'}\langle kt\omega + k't'\omega'|nT\mu\rangle^2$$

$$= \binom{n}{k}\frac{f(k,t)f(k',t')}{f(n,T)}\Delta'(tt'T), \tag{34}$$

with

$$f(n,T) = (2T+1)\binom{n+1}{n/2-T}(n+1)^{-1}, \tag{35}$$

where Δ' is the triangle condition among T, t, and t'.

Proof: We use induction on n with k, t, and ω being fixed. The expression (34) in the case of $n = k$ properly reduces to unity. In the case of $n = k+1$, it agrees with (28). Supposing (34) for $n = p$, let us deduce it for $n = p+1$. We substitute (34) for $n = p$ into the right-hand side (rhs) of (31) for $n = p+1$. After applying the orthogonality for the Racah coefficient to the sum over T' in (31), we obtain (34) for $n = p+1$.

The quantity $f(n,T)$ is just the dimensionality of the representations described by the same Young tableau[11] $[h_1,h_2]$, where $h_1 - h_2 = 2T$ and $h_1 + h_2 = n$. In the branching diagram,[4] it is illustrated as the number of allowed paths between (n,T) and $(1,\frac{1}{2})$. We can elucidate (34) by the branching rule of the Young tableau.

The relation (34) is viewed as the sum rule on spectroscopic factors with the involving isospin quantum numbers being fixed. Rewriting it in terms of intrinsic spins in place of isospins, we get the fixed-intrinsic-spin sum rule, which gives much fine information in comparison with the sum rule discussed in Ref. 15.

The relation (34) incorporated into (3) and (23), respectively, yields

$$Z(nTT_z,ktt_z) = \sum_{t'}(tt't_zt_z'|TT_z)^2\binom{n}{k}\frac{f(k,t)f(k',t')}{f(n,T)} \tag{36}$$

and

$$R(nT,kt,r)$$

$$= (2T+1)\sum_{t'}W(Trt't;Tt)\binom{n}{k}\frac{f(k,t)f(k',t')}{f(n,T)}. \tag{37}$$

We see from (34), (36), and (37) that all of S, Z, and R are independent of N, the number of the single-particle states. The relation (36) shows that the quantity

$$Z(nTT_z,ktt_z)f(n,T)/\{(2T+1)f(k,t)\} \tag{38}$$

is symmetric with respect to the simultaneous interchanges of T and t and of T_z and t_z. The similar symmetry property of R is found from (37). The expression of R given by (37) should be a solution of (30), though it is far from obvious.

IV. THE PROPAGATION COEFFICIENT Z

In this section, the propagation coefficient Z is shown to be a very simple form in the case of $|T_z| = T$, regardless of the magnitude of T relative to n.

In order to deduce another type of expression (equation) for Z other than (36) [(26)], we utilize the p-n formalism that deals with protons and neutrons separately. The state operator of creating k_1 neutrons and k_2 protons is written as

$$A^+(k_1k_2\omega_1\omega_2) = A^+(k_1\omega_1)A^+(k_2\omega_2), \tag{39}$$

where ω_1 stands for all the quantum numbers specifying the k_1-body state with isospin $t_1 = t_{1z} = k_1/2$, and ω_2 for the quantum numbers specifying the k_2-body state with $t_2 = -t_{2z} = k_2/2$. The operator defined by (39) produces the state $|k_1k_2\omega_1\omega_2\rangle$ of k_1 neutrons and k_2 protons. Notice that this state does not have a definite isospin.

The operator specified in the framework of isospin formalism is expanded as

$$A^+(ktt_z\omega)A(kt't_z'\omega')$$

$$= \sum A^+(k_1k_2\omega_1\omega_2)A(k_1'k_2'\omega_1'\omega_2')$$

$$\times\langle k_1k_2\omega_1\omega_2|A^+(ktt_z\omega)A(kt't_z'\omega')|k_1'k_2'\omega_1'\omega_2'\rangle, \tag{40}$$

where

$$k_1 + k_2 = k \quad \text{and} \quad (k_1 - k_2)/2 = t_z, \tag{41}$$

and the same condition is applied to k_1' and k_2'. The sum on the rhs of (40) is taken over ω_1, ω_2, ω_1', and ω_2'.

There exists the completeness relation

$$\sum_{\mu T}|nTT_z\mu\rangle\langle nTT_z\mu|$$

$$= \sum_{\mu_1\mu_2}|n_1n_2\mu_1\mu_2\rangle\langle n_1n_2\mu_1\mu_2|, \tag{42}$$

where T on the lhs runs over $|T_z|$, $|T_z|+1$, ..., and $n/2$. The condition (41) is applied to n_1 and n_2. By virtue of (42), the transformation coefficient on the rhs of (40) fulfills

$$\sum_{\omega_1\omega_2}\langle k_1k_2\omega_1\omega_2|A^+(ktt_z\omega)A(kt't_z'\omega')|k_1k_2\omega_1\omega_2\rangle$$

$$= \delta(\gamma,\gamma'), \tag{43}$$

where $\gamma=(tt_z\omega)$, etc.

By virtue of (40) and (42), the T_z-fixed trace in the framework of the isospin formalism is transformed as

$$\sum_{\mu T}\langle nTT_z\mu|\,A^+(ktt_z\omega)\,A(kt't'_z\omega')|nTT_z\mu\rangle$$
$$=\sum\langle n_1n_2\mu_1\mu_2|\,A^+(k_1k_2\omega_1\omega_2)$$
$$\times A(k'_1k'_2\omega'_1\omega'_2)|n_1n_2\mu_1\mu_2\rangle$$
$$\times\langle k_1k_2\omega_1\omega_2|\,A^+(ktt_z\omega)\,A(kt't'_z\omega)|k'_1k'_2\omega'_1\omega'_2\rangle\,, \tag{44}$$

where the sum on the rhs is taken over $\mu_1,\mu_2,\omega_1,\omega_2,\omega'_1$, and ω'_2. The sum over μ_1 (μ_2) is done by virtue of

$$\sum_{\Gamma}\langle n\Gamma|\,A^+(k\gamma)\,A(k\gamma')|n\Gamma\rangle=\delta(\gamma,\gamma')\binom{N/2-k}{n-k}, \tag{45}$$

the well-known relation[1] that describes the propagation of the operator average with all the quantum numbers being summed. The subsequent sum over ω_1 and ω_2 on the rhs of (44) is done by virtue of (43). Then, we get

$$\sum_{\mu T}\langle nTT_z\mu|\,A^+(ktt_z\omega)\,A(kt't'_z\omega')|nTT_z\mu\rangle$$
$$=\delta(\gamma,\gamma')\binom{N/2-k_1}{n_1-k_1}\binom{N/2-k_2}{n_2-k_2}. \tag{46}$$

Applying (9) to the lhs and using (41) to rewrite the rhs in the framework of the isospin formalism, we get the basic equation for Z, in a form different from (26),

$$\sum_T Z(nTT_z,ktt_z)d(nTT_z)$$
$$=d(ktt_z)\binom{(N-k)/2-t_z}{(n-k)/2+T_z-t_z}$$
$$\times\binom{(N-k)/2+t_z}{(n-k)/2-T_z+t_z}\equiv F(N). \tag{47}$$

The isospin T in any of (44), (46), and (47) runs over $|T_z|$, $|T_z|+1,\ldots$, and $t+(n-k)/2$.

It is a little too involved to solve (47) in a direct way. A device to circumvent the difficulty is to substitute $N=n+2T_0$ into (47), where T_0 is a parameter fulfilling

$$|T_z|\leqslant T_0<t+(n-k)/2,$$

and

$$T_0+n/2=\text{an integer}. \tag{48}$$

This device yields

$$\sum_T Z(nTT_z,ktt_z)d(nTT_z)_{N=n+2T_0}=F(n+2T_0), \tag{49}$$

where T runs only over $|T_z|$, $|T_z|+1,\ldots$, and T_0. Derivation of (49) relies first on N independence of Z, and second on vanishing of $d(nTT_z)$ with $N=n+2T_0$ in the case of $T_0<T$: These are distinctive features seen in (36) and (10), respectively.

Let us put $T_0=T_z$ ($\geqslant 0$) in (49) so that T can be fixed uniquely to T_z. Then, it follows that

$$Z(nTT,ktt_z)$$
$$=(2t+1)\binom{(n-k)/2+T+t_z}{2T}\binom{n/2+T+1}{k/2+t+1}$$
$$\times\binom{n/2+T+1}{k/2-t}\left\{(2T+1)\binom{n/2+T+1}{2T+1}\right\}^{-1}$$
$$\times\Delta(t-|t_z|). \tag{50}$$

The simple form on the rhs is worth notice. The rhs should be identical to that of (36) with $T_z=T$, though apparently very different. Combining (36) and (50) yields a novel identity involving a CG coefficient squared. The rhs of (50) with $k=1$ reduces to $n/2\pm T$ for $t_z=\pm\frac{1}{2}$, respectively, as is expected from the numbers of neutrons and protons of the state with $T_z=T$. Using (13) and (50), we get

$$Z(nTT_z,ktt)=\binom{(n-k)/2+t+T_z}{2t}\binom{n/2-T}{k/2-t}$$
$$\times\binom{n/2+T+1}{k/2-t}\binom{k/2+t+1}{2t+1}^{-1}$$
$$\times\Delta(T-|T_z|)\,. \tag{51}$$

In this derivation, the N independence of the rhs is associated with the property that the rhs of (13), comprised of N-dependent factors, does not depend on N. The symmetry property of the quantity (38) can be checked from (50) and (51).

The expression (50) together with the symmetry

$$Z(nT(-T_z),kt(-t_z))=Z(nTT_z,ktt_z), \tag{52}$$

seen from (3) or (36), will facilitate the statistical treatment of nuclear spectroscopy, since low-lying states in the nucleus are characterized by $T=|T_z|$, i.e., the lowest isospin.

The propagation coefficient $Z(nTT_z,ktt_z)$ in case T_z is close to T is easily deduced from (50) being incorporated with the recurrence

$$(T+T_z)(T-T_z+1)Z(nT(T_z-1),ktt_z)$$
$$=2\{T(T+1)-T_z^2-t(t+1)+t_z^2\}Z(nTT_z,ktt_z)$$
$$-(T-T_z)(T+T_z+1)Z(nT(T_z+1),ktt_z)$$
$$+(t-t_z)(t+t_z+1)Z(nTT_z,kt(t_z+1))$$
$$+(t+t_z)(t-t_z+1)Z(nTT_z,kt(t_z-1)), \tag{53}$$

which results from the n-body average of the identity

$$T_+A^+AT_-$$
$$=T_-T_+A^+A-T_-A^+AT_++A^+AT_+T_-$$
$$-[T_-[T_+,A^+]]A-A^+[T_-[T_+,A]]$$
$$-[T_+,A^+][T_-,A]-[T_-,A^+][T_+,A]. \tag{54}$$

V. THE PROPAGATION COEFFICIENT R

In this section, types of expressions for R are shown, which are apparently different from (37).

Substituting (50) into (21) yields

$$R(nT,kt,r)$$
$$=\sqrt{(2t+1)(2T-r)!(2T+r+1)!}$$
$$\times\left(\frac{n}{2}-T\right)!\left\{\left(\frac{k}{2}-t\right)!\left(\frac{n-k}{2}+T+t+1\right)!\right\}^{-1}$$
$$\times\binom{n/2+T+1}{k/2+t+1}\sum_{t_z}(trt_z0|tt_z)$$
$$\times\binom{(n-k)/2+T+t_z}{2T}, \tag{55}$$

where the CG coefficient is explicitly written as[16]

$$(trt_z0|tt_z)$$
$$=\sqrt{(2t+1)}\{(2t-r)!(2t+r+1)!\}^{-1/2}$$
$$\times\sum_u(-1)^u\frac{(r+u)!(2t-u)!}{u!(r-u)!}\binom{t-t_z}{u}\Delta(t+t_z). \tag{56}$$

The explicit form of $(TrT0|TT)$ is used to get (55). We notice that the CG coefficient of the form (56) appears also in the Gram expansion of the finite rotation operator.[17]

French[2] got the following result [(7.40) of Ref. 2 being multiplied by $\sqrt{2t+1}$]:

$$R(nT,kt,r)=\frac{(2t+1)\,(n/2-T)!(n/2+T+1)!}{(n-k)!(k/2-t)!(k/2+t+1)!}\sqrt{\frac{(2t-r)!(2T+r+1)!}{(2T-r)!(2t+r+1)!}}$$
$$\times\sum_p\frac{(-1)^p(2t-p)!(n+2t-r-k-2p)!}{p!(2t-r-2p)!((n-k)/2+t-T-p)!((n-k)/2+T+t+1-p)!}. \tag{57}$$

Comparing (55) with (57), we find that the involved sum in the latter is rewritten in the former as the sum of products comprised of a CG coefficient and a binomial coefficient.

Let us carry out the sum over t_z on the rhs of (55). We discuss first the case when the value of r is only a few. Expanding the CG coefficient as a polynomial in t_z, we make use of the identity

$$\sum_{k=-t}^{t}\binom{n+k}{r}=\binom{n+1+t}{r+1}-\binom{n-t}{r+1}, \tag{58}$$

etc. to sum over t_z. The result for $r\leqslant 2$ is summarized as

$$R\,(nT,kt,r)$$
$$=(2t+1)(2T+1)\sqrt{\frac{(2t-r)!(2T-r)!}{(2t+r+1)!(2T+r+1)!}}$$
$$\times R'(nT,kt,r), \tag{59a}$$

$$R'(nT,kt,r)$$
$$=\left\{\binom{n/2+T+1}{k/2+t+1}\binom{n/2-T}{k/2-t}B_+\right.$$
$$\left.-\binom{n/2+T+1}{k/2-t}\binom{n/2-T}{k/2+t+1}B_-\right\}(2T+1)^{-1}, \tag{59b}$$

$$B_\pm=\begin{cases}1, & \text{for } r=0,\\ -(n-k+1)\pm(2t+1)(2T+1), & \text{for } r=1,\\ 3(n-k+1)^2\mp 3(n-k+1)(2t+1)(2T+1)\\ \quad+16t(t+1)T(T+1), & \text{for } r=2.\end{cases} \tag{59c}$$

The expression for the case of $r=0$ is particularly simple. It was deduced previously in Ref. 4 as the propagation coefficient for the average with T (but not T_z) being fixed. We see from (59c) that the coefficient B_- is related to B_+ by

$$B_-=(B_+)_{T\to -T-1}. \tag{60}$$

From this we find that two terms in the bracket on the rhs of (59b) are interchanged under the replacement of T by $-T-1$ and that the quantity R' is a polynomial in $T(T+1)$.

Next, let us extend (59c) to a general r, retaining (59a) and (59b) as they are. In order to get the form of B_+, we consider the case of $(n-k)/2\leqslant T+t$, under which the term containing B_- vanishes. Substituting (56) into (55), let us sum over t_z as

$$\sum_{t_z}\binom{t-k}{u}\binom{(n-k)/2+T+t_z}{2T}\Delta(t+t_z)$$
$$=\binom{(n-k)/2+T+t+1}{2T+u+1}, \tag{61}$$

valid only for $(n-k)/2\leqslant T+t$ because of existence of $\Delta(t+t_z)$ on the lhs. We then get

$$B_+=\sum_{u=0}^{r}(-1)^u\frac{(2t-u)!(r+u)!}{(2t-r)!}$$
$$\times\binom{(n-k)/2-T+t}{u}\binom{2T+r+1}{r-u}. \tag{62}$$

We easily deduce the expression of B_-, applying (60) to (62).

The implication of (59a)–(59c) as well as (57) is elucidated in terms of solutions of the difference equation (30). As discussed in Ref. 4, every term with the definite p on the rhs of (57) fulfills (30). The linear combination of terms with various p's is to reproduce the condition (29). Each term with the definite p describes information propagation between the points (n,T) and $(k+r-2t+2p,r/2)$ in the branching diagram.

The expressions (59a)–(59c) are split into two parts, one containing only B_+ and the other containing B_-. We find that each part fulfills (30). The part containing B_+ can be interpreted as information propagation between the points (n,T) and (k,t) in the branching diagram, as is seen from the vanishing of the expression in the case of $n=k$ and $T\neq t$. Similarly, the part containing B_- describes the connection between (n,T) and $(k,-t-1)$. Notice that the point $(k,-t-1)$ is unphysical, since the isospin is negative. It is a mirror image of (k,t) with respect to the isospin $=-\frac{1}{2}$ axis. The mirror image takes the place of the condition that prop-

agation should be confined to the region of isospin $\geqslant r/2$. The symmetry (60) implies that the propagation between (n,T) and $(k,-t-1)$ is the same as that between $(n,-T-1)$ and (k,t).

New expressions (37), (55), and (59) as well as the previous result (57), which are given to the same quantity R, are to be algebraically identical to each other, though we hardly see it except in the case of $r=0$, the isoscalar operator. Combining two of them leads to a novel identity among n-j symbols and factorials. From the first two, for example, we get, for any integer p,

$$\sum_{t'} \frac{(2t'+1)W(Trt't;Tt)}{(p-t')!(p+t'+1)!} = \sqrt{\frac{(2T-r)!(2T+r+1)!}{(p+T+t+1)!(p+T-t)!}} \sum_{t_z} (trt_z 0|tt_z) \times \binom{p+T+t_z}{2T} (2t+1)^{-1/2}, \quad (63)$$

a relation somewhat akin to the novel identity (1) of Ref. 18.

VI. THE FIXED-ISOSPIN AVERAGE OF THE OPERATOR PRODUCT

The operator subject to the average is often a power of an operator or a product of several operators so that the direct use of (2) would not be available. Here, we present a prescription to deal with the n-body average of the operator

$$O \equiv A^+(n_1\gamma_1)A(n_1'\gamma_1')A^+(n_2\gamma_2)\cdots A(n_f'\gamma_f'), \quad (64)$$

where $\gamma_1 = T_1 T_{1z} \mu_1$, etc. and any of n_1, n_2, etc. can be zero. Usually, it is easy to express a given operator as a sum of O's.

Averaging the operator O in the n-body space requires three steps. First, the extended Wick's theorem[13] is applied to O to rewrite it as a sum of normal products. The resultant expression is summarized in (38) of Ref. 13. The normal product that is relevant to the operator average is of the form (7) being subdivided as

$$O(k) = O_+(k)O_-(k) \quad (65)$$

with

$$O_+(k) = A^+(k_1\alpha_1)A^+(k_2\alpha_2)\cdots A^+(k_f\alpha_f) \quad (66a)$$

and

$$O_-(k) = A(k_1'\alpha_1')A(k_2'\alpha_2')\cdots A(k_f'\alpha_f'), \quad (66b)$$

where $\alpha_1 = t_1 t_{1z}\omega_1$, etc. The sum of $k_1, k_2, \cdots$, and k_f is equal to k, the sum of $k_1', k_2', \cdots$, and k_f'. Next, the reduction relation (2) is applied to the n-body average of $O(k)$ so as to express it as a sum of k-body averages. The last step concerns modification of the resultant k-body average, which is rearranged as

$$\langle O(k)\rangle^{ktt_z} = \langle 0|O_-(k)P(tt_z)O_+(k)|0\rangle / d(ktt_z), \quad (67)$$

where $P(tt_z)$ stands for projection operator onto the isospin-tt_z space. Operating $P(tt_z)$ to the right, we state it in terms of CG coefficients so that Wick's theorem can be applied to the antinormal ordering form O_-O_+ in the matrix element. The n-body average of O is finally given as a sum of vacuum expectation values.

Notice that rearranging any operator as a sum of isotensors is not required in the first two steps. This advantage becomes remarkable when the number of A^+'s and A's consisting of O increases.

VII. USE OF Z AND S AS SPECTROSCOPIC FACTORS FOR NUCLEAR FRAGMENTATION

Here, we point out that it is available to use Z and S' as formation factors of target (or projectile) residues in high energy nuclear reaction.

Let us consider the case when the mass-n target nucleus in any state with isospin (T,T_z) is fragmented into two pieces (residues), one being a cluster of k nucleons coupled to isospin (t,t_z) and the other a cluster of k' $(=n-k)$ nucleons coupled to (t', t_z'). We assume that specification of the target residue by angular momentum, etc. other than isospin is not a matter of concern, which will be meaningful in analyses of mass and charge distributions of target (projectile) residues.[19]

We set up three types of spectroscopic factors for the fragmentation. The first, used in the prevalent treatment,[19] is of the hypergeometric expression

$$\binom{n_1}{k_1}\binom{n_2}{k_2}. \quad (68)$$

The condition (41) is applied to n_1, k_1, etc. That is, the argument n_1 (n_2) implies the number of neutrons (protons) of the target nucleus and k_1 (k_2) the number of neutrons (protons) of one of residues. The second and the third types of spectroscopic factors are $Z(nTT_z, ktt_z)$ and $S'(nTT_z, ktt_z + k't't_z')$. These are to be used for inclusive and exclusive analyses, respectively, regarding isospin t.

None of the isospins t and t' are taken into account in (68) contrary to Z or S'. From this, we infer the spectroscopic sum rule

$$\sum_t Z(nTT_z, ktt_z) = \binom{n_1}{k_1}\binom{n_2}{k_2}. \quad (69)$$

To prove it, we have only to substitute (13) into the lhs of (47) after replacing n and k in (47) by $N-n$ and $N-k$, respectively. The relation (69) linked with (36) leads to a novel identity containing a CG coefficient squared.

It has little been discussed how the fragmentation process is influenced by the magnitude of t relative to that of t_z. It is, however, known that the nucleus with $t > |t_z|$ has fairly large excitation energy and is unstable as to particle decays. Considering this, we present an extreme model that target residues are formed mainly under the condition $|t_z| = t$. The relevant spectroscopic factor is expressed as

$$Z(nTT,kt\,|t_z| = t) = \frac{(n_1+1)(k'-k''+1)}{(k'+1)(n_1-k''+1)}\binom{n_1}{k_1}\binom{n_2}{k_2}, \quad (70)$$

where k' (k'') indicates the larger (smaller) number between k_1 and k_2. The relation (50) is used to get (70). Comparing (70) with (68), we find that the former leads to a slight enhancement of production of β-unstable nuclei. Existing experimental data is not sufficient to support further discussion on this point.

It is easy to extend (34) so as to get spectroscopic factors for a nucleus being fragmented into several clusters. We present here only the expression

$$S(nTT_z,k_1t_1t_1+k_2t_2t_2+\cdots+k_1't_1'(-t_1')+k_2't_2'(-t_2')+\cdots)=(tt't-t'|TT_z)^2\times\frac{n!f(k_1,t_1)f(k_2,t_2)\cdots f(k_1',t_1')f(k_2',t_2')\cdots}{k_1!k_2!\cdots k_1'!k_2'!\cdots f(n,T)},\quad(71)$$

where t (t') stands for the sum of $t_1, t_2, \cdots (t_1', t_2', \cdots)$.

[1]J. B. French, in *Nuclear Structure*, edited by A. Hossain, Harum-ar-Raschid, and M. Islam (North-Holland, Amsterdam, 1967), p. 85.
[2]J. B. French, in *Isospin in Nuclear Physics*, edited by D. H. Wilkinson (North-Holland, Amsterdam, 1969), p. 259.
[3]M. Nomura, Prog. Theor. Phys. **47**, 1858 (1972).
[4]M. Nomura, Prog. Theor. Phys. **51**, 489 (1974).
[5]K. T. Hecht and J. P. Draayer, Nucl. Phys. A **223**, 285 (1974).
[6]J. P. Draayer, in *Theory and Applications of Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980), p. 213.
[7]C. Quesne, in Ref. 6, p. 257.
[8]M. Nomura, J. Math. Phys. **26**, 965 (1985).
[9]D. Walikainen and R. B. McQuistan, J. Math. Phys. **26**, 815 (1985).
[10]de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[11]I. G. Kaplan, *Symmetry of Many-Electron Systems*, translated by J. Gerratt (Academic, New York, 1975).
[12]M. Nomura, Prog. Theor. Phys. **59**, 451 (1978).
[13]M. Nomura, J. Math. Phys. **26**, 732 (1985).
[14]G. Racah, Phys. Rev. **76**, 1352 (1949).
[15]E. McGuire, Phys. Rev. A **10**, 32 (1974).
[16]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957), p. 44.
[17]W. Happer, Ann. Phys. NY **48**, 579 (1968).
[18]B. I. Dunlap and B. R. Judd, J. Math. Phys. **16**, 318 (1975).
[19]D. J. Morrissey, W. R. Marsh, R. J. Otto, W. Loveland, and G. T. Seaborg, Phys. Rev. C **18**, 1267 (1978).

Summation-free expression for some special Clebsch–Gordan coefficients

M. A. Rashid
Mathematics Department, Ahmadu Bello University, Zaria, Nigeria

(Received 22 August 1984; accepted for publication 3 October 1985)

The known general formula for the Clebsch–Gordan coefficients of the three-dimensional rotation group involves one summation that results in explicit summation-free expressions for the coefficients where either one of the angular momenta is the sum of the other two or the magnetic quantum number corresponding to one of the angular momenta takes its maximum value in magnitude. By using very different techniques, explicit expressions for the coefficients $\langle j_1, 0; j_2, 0 | j, 0\rangle$, $\langle j_1, \frac{1}{2}; j_2, -\frac{1}{2} | j, 0\rangle$ are also obtained where the integral or half-integral nature of the j's is indicated by the magnetic quantum number involved. Here the expressions depend upon whether $j_1 + j_2 + j$ is an even or an odd integer. For these coefficients, the magnetic quantum numbers involved take their minimum value in magnitude. By using the recursion relation for the coefficients of the form $\langle j_1, m, j_2, -m, | j, 0\rangle$, these coefficients can be calculated in terms of the above known ones provided the explicit value of the coefficient $\langle j_1, 1; j_2, -1 | j, 0\rangle$ is known, where $j_1 + j_2 + j$ is an odd integer. (The recursion relation for these coefficients in terms of $\langle j_1, 0; j_2, -1 | j, 0\rangle$ becomes a triviality since $\langle j_1, 0; j_2, 0 | j, 0\rangle$ vanishes when $j_1 + j_2 + j$ is an odd integer.) The main purpose of this paper is to give an explicit expression for the coefficients $\langle j_1, 1; j_2, -1 | j, 0\rangle$, where $j_1 + j_2 + j$ is an odd integer. This expression is obtained by using a complicated transformation between hypergeometric functions, which seems to have been neglected so far. For the coefficients where the magnetic quantum numbers have their minimum value in magnitude, this transformed expression becomes summation-free and the explicit values of the three already known coefficients and the fourth so far unknown are obtained. Further study of this transformation may be useful on its own because it provides a link between very different types of expressions. For completeness, explicit expressions for the coefficient $\langle j_1, 1; j_2, -1 | j, 0\rangle$, where $j_1 + j_2 + j$ is an even integer, and of $\langle j_1, 1; j_2, 1 | j, 2\rangle$ and $\langle j_1, \frac{1}{2}; j_2, -\frac{1}{2} | j, 0\rangle$, where $j_1 + j_2 + j$ is an even or an odd integer, are given.

I. INTRODUCTION

Many equivalent expressions for the general Clebsch–Gordan (CG) coefficient $\langle j_1, m_1; j_2, m_2 | j, m\rangle$ of the three-dimensional rotation group are known.[1–3] All of these involve a single summation which degenerates in special cases of types I and II, which are exemplified by $j = |j_1 \pm j_2|$ and $j = |m|$, respectively, as is clear from the Van der Waerden[4] form for these coefficients. Van der Waerden[4] shows that there are nine quantities that appear in the form of a symmetric 3×3 array (the Regge array)[5] and the coefficient is given in a closed form whenever any of the elements of this 3×3 array vanishes. In the special CG coefficients of type I, one of the angular momenta takes its extreme value (maximum or minimum), which is simultaneously an extreme value of its magnitude since the value of any angular momentum is always non-negative. However, for the CG coefficients of type II, the magnetic quantum number corresponding to one of the angular momenta involved takes its maximum value *only in magnitude*. A question naturally arises whether we can have summation-free explicit expressions for *at least* those CG coefficients where the magnetic quantum numbers of the angular momenta involved take their *minimum possible values in magnitude*. Such an expression for the coefficients $\langle j_1, 0; j_2, 0 | j, 0\rangle$ is already known.[6] This is obtained by using very different techniques than those used for obtaining the general expression. These coefficients are nonzero only when $j_1 + j_2 + j =$ an even integer, as follows easily from a known symmetry of the CG coefficients. In addition, explicit expressions for the CG coefficients $\langle j_1, \frac{1}{2}; j_2, -\frac{1}{2} | j, 0\rangle$ for the two cases when $j_1 + j_2 + j$ is an even or an odd integer are also known.[7,8] In this paper, we give such an expression for the CG coefficients $\langle j_1, 1; j_2, -1 | j, 0\rangle$, where $j_1 + j_2 + j$ is an odd integer, which are then the nonzero CG coefficients with *least value in magnitude* of the three magnetic quantum numbers. The three already-known coefficients and the fourth one given in this paper are needed as starting coefficients in a recursion scheme for the coefficients $\langle j_1, m; j_2, -m | j, 0\rangle$. For these coefficients, we arrive at explicit expressions (for different cases) involving one summation only, which is of a very different type than the known general expression using a transformation of hypergeometric functions. It seems that this transformation has not been applied in the past. Since, in the expressions for $\langle j_1, m; j_2, -m | j, 0\rangle$ obtained in this paper, the range of the single summation is restricted by m $(m - \frac{1}{2})$ for a positive integral (half-integral) value of m, for the coefficients for which m is small in magnitude, these formulas might be computationally easier to operate with even when the j's are larger.

For completeness, we also have given closed form expressions for the coefficient $\langle j_1, 1; j_2, -1 | j, 0\rangle$, where $j_1 + j_2 + j$ is an even integer and for the coefficients $\langle j_1, 1; j_2, 1 | j, 2\rangle$ and $\langle j_1, \frac{1}{2}; j_2, \frac{1}{2} | j, 1\rangle$, where $j_1 + j_2 + j$ is either an odd or an even integer. The last two cases are obtained from simple recursion relations in terms of $\langle j_1, 1; j_2, -1 | j, 0\rangle$ and $\langle j_1, \frac{1}{2}; j_2, -\frac{1}{2} | j, 0\rangle$, respectively.

 0022-2488/86/020544-05$02.50 © 1986 American Institute of Physics

The special coefficients $\langle j_1, 0; j_2, 0|j, 0\rangle$, $\langle j_1, 1; j_2, -1|j, 0\rangle$, and $\langle j_1, \frac{1}{2}; j_2, -\frac{1}{2}|j, 0\rangle$, and the ones that can be obtained from these using the various symmetries, are useful from the point of view of the theoretical discussion of the physical problem of angular correlations. Indeed, the above three coefficients occur in the study of angular correlations of spinless particles, of photons, and of conversion electrons, respectively. For unpolarized correlation measurements, we require these coefficients for $j_1 + j_2 + j =$ an even integer only. (The coefficient $\langle j_1, 0; j_2, 0|j, 0\rangle$ vanishes otherwise as follows from its symmetries, but the other two do not.) For polarized correlation measurements, we require, in addition, the latter two coefficients for $j_1 + j_2 + j =$ an odd integer and also the coefficients $\langle j_1, 1; j_2, 1|j, 2\rangle$ and $\langle j_1, \frac{1}{2}; j_2, \frac{1}{2}|j, 1\rangle$, where $j_1 + j_2 + j$ is an even or an odd integer.[9] This paper is organized as follows. In Sec. II, we present the results, whereas in Sec. III, the derivation of the results is summarized.

II. RESULTS

(a)

(i) $$\langle j_1,0;j_2,0|j,0\rangle = (-1)^{(1/2)(j_1+j_2-j)}\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j_1+j_2+j)!}{(j_1+j_2+j+1)!}}$$

$$\times\frac{(\frac{1}{2}(j_1+j_2+j))!}{(\frac{1}{2}(j_1+j_2-j))!(\frac{1}{2}(j_1-j_2+j))!(\frac{1}{2}(-j_1+j_2+j))!}, \tag{1}$$

if $j_1 + j_2 \pm j$ is an even integer, or $= 0$ if $j_1 + j_2 \pm j$ is an odd integer. In the above, all three j's are integral.[6]

(ii) $$\langle j_1,1;j_2,-1|j,0\rangle = (-1)^{(1/2)(j_1+j_2-j-1)}\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j+j_2+j)!}{(j_1+j_2+j+1)!j_1(j_1+1)j_2(j_2+1)}}$$

$$\times\frac{2(\frac{1}{2}(j_1+j_2+j+1))!}{(\frac{1}{2}(j_1+j_2-j-1))!(\frac{1}{2}(j_1-j_2+j-1))!(\frac{1}{2}(-j_1+j_2+j-1))!}. \tag{2}$$

In the above, the three j's are integral and $j_1 + j_2 \pm j$ is an odd integer. This result seems to be new.

(iii) $$\langle j_1,1;j_2,-1|j,0\rangle = \langle j_1,-1,j_2,1|j,0\rangle = \frac{1}{2}\frac{j(j+1)-j_1(j_1+1)-j_2(j_2+1)}{j_1(j_1+1)j_2(j_2+1)}\langle j_1,0;j_2,0|j,0\rangle, \tag{3}$$

where the three j's are integral, $j_1 + j_2 \pm j$ is an even integer, (see Ref. 10) and the coefficient $\langle j_1, 0; j_2, 0|j, 0\rangle$ is given in (a) (i).

(b)

(i) $$\langle j_1,\tfrac{1}{2};j_2,-\tfrac{1}{2}|j,0\rangle = (-1)^{(1/2)(j_1+j_2-j)}\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j_1+j_2+j)!}{(j_1+j_2+j+1)!(j_1+\frac{1}{2})\,(j_2+\frac{1}{2})}}$$

$$\times\frac{(\frac{1}{2}(j_1+j_2+j))!}{(\frac{1}{2}(j_1+j_2-j))!(\frac{1}{2}(j_1-j_2+j-1))!(\frac{1}{2}(-j_1+j_2+j-1))!}. \tag{4}$$

In the above, j is integral whereas j_1, j_2 are both half-integral, $j_1 + j_2 \pm j$ is an even integer, and $\pm j_1 \pm j_2 + j$ is an odd positive integer.

(ii) $$\langle j_1,\tfrac{1}{2};j_2,-\tfrac{1}{2}|j,0\rangle = (-1)^{(1/2)(j_1+j_2-j-1)}\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j_1+j_2+j)!}{(j_1+j_2+j+1)!(j_1+\frac{1}{2})\,(j_2+\frac{1}{2})}}$$

$$\times\frac{(\frac{1}{2}(j_1+j_2+j+1))!}{(\frac{1}{2}(j_1+j_2-j-1))!(\frac{1}{2}(j_1-j_2+j))!(\frac{1}{2}(-j_1+j_2+j))!}, \tag{5}$$

where j is integral whereas j_1, j_2 are both half-integral and $j_1 + j_2 \pm j$ is an odd integer but $\pm j_1 \mp j_2 + j$ is an even positive integer.

All the above results are invariant under the interchange of j_1 and j_2 as follows from the known symmetries of the CG coefficients.[7,8]

(c)

(i) $$\langle j_1,\tfrac{1}{2};j_2,\tfrac{1}{2}|j,1\rangle = [(j_1+j_2+1)/\sqrt{j(j+1)}\,]\langle j_1,\tfrac{1}{2};j_2,-\tfrac{1}{2}|j,0\rangle, \quad \text{if } j_1+j_2\pm j \text{ is an even integer,} \tag{6}$$

$$= [(j_1-j_2)/\sqrt{j(j+1)}\,]\langle j_1,\tfrac{1}{2};j_2,-\tfrac{1}{2}|j,0\rangle, \quad \text{if } j_1+j_2\pm j \text{ is an odd integer.} \tag{7}$$

(ii) $$\langle j_1,1;j_2,1|j,2\rangle = \frac{j(j+1)[j_1(j_1+1)+j_2(j_2+1)]-[(j_1-j_2)(j_1+j_2+1)]^2}{\sqrt{(j-1)j(j+1)(j+2)}[j(j+1)-j_1(j_1+1)-j_2(j_2+1)]}\langle j_1,1;j_2,-1|j,0\rangle,$$

if $j_1+j_2\pm j$ is an even integer, (8)

$$= \frac{(j_1-j_2)(j_1+j_2+1)}{\sqrt{(j-1)j(j+1)(j+2)}}\langle j_1,1;j_2,-1|j,0\rangle, \quad \text{if } j_1+j_2\pm j \text{ is an odd integer.}^{11} \tag{9}$$

III. SUMMARY OF THE DERIVATION OF THE RESULTS IN SEC. II

From the general expression for the CG coefficient, we have

$$\langle j_1,m;j_2-m|j,0\rangle = (-1)^{j_1+j_2-j}\sqrt{\frac{(2j+1)(j_1+j_2-j)!(j_1+m)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1-j_2+j)!(-j_1+j_2+j)!(j_1-m)!(j_2+m)!}}$$

$$\times j!\sum_s (-1)^s \frac{(2j_1-s)!(-j_1+j_2+j+s)!}{s!(j_1+m-s)!(-j_1+j-m+s)!(j_1+j_2-j-s)!} \tag{10}$$

$$= (-1)^{j_1+j_2-j}\sqrt{\frac{(2j+1)(-j_1+j_2+j)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+j_2-j)!(j_1-j_2+j)!(j_1+m)!(j_1-m)!(j_2+m)!}}$$

$$\times \frac{j!(2j_1)!}{\Gamma(-j_1+j_2-m+1)}\,{}_3F_2\left(\begin{matrix}-j_1+j_2+j+1, & -j_1-j_2+j, & -j_1-m\\ -2j_1, & -j_1+j-m+1\end{matrix};1\right). \tag{11}$$

Using the transformation[12]

$${}_3F_2\left(\begin{matrix}2a, & 2b, & -n\\ 2c, & \frac{1}{2}+a+b-c-n\end{matrix};1\right) = \frac{(c-a+\frac{1}{2})_n(c-b+\frac{1}{2})_n}{(c+\frac{1}{2})_n(c+\frac{1}{2}-a-b)_n}\,{}_4F_3\left(\begin{matrix}a,b, & -c-n, & -n\\ c,\frac{1}{2}+a-c-n, & \frac{1}{2}+b-c-n\end{matrix};1\right) \tag{12}$$

and the duplication formula[13]

$$\Gamma(2x) = \left[\Gamma(x)\Gamma(x+\tfrac{1}{2})/\sqrt{\pi}\right]2^{2x-1} \tag{13}$$

for the gamma functions, we arrive at

$$\langle j_1,m,j_2,-m|j,0\rangle = (-1)^{j_1+j_2-j}\sqrt{\frac{(2j+1)(-j_1+j_2+j)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+j_2-j)!(j_1-j_2+j)!(j_1+m)!(j_1-m)!(j_2+m)!}}$$

$$\times(-1)^{j_1+m}\frac{2^{2j_1}}{\sqrt{\pi}}\frac{\Gamma(j_1+1)\Gamma(1+\frac{1}{2}(j_1+j_2+j))\Gamma(\frac{1}{2}(j_1+j_2-j+1)+m)\Gamma(\frac{1}{2}-m)}{\Gamma(1+\frac{1}{2}(-j_1+j_2+j)-m)\Gamma(\frac{1}{2}(-j_1+j_2+j+1))}$$

$$\times\,{}_4F_3\left(\begin{matrix}\frac{1}{2}(1-j_1+j_2+j), & \frac{1}{2}(-j_1-j_2+j), & -j_1-m, & -m\\ -j_1, & 1+\frac{1}{2}(-j_1+j_2+j)-m, & \frac{1}{2}(1-j_1-j_2+j)-m\end{matrix};1\right). \tag{14}$$

Case I: $j_1+j_2\pm j$ *is an even integer.* We first assume that m is a non-negative integer. The hypergeometric function ${}_4F_3$ in Eq. (14) is well defined and is a terminating Saalchützian, for which we have the transformation

$${}_4F_3\left(\begin{matrix}x, & y, & z, & -n\\ u, & v, & w\end{matrix};1\right) = \frac{(v-z)_n(w-z)_n}{(v)_n(w)_n}\,{}_4F_3\left(\begin{matrix}u-x, & u-y, & z, & -n\\ u, & 1-v+z-n, & 1-w+z-n\end{matrix};1\right). \tag{15}$$

We use the above transformation for the ${}_4F_3$ in Eq. (14), taking

$$x=(1-j_1+j_2+j)/2,\quad y=-j_1-m,\quad z=(-j_1-j_2+j)/2,\quad n=m,$$

$$u=1+(-j_1+j_2+j)/2-m,\quad v=-j_1,\quad w=(1-j_1-j_2+j)/2-m,$$

to arrive at

$$\langle j_1,m;j_2,-m|j,0\rangle$$

$$= (-1)^{(1/2)(j_1+j_2-j)}\sqrt{\frac{(2j+1)(j_1+j_2-j)(j_1-j_2+j)!(-j_1+j_2+j)!(j_1-m)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+m)!(j_2+m)!}}$$

$$\times\frac{(\frac{1}{2}(j_1+j_2+j))!}{(\frac{1}{2}(j_1-j_2+j)-m)!(\frac{1}{2}(-j_1+j_2+j)-m)!(\frac{1}{2}(j_1+j_2-j))!}$$

$$\times{}_4F_3\left(\begin{matrix}1+\frac{1}{2}(j_1+j_2+j), & \frac{1}{2}(-j_1-j_2+j), & \frac{1}{2}-m, & -m\\ 1+\frac{1}{2}(-j_1+j_2+j)-m, & 1+\frac{1}{2}(j_1-j_2+j)-m, & \frac{1}{2} & \end{matrix};1\right). \tag{16}$$

The advantage of the above form is that it applies equally when m is a non-negative integer or a non-negative half-integer, although, in the above derivation, we assumed, to start with, that m is a non-negative integer. In fact, had we started with the transformation[14]

$${}_3F_2\left(\begin{matrix}2a, & 2b, & -n\\ 2c, & \frac{1}{2}+a+b-c-n & \end{matrix};1\right)$$

$$=\frac{(c-a+\frac{1}{2})_n(c-b)_n}{(c)_n(c+\frac{1}{2}-a-b)_n}\,{}_4F_3\left(\begin{matrix}a+\frac{1}{2}, & b, & \frac{1}{2}-c-n, & -n\\ c+\frac{1}{2}, & \frac{1}{2}+a-c-n, & 1+b-c-n & \end{matrix};1\right), \tag{17}$$

and followed the steps identical to the ones mentioned above, we would have gotten exactly the expression in Eq. (16) under the assumption that m is a non-negative half-integer.

Finally expanding the hypergeometric function ${}_4F_3$ in Eq. (16), we can write

$$\langle j_1,m;j_2,-m|j,0\rangle$$

$$=(-1)^{(1/2)(j_1+j_2-j)}\sqrt{\frac{(2j+1)\,(j_1-j_2+j)!(-j_1+j_2+j)!(j_1+j_2-j)!(j_1-m)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+m)!(j_2+m)!}}$$

$$\times(2m)!\sum_s(-1)^s\frac{(\frac{1}{2}(j_1+j_2+j)+s)!}{(2s)!(2m-2s)!(\frac{1}{2}(j_1+j_2-j)-s)!(\frac{1}{2}(-j_1+j_2+j)-m+s)!(\frac{1}{2}(j_1-j_2+j)-m+s)!}, \tag{18}$$

which obviously satisfies the symmetry

$$\langle j_1,m;j_2,-m|j,0\rangle=\langle j_2,m;j_1,-m|j,0\rangle, \tag{19}$$

since in the above, $j_1+j_2\pm j$ is an even integer.

Case II: $j_1+j_2\pm j$ *is an odd integer*. In this case, the ratio of the ${}_4F_3$ appearing in Eq. (14) to $\Gamma((-j_1+j_2-j+1)/2)$ is a finite quantity that shows, after a careful analysis, that (assuming that m is a non-negative integer)

$$\langle j_1,m;j_2,-m|j,0\rangle$$

$$=(-1)^{j_1+j_2-j+m}\frac{2^{2j_1}}{\sqrt{\pi}}\frac{m\Gamma(\frac{1}{2}-m)\Gamma(1+\frac{1}{2}(j_1+j_2+j))\Gamma(\frac{1}{2}(3+j_1+j_2+j))}{\Gamma(\frac{1}{2}(1-j_1+j_2+j))\Gamma(\frac{1}{2}(-j_1-j_2+j))\Gamma(\frac{1}{2}(3+j_1-j_2+j)-m)}$$

$$\times\frac{\Gamma(1+\frac{1}{2}(j_1-j_2+j))\Gamma(\frac{1}{2}(1+j_1-j_2+j))}{\Gamma(j_1+2)\Gamma(2+\frac{1}{2}(j_1+j_2+j)-m)}\sqrt{\frac{(2j+1)\,(-j_1+j_2+j)!(j_1-m)!(j_2-m)!(j_1+m)!}{(j_1+j_2+j+1)!(j_1+j_2-j)!(j_1-j_2+j)!(j_2+m)!}}$$

$$\times{}_4F_3\left(\begin{matrix}\frac{1}{2}(3+j_1+j_2+j), & 1+\frac{1}{2}(j_1-j_2+j), & j_1+1-m, & 1-m\\ j_1+2, & 2+\frac{1}{2}(j_1+j_2+j)-m, & \frac{1}{2}(3+j_1-j_2+j)-m & \end{matrix};1\right). \tag{20}$$

We now follow the same procedure as for case I to arrive at

$$\langle j_1,m;j_2,-m|j,0\rangle$$

$$=(-1)^{(1/2)(j_1+j_2-j-1)}(2m)\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j_1+j_2+j)!(j_1-m)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+m)!(j_2+m)!}}$$

$$\times\frac{(\frac{1}{2}(j_1+j_2+j+1))!}{(\frac{1}{2}(j_1-j_2+j+1)-m)!(\frac{1}{2}(-j_1+j_2+j+1)-m)!(\frac{1}{2}(j_1+j_2-j-1))!}$$

$$\times{}_4F_3\left(\begin{matrix}\frac{1}{2}(j_1+j_2+j+3), & \frac{1}{2}(-j_1-j_2+j+1), & 1-m, & \frac{1}{2}-m\\ 3+\frac{1}{2}(j_1-j_2+j)-m, & \frac{1}{2}(3-j_1+j_2+j)-m, & \frac{3}{2} & \end{matrix};1\right), \tag{21}$$

which can be expanded in the form

$$\langle j_1,m;j_2,-m|j,0\rangle$$
$$=(-1)^{(1/2)(j_1+j_2-j-1)}\sqrt{\frac{(2j+1)\,(j_1+j_2-j)!(j_1-j_2+j)!(-j_1+j_2+j)!(j_1-m)!(j_2-m)!}{(j_1+j_2+j+1)!(j_1+m)!(j_2+m)!}}$$
$$\times(2m)!\sum_s(-1)^s$$
$$\times\frac{(\frac{1}{2}(j_1+j_2+j+1)+s)!}{(2s+1)!(2m-2s-1)!(\frac{1}{2}(j_1+j_2-j-1)-s)!(\frac{1}{2}(j_1-j_2+j+1)-m+s)!(\frac{1}{2}(-j_1+j_2+j+1)-m+s)!}, \tag{22}$$

where $j_1+j_2\pm j$ is an odd integer. Note that as for case I, the expressions in Eqs. (21) and (22) apply when m is a non-negative integer or a non-negative half-integer.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste, where the first draft of the paper was prepared.

[1]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1960).
[2]A. Jucys and A. Bandzaitis, *Judijemo Kiekio Momento Teorija Kuantineje Mechanikoje* (Vilnius, Mintes, 1965).
[3]R. T. Sharp, Nucl. Phys. A **95**, 222 (1967).
[4]B. Van der Waerden, *Group Theory and Quantum Mechanics* (Springer, Berlin, 1980), p. 105, Eq. 21.9. The German edition of this book appeared in 1937. Van der Waerden's "highly symmetrical form" appears as Eq. (3.6.11) in Ref. 1, wherein Racah's derivation of it from his form is also given.
[5]T. Regge, Nuovo Cimento **10**, 544 (1959).
[6]G. Racah, Phys. Rev. **62**, 438 (1942).
[7]R. C. Young, Phys. Rev. **115**, 577 (1959).
[8]Following Ref. 7, a systemic derivation of the results for $\langle j_1,\frac{1}{2},j_2,-\frac{1}{2}|j,0\rangle$ is given in L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum mechanics," in *Encyclopaedia of Mathematics and its Applications*, edited by G -C. Rota, Vol. 8 (Addison–Wesley, Reading, MA, 1981), p. 361, Appendix h.
[9]L. C. Biedenharn and M. E. Rose, Rev. Mod. Phys. **25**, 729 (1953). This reference could be considered as a "handbook of angular correlations" as far as the theory is concerned.
[10]See Eq. (A3), p. 776 in Ref. 9.
[11]The result (c) (ii) for $j_1+j_2\pm j=$ an odd integer appears in Eq. (73j) and the special case with $j_1=j_2$ of (c) (ii) with $j_1+j_2\pm j=$ an even integer appears in Eq. (74) in Ref. 9.
[12]W. M. Bailey, *Generalised Hypergeometrics Series, Cambridge Mathematical Tract No. 32* (Cambridge U. P., Cambridge, 1964), Sec. (10.1), Eq. (2).
[13]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Eq. (8.335(i)).
[14]Section (10.1), Eq. (3) in Ref. 12.

Simple expressions for radial functions appearing in the expansions of $\mathscr{Y}_{l_1}^{m_1}(\nabla)F_{l_2}^{m_2}(\mathbf{r})$ and $\nabla^{2n}\mathscr{Y}_{l_1}^{m_1}(\nabla)F_{l_2}^{m_2}(\mathbf{r})$

M. A. Rashid
Mathematics Department, Ahmadu Bello University, Zaria, Nigeria

(Received 28 March 1985; accepted for publication 16 October 1985)

In this paper, simple analytic expressions for the radial functions $r^l_{l_1l_2}(f(r))$ and $G^{ln}_{l_1l_2}(f(r))$, which appear in the expressions for $\mathscr{Y}_{l_1}^{m_1}(\nabla)f_{l_2}(\mathbf{r})Y_{l_2}^{m_2}(\hat{\mathbf{r}})$ and $\nabla^{2n}\mathscr{Y}_{l_1}^{m_1}(\nabla)f_{l_2}(r)Y_{l_2}^{m_2}(\hat{\mathbf{r}})$ when expressed as linear combinations of $Y_l^m(\hat{\mathbf{r}})$, where $\mathscr{Y}_{l_1}^{m_1}(\nabla)$ is obtained from the solid harmonic $\mathscr{Y}_{l_1}^{m_1}(\mathbf{r}) = r^{l_1}Y_{l_1}^{m_1}(\hat{\mathbf{r}})$, are derived by replacing x, y, and z in its representation as a polynomial of degree l_1 by $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$, respectively. Writing $r^l_{l_1l_2}(f(r)) = r^l_{l_1l_20}(r)f(r)$ and $G^{ln}_{l_1l_2}(f(r)) = G^{ln}_{l_1l_20}(r)f(r)$, the expression for $r^l_{l_1l_20}(r)(G^{ln}_{l_1l_20}(r))$ is found to be $(2/r)^{l_1}((2/r)^{l_1+2n})$ times a product of commuting factors containing the single operator $\frac{1}{2}r(d/dr)$ and not two noncommuting operators r and $(1/r)(d/dr)$ in the many equivalent previously obtained results which are thus synthesized in this approach. Also, expressions for these operators manifest the symmetries in the problem. In addition, a simple connection $G^{ln}_{l_1l_20}(r) = r^l_{l_1+2nl_20}(r)$ between these two operators and the corresponding radial functions $G^{ln}_{l_1l_2}(f(r))$ and $r^l_{l_1l_2}(f(r))$ is found.

I. INTRODUCTION

Because of its usefulness in many problems of theoretical physics,[1-5] many authors have calculated[3,6,7] the effect of applying the spherical tensor gradient $\mathscr{Y}_{l_1}^{m_1}(\nabla)$ of rank l_1 on a spherical tensor $F_{l_2}^{m_2}(\mathbf{r})$ of rank l_2. Here $\mathscr{Y}_{l_1}^{m_1}(\nabla)$ is obtained by replacing x, y, and z in the polynomial of degree l_1 in x, y, and z for the solid harmonic $\mathscr{Y}_{l_1}^{m_1}(\mathbf{r}) = r^{l_1}Y_{l_1}^{m_1}(\hat{\mathbf{r}})$ by $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$, respectively. Thus $\mathscr{Y}_{l_1}^{m_1}(\nabla)$, like the solid harmonic $\mathscr{Y}_{l_1}^{m_1}(\mathbf{r})$, is a spherical tensor of rank l_1. Also

$$F_{l_2}^{m_2}(\mathbf{r}) = f_{l_2}(r)Y_{l_2}^{m_2}(\hat{\mathbf{r}}). \tag{1}$$

Indeed, from angular momentum considerations, we can write

$$\mathscr{Y}_{l_1}^{m_1}(\nabla)F_{l_2}^{m_2}(\mathbf{r}) = \sum_l \langle l_1,m_1|l_2,m_2|l,m\rangle \times r^l_{l_1l_2}(f_{l_2}(r))Y_l^{m_1+m_2}(\hat{\mathbf{r}}), \tag{2}$$

where $\langle l_1,m_1|l_2,m_2|l,m\rangle$ is the Gaunt coefficient given by

$$\langle l_1,m_1|l_2,m_2|l,m_1+m_2\rangle = \sqrt{(2l_1+1)(2l_2+1)/4\pi(2l+1)} \times \langle l_1,m_1;l_2,m_2|l,m\rangle\langle l_1,0;l_2,0|l,0\rangle \tag{3}$$

in terms of the Clebsch–Gordan coefficients as given in Eq. (2.5) in Ref. 6.

In Eq. (2), the radial function $r^l_{l_1l_2}(f(r))$ is independent of the magnetic quantum numbers m_1 and m_2 and the summation variable l takes the values l_1+l_2, $l_1+l_2-2,\ldots,|l_1-l_2|$. Thus $\frac{1}{2}(l_1+l_2-l)$ and $\frac{1}{2}(l_1-l_2+l)$ are both *non-negative integers.*

Bayman[3] derived an explicit analytic expression for $r^l_{l_1l_2}(f(r))$, wherein he factored out only the Clebsch–Gordan coefficient $\langle l_1,m_1;l_2,m_2|l,m\rangle$ in Eq. (2). Weniger and Steinborn have recently obtained an equivalent but somewhat simpler expression[8]

$$r^l_{l_1l_2}(f(r)) = \sum_{q=0}^{(l_1+l_2-l)/2} \frac{(\frac{1}{2}(l_1+l_2-l))!\Gamma(\frac{1}{2}(l_1+l_2+l+3))}{q!(\frac{1}{2}(l_1+l_2-l)-q)!\Gamma(\frac{1}{2}(l_1+l_2+l+3)-q)} 2^q r^{l_1+l_2-2q}\left(\frac{1}{r}\frac{d}{dr}\right)^{l_1-q}\frac{f(r)}{r^{l_2}} \tag{4}$$

by factoring out the Gaunt coefficient as shown in Eq. (2). The above expression for $r^l_{l_1l_2}(f(r))$ contains a single summation over q. By utilizing Fourier transform techniques, they were able to arrive at many equivalent expressions for the same radial function that do not involve any summation. These could be exemplified by[9]

$$r^l_{l_1l_2}(f(r)) = \frac{1}{r^{l+1}}\left(\frac{1}{r}\frac{d}{dr}\right)^{(l_1+l_2-l)/2} r^{l_1+l_2+l+1} \times\left(\frac{1}{r}\frac{d}{dr}\right)^{(l_1-l_2+l)/2}\frac{f_{l_2}(r)}{r^{l_2}} \tag{5}$$

$$= r^l\left(\frac{1}{r}\frac{d}{dr}\right)^{(l_1-l_2+l)/2} r^{l_1-l_2-l-1} \times\left(\frac{1}{r}\frac{d}{dr}\right)^{(l_1+l_2-l)/2} r^{l_2+1}f_{l_2}(r). \tag{6}$$

In addition, the same technique gave them similar but more complicated expressions for the radial function, which appears in the equation [10]

$$\nabla^{2n}\mathscr{Y}_{l_1}^{m_1}(\nabla)F_{l_2}^{m_2}(\mathbf{r}) = \sum_l \langle l_1,m_1|l_2,m_2|l,m\rangle \times G^{ln}_{l_1l_2}(f_{l_2}(r))Y_l^{m_1+m_2}(\hat{\mathbf{r}}). \tag{7}$$

 0022-2488/86/020549-03$02.50 © 1986 American Institute of Physics

We note that in Eqs. (4)–(6) above, $r^{l}_{l_1 l_2}(f(r))$ is expressed in terms of products of *two noncommuting* operators r and $(1/r)(d/dr)$ operating on $f(r)$. In this paper, we have expressed $r^{l}_{l_1 l_2}(f(r))$ as $(2/r)^{l_1}$ times a product of commuting factors involving the operator $\frac{1}{2}r(d/dr)$ only from Eq. (4) directly and without application of Fourier transform techniques. This operator $\frac{1}{2}r(d/dr)$ has the additional advantage that it does not change the degree of any term in $f(r)$. Thus each term in the expansion of $f(r)$ is acted upon independently. Exactly the same remarks apply to our expression for $G^{ln}_{l_1 l_2}(f(r))$. If we write

$$r^{l}_{l_1 l_2}(f(r)) = r^{l}_{l_1 l_2 0}(r) f(r), \tag{8}$$

$$G^{ln}_{l_1 l_2}(f(r)) = G^{ln}_{l_1 l_2 0}(r) f(r), \tag{9}$$

our expression for the *radial operators* $r^{l}_{l_1 l_2 0}(r)$ and $G^{ln}_{l_1 l_2 0}(r)$ manifestly exhibit their symmetries, unlike the expressions of Bayman and Weniger and Steinborn. Also, we find the relationship

$$G^{ln}_{l_1 l_2 0}(r) = r^{l_n}_{l_1 + 2n l_2 0}(r), \tag{10}$$

which is not evident in the expressions derived by Weniger and Steinborn.

This paper is organized as follows. In Sec. II, we present the derivation of our results for $r^{l}_{l_1 l_2}(f(r))$ and $G^{ln}_{l_1 l_2}(f(r))$. From our results, the symmetries of the corresponding operators $r^{l}_{l_1 l_2 0}(r)$ and $G^{ln}_{l_1 l_2 0}(r)$ are evident. In Sec. III we summarize our conclusions.

II. DERIVATION OF THE ANALYTIC EXPRESSION FOR $r^{l}_{l_1 l_2}(f(r))$ AND $G^{ln}_{l_1 l_2}(f(r))$

From

$$\frac{1}{r}\frac{d}{dr}\left(\frac{1}{r^{l_2}}f(r)\right) = \frac{2}{r^{l_2+2}}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2\right)f(r),$$

we note that the operator $(1/r)(d/dr)$ replaces $1/r^{l_2}$ by $1/r^{l_2+2}$ and $f(r)$ by $2(\frac{1}{2}r(d/dr) - \frac{1}{2}l_2)f(r)$. Repeating the operation of $(1/r)(d/dr)$ then gives

$$\left(\frac{1}{r}\frac{d}{dr}\right)^2 \frac{1}{r^{l_2}}f(r) = \frac{1}{r}\frac{d}{dr}\left[\frac{1}{r^{l_2+2}}\left\{2\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2\right)f(r)\right\}\right] = \frac{2^2}{r^{l_2+4}}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2 - 1\right)\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2\right)f(r)$$
$$= \frac{2^2}{r^{l_2+4}}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2 - 1\right)_2 f(r),$$

where we have used the Pochhammer symbol defined by

$$(a)_n = \Gamma(a+n)/\Gamma(a).$$

The above procedure results in the generalization

$$\left(\frac{1}{r}\frac{d}{dr}\right)^k \left(\frac{1}{r^{l_2}}f(r)\right) = \frac{2^k}{r^{l_2+2k}}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}l_2 - k + 1\right)_k f(r).$$

In the above, we take $k = l_1 - q$ and multiply the result by $2^q r^{l_1+l_2-2q}$ to arrive at

$$2^q r^{l_1+l_2-2q}\left(\frac{1}{r}\frac{d}{dr}\right)^{l_1-q}\frac{1}{r^{l_2}}f(r) = \left(\frac{2}{r}\right)^{l_1}\left(\frac{1}{2}r\frac{d}{dr} - l_1 - \frac{1}{2}l_2 + q + 1\right)_{l_1-q} f(r). \tag{11}$$

Thus Eq. (4) becomes

$$r^{l}_{l_1 l_2}(f(r)) = \left(\frac{2}{r}\right)^{l_1}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}(l_1+l) + 1\right)_{(l_1-l_2+l)/2} \times\left(\frac{1}{2}r\frac{d}{dr} + \frac{1}{2}(-l_1+l+3)\right)_{(l_1+l_2-l)/2} f(r). \tag{12}$$

In deriving the above result, we have performed the formal finite summation in q in Eq. (4) by using a special case of Vandermonde's summation theorem for the terminating hypergeometric function ${}_2F_1$ in the form[11]

$${}_2F_1(-n,a;c;1) = (c-a)_n/(c)_n. \tag{13}$$

Note that in Eq. (12) both the Pochhammer symbols containing $\frac{1}{2}r(d/dr)$ are well defined since $\frac{1}{2}(l_1+l_2-l)$ and $\frac{1}{2}(l_1-l_2+l)$ are integers and each symbol can be expressed as a product of factors containing the operator $\frac{1}{2}r(d/dr)$. Naturally all these factors commute and we can place them in any order, which results in many different equivalent expressions of the type given in Eqs. (5) and (6). In our single equation (12), we have synthesized all the different equivalent expressions obtained by Weniger and Steinborn[9] for $r^{l}_{l_1 l_2}(f(r))$.

To exemplify the usefulness of our result, we rederive Hobson's result by specifying[12]

$$F^{m_2}_{l_2}(\mathbf{r}) = 1/r,$$

i.e., $l_2 = m_2 = 0$, $f_{l_2}(r) = f_0(r) = 1/r$. Consequently $l = l_1$, and we obtain

$$r^{l}_{l0}(1/r) = (-1)^l (2l-1)!!/r^{l+1},$$

which finally results in

$$\mathscr{Y}^{m}_{l}(\nabla)(1/r) = (-1)^l [(2l-1)!!/r^{l+1}] Y^{m}_{l}(\hat{\mathbf{r}}). \tag{14}$$

Next we derive an analytic expression for $G^{ln}_{l_1 l_2}(f(r))$, which appears in Eq. (7) above. For this purpose, we note that

$$\nabla^2\left[\left(\frac{2}{r}\right)^{l_1+2k} f(r) Y^{m}_{l}(\hat{\mathbf{r}})\right] = \left(\frac{2}{r}\right)^{l_1+2k+2} Y^{m}_{l}(\hat{\mathbf{r}})\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}(l_1+l) - k\right) \times\left(\frac{1}{2}r\frac{d}{dr} + \frac{1}{2}(-l_1+l+1) - k\right) f(r). \tag{15}$$

We apply the above result repeatedly on both sides in Eq. (2), and utilizing Eq. (12) we finally arrive at

$$G^{ln}_{l_1 l_2}(f(r)) = \left(\frac{2}{r}\right)^{l_1+2n}\left(\frac{1}{2}r\frac{d}{dr} - \frac{1}{2}(l_1+l) - n + 1\right)_{(l_1+l-l_2)/2+n}\left(\frac{1}{2}r\frac{d}{dr} + \frac{1}{2}(-l_1+l+3) - n\right)_{(l_1+l_2-l)/2+n} f(r). \tag{16}$$

In Eqs. (8) and (9) above, we defined the radial operators $r^l_{l_1 l_2 0}(r)$ and $G^{ln}_{l_1 l_2 0}(r)$, which are just the right-hand sides of Eqs. (12) and (16) without the $f(r)$. Evidently, these operators satisfy the symmetries

$$r^l_{l_1 l_2 0}(r) = r^{-l-1}_{l_1 l_2 0}(r) = r^l_{l_1 -l_2-1 0}(r), \tag{17}$$

and the relationship

$$G^{ln}_{l_1 l_2 0}(r) = r^l_{l_1+2nl_2 0}(r), \tag{18}$$

from which the symmetries given in Eq. (17) are also evident for $G^{ln}_{l_1 l_2}(r)$.

We can now obtain expressions similar to those given in Eqs. (5) and (6) for $G^{ln}_{l_1 l_2}(f(r))$ by just replacing l_1 by $l_1 + 2n$. These expressions are to be compared with the more complicated ones given in Eqs. (4.28) and (4.29) in Weniger and Steinborn[6] and in Eqs. (3)–(6) in Niukkanen,[7] which give an expression obtained from the one given in Eq. (4) above with the replacements of $r^l_{l_1 l_2}\ f(r)$, $f(r)/r^{l_2}$, l_1, and q by $r^l_{l_1+2nl_2}(f(r))$, $f(r)$, l_1+2n, and $\frac{1}{2}(l_1+l_2-l)-q$, respectively. Note that while considering powers of $(-\frac{1}{2}r^2\mathscr{D})$ in the expansion of the Laguerre polynomial $L^{\lambda+1/2}_{\nu}(-\frac{1}{2}r^2\mathscr{D})$ in Eq. (3) of Niukkanen,[7] we assume that r^2 and $\mathscr{D} = (1/r)(d/dr)$ commute, i.e., $(-\frac{1}{2}r^2\mathscr{D})^k = (-r^2/2)^k\mathscr{D}^k$.

III. DISCUSSION

The main results of this paper are contained in Eqs. (12) and (16), which express the radial functions $r^l_{l_1 l_2}(f(r))$ and $G^{ln}_{l_1 l_2}(f(r))$ in terms of the repeated operations of factors containing $\frac{1}{2}r(d/dr)$ on $f(r)$. Since no other operator exists between these factors, the factors can be permuted at will. This results in a synthesis of many expressions derived by Weniger and Steinborn using Fourier transform techniques. Also our expressions for the radial operators $r^l_{l_1 l_2 0}(r)$ and $G^{ln}_{l_1 l_2 0}(r)$ exhibit the symmetries present in the problem, i.e., these operators are manifestly invariant under the transformations $l \leftrightarrow -l-1$, and $l_2 \leftrightarrow -l_2-1$. We have also obtained a relationship between $r^l_{l_1 l_2 0}(r)$ and $G^{ln}_{l_1 l_2 0}(r)$ that shows that the second operator can be obtained from the first by replacing l_1 by l_1+2n. We have not been able to find any physical reason for the existence of this relationship.

Since our results are expressed in terms of the operator $\frac{1}{2}r(d/dr)$, operation on any function that can be expressed as a power series in r becomes trivial—each term of the power series being operated upon independently. We have exhibited the usefulness of our approach by rederiving the classical Hobson result.

[1] F. D. Santos, Nucl. Phys. A **212**, 341 (1973).
[2] E. G. Rowe, J. Math. Phys. **19**, 1962 (1978).
[3] B. F. Bayman, J. Math. Phys. **19**, 2558 (1978).
[4] S. N. Stuart, J. Austral. Math. Soc. Ser. B **22**, 368 (1981).
[5] E. J. Weniger and E. O. Steinborn, J. Chem. Phys. **78**, 6121 (1983).
[6] E. J. Weniger and E. O. Steinborn, J. Math. Phys. **24**, 2553 (1983). This reference contains many useful references.
[7] A. W. Niukkanen, J. Math. Phys. **24**, 1989 (1983).
[8] See Ref. 6, Eq. (3.29).
[9] See Ref. 6, Eqs. (4.15)–(4.18).
[10] See Ref. 6, Eqs. (4.28)–(4.29).
[11] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, England, 1966), Eq. (III.4), p. 243.
[12] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea, New York, 1965).

Confining optical potentials

F. J. Gomez and M. F. Rañada
Departamento de Fisica Teorica, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain
J. Sesma[a)]
Max-Planck-Institut für Kernphysik, Heidelberg, Federal Republic of Germany

(Received 28 January 1985; accepted for publication 27 September 1985)

The possibility of confining a quantum particle in an optical potential, the imaginary part of which would be related to the existence of inelastic reaction channels coupled to the elastic one, is considered. Pure imaginary potentials of various (constant, linear, and parabolic) radial shapes are studied. The numerical dependence of the energies and mean-lives of the confined states is explicitly shown. Analytical expressions for such dependence are also given in some cases.

I. INTRODUCTION

Since the discovery of the J/ψ (3095) and its interpretation as a charmed quark–antiquark bound state, there has been a considerable interest, increased by the subsequent detection of more and more members of the charmonium and bottomonium families, in confinement mechanisms. Efforts to understand the nonobservability of free quarks have given rise to two kinds of models: bag and potential models (for recent reviews, see Refs. 1 and 2, respectively). In the first kind of model, confinement is imposed via boundary conditions on the surface of the bag; in the second, confinement is produced by the tail of the potential, which goes to infinity as r increases. Both confinement mechanisms are rather *ad hoc*, although they can be made plausible in terms of color flux bubbles or strings.

The physical idea behind those models of confinement is very simple: Free quarks cannot be observed, for instance, by breaking the quark–antiquark pair in a meson because new quark–antiquark pairs are created as the original quark and antiquark are pulled out. In other words, reaction channels are opened as the quark and antiquark in the elastic channel are separated. In nuclear physics, the presence of coupled reaction channels is commonly represented in the elastic one by means of an optical (complex) potential, the imaginary part of which is related to the channel coupling.

Motivated by those considerations, we forward the possibility of producing $q\bar{q}$ confinement by adding to the quantum chromodynamics (QCD)-supported Coulombian potential an imaginary part acting only at distances larger than the "radius" of the meson. Such an imaginary part would increase with the distance, so as to represent the fact that more and more inelastic channels are opened.

Before trying to fit the masses of the hadrons by assuming a complex phenomenological $q\bar{q}$ potential, it has seemed to us worthwhile to explore some specific examples in order to understand how those optical potentials, of various shapes, would work to produce confinement. Moreover, we have taken purely imaginary potentials to avoid the effect of the imaginary part of the potential being masked or distorted by that of the real part.

In the examples discussed in Secs. II–VI, spherical symmetry is assumed. To facilitate the solution of the Schrödinger equation, we have considered only constant, linear, or quadratic dependence on the radius. The information so obtained allows us to draw some conclusions in Sec. VII.

II. PARABOLIC SHAPE

Let us start by considering an analytically solvable case, namely, an imaginary potential with quadratic dependence on the radius,

$$V(r) = \tfrac{1}{2}m\omega^2 r^2, \quad \mathrm{Arg}\,\omega^2 = -\pi/2. \tag{2.1}$$

This is a very simple example, the solutions of which are the same as for the harmonic oscillator, with the only difference being that the frequency ω is now complex. The physically acceptable solutions for the l-wave radial function are written, in terms of the dimensionless variable $z \equiv r^2 m\omega/\hbar$, in the form[3]

$$\psi_{n,l}(r) = z^{l/2}e^{-z/2}\,{}_1F_1(-n;l+\tfrac{3}{2};z), \quad n,l = 0,1,2,..., \tag{2.2}$$

where ${}_1F_1$ denotes the confluent hypergeometric function.[4] The corresponding values of the energy are given by

$$E_{n,l} = (2n + l + \tfrac{3}{2})\hbar\omega. \tag{2.3}$$

The energy spectrum is, therefore, trivially related to the real harmonic oscillator one: it can be obtained from this one by a rotation of angle

$$\mathrm{Arg}\,E_{n,l} = \mathrm{Arg}\,\omega = -\pi/4 \tag{2.4}$$

in the complex-energy plane.

The potential given by Eq. (2.1) produces, therefore, an infinite number of quasibound states, of equally spaced energies, whose mean-lives are inversely proportional to their energies.

III. LINEAR SHAPE

As a second example, let us consider an imaginary potential of the form

$$V(r) = ar, \quad \mathrm{Arg}\,a = -\pi/2. \tag{3.1}$$

Real potentials with a linear dependence on the radial variable have been widely used to explain the quark–antiquark bound states, mainly due to the fact that such a dependence is supported by (nonrigorous) QCD calculations on a lattice.

[a)] On leave of absence from Universidad de Zaragoza.

 0022-2488/86/020552-07$02.50 © 1986 American Institute of Physics

The Schrödinger equation for such a radial dependence can be solved[5] exactly for angular momentum $l = 0$ and approximately for $l \neq 0$. The resulting eigenenergies $E_{n,l}$ are proportional to $a^{2/3}$, with coefficients of proportionality depending on the label n and on the angular momentum l. The treatment made for real a can be immediately extended to complex a. We can conclude in this way that

$$\text{Arg } E_{n,l} = -\pi/3, \tag{3.2}$$

which shows that the infinity of quasibound states originated by the imaginary linear potential are even shorter lived than those due to a parabolic one.

IV. STEP SHAPE

The potential to be considered now is of the form

$$V(r) = \theta(r-b)V, \quad \text{Arg } V = -\pi/2. \tag{4.1}$$

The solutions of the resulting Schrödinger equation can be expressed in terms of Bessel functions in both inner ($r<b$) and outer ($r>b$) regions. If we denote by $k \equiv (2mE)^{1/2}/\hbar$ and $k' \equiv [2m(E-V)]^{1/2}/\hbar$, respectively, the wave numbers in these two regions, then the radial wave function for angular momentum l can be written as

$$\begin{aligned} &\psi(r) \propto j_l(kr), \quad r<b, \\ &\psi(r) \propto h_l^{(1)}(k'r), \quad r>b. \end{aligned} \tag{4.2}$$

The continuity at $r = b$ of the reduced logarithmic derivative of the wave function requires

$$\alpha j_l'(\alpha)/j_l(\alpha) = \alpha' h_l^{(1)\prime}(\alpha')/h_l^{(1)}(\alpha'), \tag{4.3}$$

with

$$\alpha \equiv kb, \quad \alpha' \equiv k'b. \tag{4.4}$$

An equivalent form of writing Eq. (4.3) is

$$\alpha J_\lambda'(\alpha)/J_\lambda(\alpha) = \alpha' H_\lambda^{(1)\prime}(\alpha')/H_\lambda^{(1)}(\alpha'), \tag{4.5}$$

with

$$\lambda \equiv l + \tfrac{1}{2}. \tag{4.6}$$

Obviously, the values of α satisfying Eq. (4.5) correspond to the poles, in the complex wave number plane, of the S matrix for the potential given in Eq. (4.1). Plots of the reduced logarithmic derivatives of the cylindrical Bessel and Hankel functions[6,7] are of great help in obtaining approximate solutions of Eq. (4.5). It can be seen that an infinity of complex values of α, lying below the real axis, satisfy that equation. We are interested only in solutions corresponding to decaying states (Im $E < 0$), that is, those with Re $\alpha > 0$.

Approximate solutions of Eq. (4.5) can be obtained analytically in the limiting cases of extremely weak or very strong potential. In the first case, $|V| \ll \hbar^2/2mb^2$, it is evident that the values of α satisfying Eq. (4.5) must go to infinity as $V \to 0$, in view of the Wronskian relation for the Bessel and Hankel functions. Therefore, asymptotic expansions[8] for these functions and their derivatives can be used to obtain

$$\tan\{\alpha - (l+1)\pi/2\} \simeq -i(1 - mVb^2/\hbar^2\alpha^2), \tag{4.7}$$

from which it follows that

$$\text{Re } \alpha \simeq n\pi + (l+1)\pi/2 - \tfrac{1}{2}\text{Arg}(mVb^2/2\hbar^2\alpha^2), \tag{4.8a}$$

$$\text{Im } \alpha \simeq \tfrac{1}{2}\log|mVb^2/2\hbar^2\alpha^2|. \tag{4.8b}$$

These equations show that, as $V \to 0$, the solutions go to infinity in the lower half of the α plane along the asymptotes

$$\text{Re } \alpha = (n + l/2 + \tfrac{1}{4})\pi. \tag{4.9}$$

In the case of very strong potential, $|V| \gg \hbar^2/2mb^2$, the right-hand side of Eq. (4.5) can be expanded[7] in the form

$$\alpha' \frac{H_\lambda^{(1)\prime}(\alpha')}{H_\lambda^{(1)}(\alpha')} = i\alpha' - \frac{1}{2} - \frac{i(4\lambda^2-1)}{8\alpha'} + O(\alpha'^{-2}), \tag{4.10}$$

which clearly shows that it tends to infinity as $|\alpha'| \to \infty$. Therefore, the solutions of Eq. (4.5) tend to that of

$$J_\lambda(\alpha) = 0 \tag{4.11}$$

as $|V| \to \infty$. That is, the imaginary potential acts in the limit as a rigid box. Let us denote by α_∞ the solutions of Eq. (4.11). It is not difficult to obtain, in the case of strong potential, an approximate expression for the difference

$$\delta \equiv \alpha - \alpha_\infty \tag{4.12}$$

between a solution of Eq. (4.5) and its limit position for $|V| \to \infty$. The left-hand side of Eq. (4.5) can be written in the form

$$\alpha J_\lambda'(\alpha)/J_\lambda(\alpha) \simeq \alpha(1 - \delta/\alpha_\infty)/\delta(1 - \delta/2\alpha_\infty) \tag{4.13}$$

by introducing Taylor expansions for J_λ and J_λ' around the point α_∞ and making use of the Bessel differential equation and of Eq. (4.11). If only lowest-order terms are retained in the right-hand sides of Eqs. (4.10) and (4.13), it turns out that

$$\delta \simeq -\alpha_\infty \hbar/b(2mV)^{1/2}. \tag{4.14}$$

Of course, more accurate expressions of δ can be obtained by retaining higher-order terms.

We have solved Eq. (4.5) numerically for intermediate values of the intensity of the potential. The results, for angular momenta 0 and 1, are shown in Figs. 1 and 2, respectively. For the reduced logarithmic derivative of the Bessel function we have used its continued fraction expansion[8]

$$\alpha J_\lambda'(\alpha)/J_\lambda(\alpha) = \lambda - \alpha J_{\lambda+1}(\alpha)/J_\lambda(\alpha)$$

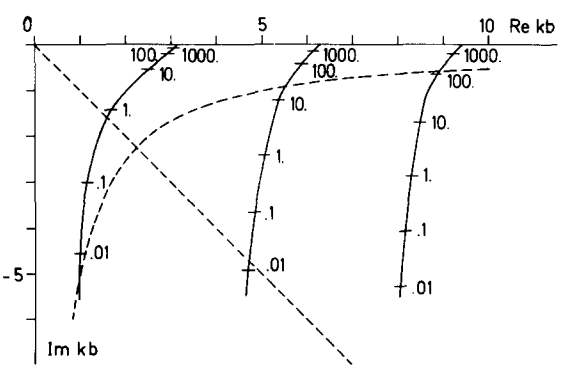


FIG. 1. Trajectories followed by the S-wave eigenvalues of the Schrödinger equation for a pure imaginary step-shaped potential as the intensity of the potential varies. Only the three lowest of an infinity of eigenvalues are shown. The numbers beside the trajectories indicate the intensity of the potential in units $i\hbar^2/2mb^2$. Eigenvalues above the bisector (dashed straight line) could be interpreted as resonances. The dashed hyperbola corresponds to an intensity of the potential equal to 10. For this intensity only one eigenvalue lies above the hyperbola and corresponds, therefore, to a confined quasistable state.

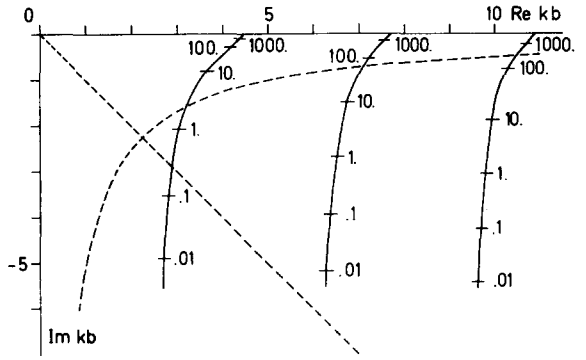


FIG. 2. Trajectories followed by the P-wave eigenvalues of the Schrödinger equation for a pure imaginary step-shaped potential as the intensity of the potential varies. All comments in caption for Fig. 1 apply also here.

$$= \lambda + \mathop{\mathbf{K}}_{n=1}^{\infty}\left(\frac{-\alpha^2}{2(\lambda + n)}\right). \tag{4.15}$$

The reduced logarithmic derivative of the Hankel function, in the case under consideration of physical values of the angular momentum, can be expressed as a quotient of polynomials

$$\alpha' H^{(1)}_{l+1/2}(\alpha')/H^{(1)}_{l+1/2}(\alpha') = -\frac{1}{2} + i\alpha' - \frac{\Sigma_{p=0}^{l-1}(l-p)(l+\tfrac{1}{2},l-p)(-2i\alpha')^p}{\Sigma_{p=0}^{l}(l+\tfrac{1}{2},l-p)(-2i\alpha')^p}, \tag{4.16}$$

with the usual notation[9]

$$(l+\tfrac{1}{2},n) \equiv (l+n)!/n!(l-n)!.$$

Those solutions located below the bisector of the fourth quadrant of the α plane do not admit a clear physical interpretation, since the real part of the corresponding energy lies below the threshold, Re $E<0$. Solutions above that bisector are commonly[10] interpreted as resonances. Some of these solutions, namely those above the hyperbola $\mathrm{Re}(kb)\times\mathrm{Im}(kb) = -mb^2|V|/\hbar^2$, do in fact correspond to confined states. Actually, for these solutions it becomes Im $k'>0$ and the external wave function, $h_l^{(1)}(k'r)$, decays exponentially with r.

Let us now prove that only a finite number of solutions can represent confined states. For large values of $|\alpha|$, the approximate version, Eq. (4.7), of Eq. (4.5) is applicable. We then immediately obtain

$$\mathrm{Re}\,\alpha \simeq n\pi + (l+1)\pi/2 + \tfrac{1}{2}\,\mathrm{Arg}(2\hbar^2\alpha^2/mVb^2 - 1), \tag{4.17a}$$

$$\mathrm{Im}\,\alpha \simeq -\tfrac{1}{2}\log|2\hbar^2\alpha^2/mVb^2 - 1|. \tag{4.17b}$$

These relations show that for the intensity of the potential being fixed, solutions corresponding to increasing values of the integer n are successively more distant from the real axis and, for n larger than a certain integer, they must lie below the hyperbola corresponding to V. Obviously, increasing the intensity of the potential makes the solutions approach the real axis and the hyperbola move downwards, the two effects contributing to increasing the number of confined states. But this number remains, nevertheless, finite except in the limit $|V|\to\infty$.

V. PARABOLIC TAIL

Let us now consider a potential given by

$$V(r) = \theta(r-b)\tfrac{1}{2}m\omega^2(r^2-b^2), \quad \mathrm{Arg}\,\omega^2 = -\pi/2, \tag{5.1}$$

that is, an imaginary potential quadratic in the distance and effective only outside a spherical surface of radius b. The radial wave function, for angular momentum l in the inner region is given by

$$\psi(r) \propto j_l(kr), \quad r<b, \tag{5.2a}$$

k being the wave number $k \equiv (2mE)^{1/2}/\hbar$, whereas in the outer region it can be written in terms of the Whittaker function,[4]

$$\psi(r) \propto \rho^{-3/4} W_{\kappa,\mu}(\rho), \quad r>b, \tag{5.2b}$$

with the notation

$$\rho \equiv r^2 m\omega/\hbar, \qquad \kappa \equiv (k^2\hbar/m\omega + b^2 m\omega/\hbar)/4, \qquad \mu \equiv (l+\tfrac{1}{2})/2. \tag{5.3}$$

The continuity of the logarithmic derivative of the wave function at $r=b$ requires

$$\alpha\frac{J'_{l+1/2}(\alpha)}{J_{l+1/2}(\alpha)} = 2z\frac{W'_{\kappa,\mu}(z)}{W_{\kappa,\mu}(z)} - 1, \tag{5.4}$$

with

$$\alpha \equiv kb, \quad z \equiv b^2 m\omega/\hbar. \tag{5.5}$$

Equation (5.4) has an infinity of solutions, some of which have been obtained numerically, in the cases of angular momentum 0 and 1, and are shown in Figs. 3 and 4, respectively, for varying intensity $|\omega^2|$ of the potential. The reduced logarithmic derivative of the Bessel function has been evaluated by means of Eq. (4.15). For the Whittaker function we have used two different algorithms, according to the magnitude of the variable z. The reduced logarithmic derivative, written as

$$z\frac{W'_{\kappa,\mu}(z)}{W_{\kappa,\mu}(z)} = \frac{z}{2} - \kappa - \frac{W_{\kappa+1,\mu}(z)}{W_{\kappa,\mu}(z)}, \tag{5.6}$$

can be expressed in terms of Kummer's functions, whose series expansions converge rapidly for small $|z|$. In the case of large $|z|$ we have preferred to use the asymptotic expansion[4,11]

$$W_{\kappa,\mu}(z) = e^{-z/2}z^{\kappa}\,{}_2F_0(\tfrac{1}{2}+\mu-\kappa,\tfrac{1}{2}-\mu-\kappa;;-1/z), \tag{5.7}$$

so as to obtain

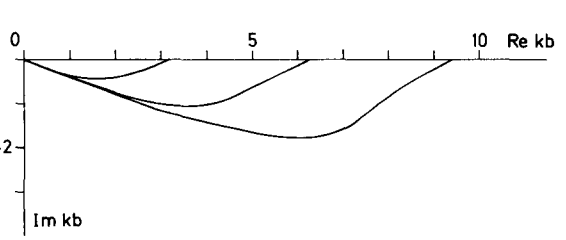


FIG. 3. S-wave eigenvalues of the Schrödinger equation for a pure imaginary parabolic tail. The first three of the infinite set of trajectories followed by the eigenvalues, as the intensity of the tail varies, are shown. All eigenvalues correspond to confined quasistable states.

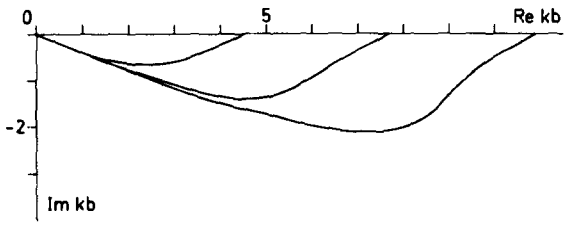


FIG. 4. P-wave eigenvalues of the Schrödinger equation for a pure imaginary parabolic tail. The comments in caption to Fig. 3 extend to this figure.

$$\frac{W_{\kappa+1,\mu}(z)}{W_{\kappa,\mu}(z)} = z\frac{{}_2F_0(-\tfrac{1}{2}+\mu-\kappa,-\tfrac{1}{2}-\mu-\kappa;;-1/z)}{{}_2F_0(-\tfrac{1}{2}+\mu-\kappa,\tfrac{1}{2}-\mu-\kappa;;-1/z)} \times\frac{{}_2F_0(\tfrac{1}{2}-\mu-\kappa,-\tfrac{1}{2}+\mu-\kappa;;-1/z)}{{}_2F_0(\tfrac{1}{2}-\mu-\kappa,\tfrac{1}{2}+\mu-\kappa;;-1/z)}. \tag{5.8}$$

The quotients of confluent hypergeometric functions can then be replaced by their continued fraction expansions[12]

$$\frac{{}_2F_0(a,b;;-1/z)}{{}_2F_0(a,b+1;;-1/z)} = 1+\mathop{\mathbf{K}}_{n=1}^{\infty}\left(\frac{a_n/z}{1}\right), \tag{5.9}$$

with

$$a_{2m-1}=a+m-1, \quad a_{2m}=b+m, \quad m \text{ an integer.} \tag{5.10}$$

Analytic approximate solutions of Eq. (5.4) can be obtained in the limit of weak potential, $|\omega|\to 0$. It can be checked that Eq. (5.4) is satisfied, for $|\omega|\to 0$, if $\alpha\to 0$ in such a way that $\kappa\to n+\mu+\frac{1}{2}$, n being a non-negative integer. In that case,

$$\lim_{\alpha\to 0} \alpha J'_{l+1/2}(\alpha)/J_{l+1/2}(\alpha) = l+\tfrac{1}{2}, \tag{5.11}$$

$$\lim_{|\omega|\to 0} W_{\kappa+1,\mu}(z)/W_{\kappa,\mu}(z) = -\mu-\kappa-\tfrac{1}{2}. \tag{5.12}$$

Going to the next order of approximation we obtain

$$\alpha J'_{l+1/2}(\alpha)/J_{l+1/2}(\alpha) \simeq l+\tfrac{1}{2}-\alpha^2/2(l+\tfrac{3}{2}), \tag{5.13}$$

$$W_{\kappa+1,\mu}(z)/W_{\kappa,\mu}(z) \simeq (-\mu-\kappa-\tfrac{1}{2})\left(1-\frac{z}{1+2\mu}-2\mu(-1)^n n!\,\frac{\Gamma(2\mu)}{\Gamma(-2\mu)}\,\Gamma\left(-\mu-\kappa-\frac{1}{2}\right)\frac{\epsilon}{z^{2\mu}}\right), \tag{5.14}$$

where ϵ is defined by

$$\kappa = n+\mu+\tfrac{1}{2}+\epsilon. \tag{5.15}$$

Equation (5.4) is fulfilled if $\epsilon = O(|\omega|^{2\mu+1})$, the eigenvalues of the energy being then given by

$$E_{n,l} \simeq (2n+l+\tfrac{3}{2})\hbar\omega - m\omega^2 b^2/2, \quad n=0,1,2,..., \tag{5.16}$$

which shows that the solutions tend to those of the parabolic potential, discussed in Sec. II, as $|\omega|\to 0$.

All solutions of Eq. (5.4) correspond to confined states. This is evident from Eq. (5.2b) and the asymptotic expansion Eq. (5.7). At large distances the radial wave function turns out to be

$$\psi(r) \underset{r\to\infty}{\sim} e^{-\rho/2}\rho^{\kappa-3/4}\{1+O(\rho^{-1})\} \tag{5.17}$$

and, as far as $\mathrm{Re}\,\omega>0$, it decreases exponentially as r increases. As shown in Eq. (5.16), those confined states are very "broad" for small $|\omega|$, since $\mathrm{Arg}\,E \simeq \mathrm{Arg}\,\omega = -\pi/4$. It can be seen in Figs. 3 and 4 that the real part of the energy of the confined states increases, for increasing $|\omega|$, whereas the relative width $\mathrm{Im}\,E/\mathrm{Re}\,E$ decreases, going to zero as $|\omega|\to\infty$. In the limit $|\omega|\to\infty$, the confined states have the (real) energies corresponding to a rigid box of radius b.

We have shown in Figs. 5 and 6 the radial wave functions of the three lowest confined states with angular momentum $l=0$ and $l=1$, respectively, for an intensity of the potential $\omega^2 = 2i$ in units $\hbar^2 m^{-2} b^{-4}$. The approximate energies of these states are, for $l=0$,

$$E_{0,0} = 1.2998 - 0.7261\,i,$$

$$E_{1,0} = 3.4082 - 2.7357\,i,$$

$$E_{2,0} = 5.4904 - 4.6606\,i,$$

and, for $l=1$,

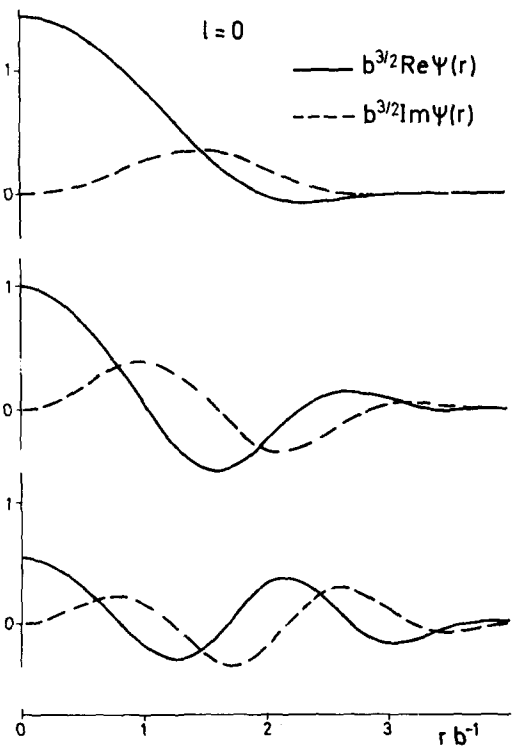


FIG. 5. Radial wave functions of the three lowest states of angular momentum $l=0$ confined by an imaginary parabolic tail. Each wave function has been properly normalized and its phase arbitrarily chosen so as to make the wave function real at the origin.

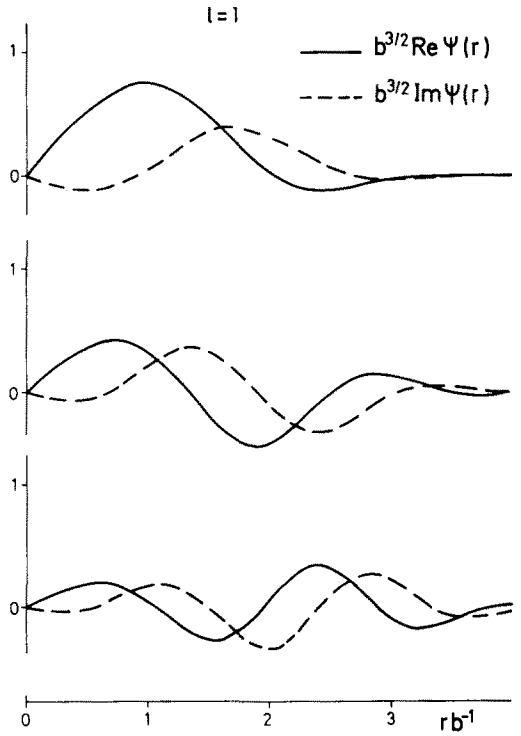


FIG. 6. Normalized radial wave functions of the three lowest states of angular momentum $l = 1$ confined by an imaginary parabolic tail. The arbitrary phase has been chosen to be that of $j_l(kr)$ for $r < b$.

$$E_{0,1} = 2.3897 - 1.5238\,i,$$
$$E_{1,1} = 4.3604 - 3.5976\,i,$$
$$E_{2,1} = 6.3985 - 5.6499\,i,$$

in units $\hbar^2 m^{-1} b^{-2}$. The radial density of probability for the same states is shown in Figs. 7 and 8.

VI. LINEAR TAIL

As a last example, let us consider an imaginary potential

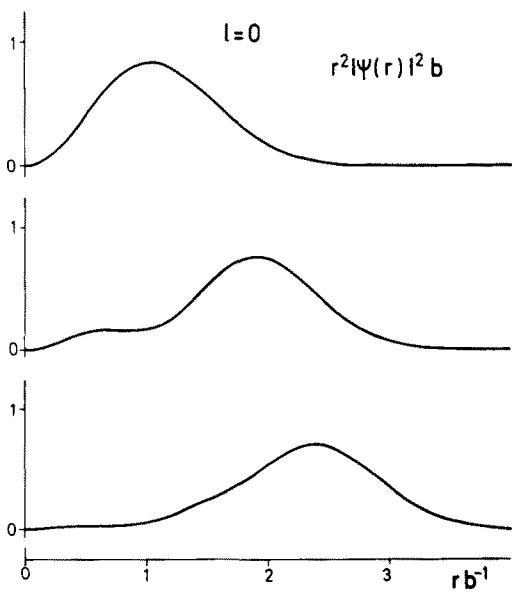


FIG. 7. Radial density of probability for the three lowest S-wave confined states in a parabolic tail.

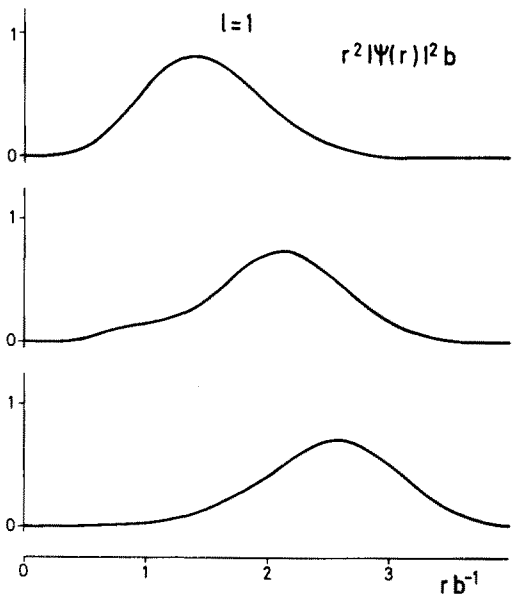


FIG. 8. Radial density of probability for the three lowest P-wave confined states in a parabolic tail.

$$V(r) = \theta(r-b)(\hbar^2/2m)ab^{-3}(r-b), \quad \text{Arg}\, a = -\pi/2, \tag{6.1}$$

the intensity of which increases linearly with the distance, starting from a given radius b. Once again, the inner radial wave function for angular momentum l is given by

$$\psi(r) \propto j_l(kr), \quad r < b, \tag{6.2}$$

with $k \equiv (2mE)^{1/2}/\hbar$. In the outer region the radial wave equation

$$\frac{d^2\psi(r)}{dr^2} + \frac{2}{r}\frac{d\psi(r)}{dr} + \left(k^2 - \frac{l(l+1)}{r^2} - ab^{-3}r + ab^{-2}\right)\psi(r) = 0, \tag{6.3}$$

when written in terms of the dimensionless variable $x \equiv r/b$, becomes

$$\frac{d^2\phi(x)}{dx^2} + \frac{2}{x}\frac{d\phi(x)}{dx} + \left(k^2b^2 - \frac{l(l+1)}{x^2} - ax + a\right)\phi(x) = 0, \tag{6.4}$$

with $\phi(x) = \psi(r)$.

In the case of zero angular momentum, the solution of Eq. (6.4) behaving regularly at infinity can be expressed in terms of the Airy function.[9] In fact,

$$\phi(x) = x^{-1}\,\text{Ai}\{a^{1/3}(x - 1 - k^2b^2/a)\}, \tag{6.5}$$

as has already been reported by several authors.[5,13] The continuity of the reduced logarithmic derivative of the wave function at $r = b$ would then require the fulfillment of the equation

$$kb\frac{j_0'(kb)}{j_0(kb)} = a^{1/3}\frac{\text{Ai}'(-k^2b^2/a^{2/3})}{\text{Ai}(-k^2b^2/a^{2/3})} - 1, \tag{6.6}$$

where the primes indicate derivatives with respect to the corresponding variable.

In the case of $l \neq 0$ the solutions of Eq. (6.4) have to be

obtained numerically. Of course, the solution regular at infinity is to be selected. Approximate expressions of such a solution, valid in the vicinity of the irregular singular point at infinity, can be obtained by standard methods.[14] A useful asymptotic expansion turns out to be

$$\phi(x)\sim x^{-5/4}\exp\{\beta_1 x^{3/2}+\beta_2 x^{1/2}+\beta_3 x^{-1/2} +\beta_4 x^{-1}+\beta_5 x^{-3/2}\}[1+O(x^{-2})], \tag{6.7}$$

with

$$\beta_1 = -2a^{1/2}/3, \tag{6.8a}$$

$$\beta_2 = (k^2b^2+a)/a^{1/2}, \tag{6.8b}$$

$$\beta_3 = -(k^2b^2+a)^2/4a^{3/2}, \tag{6.8c}$$

$$\beta_4 = (k^2b^2+a)/4a, \tag{6.8d}$$

$$\beta_5 = [l(l+1)-\tfrac{5}{16}-(k^2b^2+a)^3/8a^2]/3a^{1/2}, \tag{6.8e}$$

which can be checked by substitution in Eq. (6.4). Alternatively, we can try to relate the solutions of this equation to those of the Bessel differential equation in the region of large values of x. Changes of both variable and function,

$$\rho = a^{1/3}(x-1-k^2b^2/a), \quad \xi(\rho) = x\phi(x), \tag{6.9}$$

in Eq. (6.4) allow us to obtain

$$\frac{d^2\xi(\rho)}{d\rho^2}-\left(\rho+\frac{l(l+1)}{\rho^2}\left(\frac{\rho}{a^{1/3}x}\right)^2\right)\xi(\rho)=0. \tag{6.10}$$

New changes of variable and function,

$$t = e^{i\pi/2}\tfrac{2}{3}\rho^{3/2}, \quad w(t) = \rho^{-1/2}\xi(\rho), \tag{6.11}$$

give finally

$$t\frac{d^2w(t)}{dt^2}+t\frac{dw(t)}{dt} +\left(t^2-\frac{1}{9}\left[1+4l(l+1)\left(\frac{\rho}{a^{1/3}x}\right)^2\right]\right)w(t)=0, \tag{6.12}$$

which can be written in the form

$$t\frac{d^2w(t)}{dt^2}+t\frac{dw(t)}{dt}+\left(t^2-\left(\frac{2l+1}{3}\right)^2 +\frac{4l(l+1)}{9}\cdot\frac{1+k^2b^2/a}{x} \times\left(2-\frac{1+k^2b^2/a}{x}\right)\right)w(t)=0. \tag{6.13}$$

Asymptotically, in the region where the term $|(1+k^2b^2/a)/x|$ is much smaller than 1, this equation tends to the Bessel one.[8] Obviously, the solution we are looking for tends asymptotically to the Hankel function

$$w(t)\sim H^{(1)}_{(2l+1)/3}(t), \tag{6.14}$$

provided the ambiguity in the definition of t, as given by Eq. (6.11), has been removed by selecting the value for which $\mathrm{Im}\, t>0$. For $l=0$ there is no need to neglect terms in Eq. (6.13) and the relation Eq. (6.14) is exact: we reencounter in this way the solution given by Eq. (6.5).

By imposing continuity at $r=b$ of the logarithmic derivative of the wave function, we obtain the equation

$$kbj_l'(kb)/j_l(kb) = \phi'(1)/\phi(1), \tag{6.15}$$

which determines the energies of the confined states. As expected, an infinity of such states results for each angular momentum. In Figs. 9 and 10 we have represented the first values of kb satisfying Eq. (6.15) for $l=0$ and $l=1$, respectively. In the case of $l=0$, the right-hand side of Eq. (6.15) becomes that of Eq. (6.6) and can be easily calculated by using ascending series or asymptotic expansions for the Airy functions.[9] For $l=1$ we have evaluated the logarithmic derivative of $\phi(x)$ at $x=1$ by means of numerical integration of Eq. (6.4) starting at large values of x, for which the expansion Eq. (6.7) is applicable.

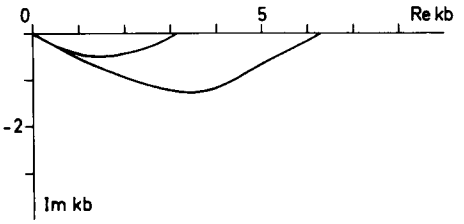


FIG. 9. Quasistable states of angular momentum $l=0$ confined by a pure imaginary linear tail. The first two trajectories of the eigenvalues of the Schrödinger equation, for varying intensity of the tail, are shown.

Approximate solutions of Eq. (6.6), corresponding to $l=0$, can be obtained in the cases of weak ($|a|\to 0$) or very strong ($|a|\to\infty$) potentials. In the first case we may assume that, as suggested by the numerical treatment, the quotient $k^2b^2/a^{2/3}$ tends to a constant as $|a|\to 0$. This implies that $kb\to 0$ and, therefore, the left-hand side of Eq. (6.6) can be approximated by $-k^2b^2/3$. It is now clear that the assumption of $k^2b^2/a^{2/3}$ tending to a constant is correct if the right-hand side of Eq. (6.6) vanishes for such constant value. Obviously, there are an infinite number of solutions corresponding to

$$k^2b^2/a^{2/3}\to -a_n, \quad \text{as } |a|\to 0, \tag{6.16}$$

where we have denoted by a_n the zeros of the Airy function.[9] By retaining the dominant terms in the two sides of Eq. (6.6), we obtain immediately

$$k^2b^2\simeq -a_n a^{2/3}-a, \quad |a|\ll 1, \tag{6.17}$$

for the energies, in units $\hbar^2/2mb^2$, of the S-wave confined states. In the case of very strong potentials, it is easy to check that the solutions kb of Eq. (6.6) tend to the zeros of $j_0(kb)$, as $|a|\to\infty$. Following a procedure similar to that used in Sec. V, we obtain

$$kb\simeq k_\infty b\,[1+a^{-1/3}\,\mathrm{Ai}(0)/\mathrm{Ai}'(0)], \tag{6.18}$$

where $k_\infty b$ takes values such that $j_0(k_\infty b)=0$, i.e.,

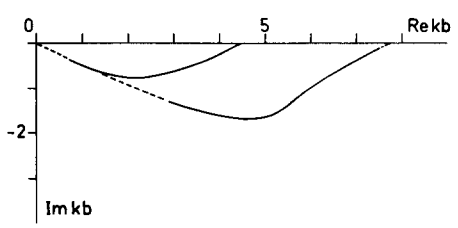


FIG. 10. Quasistable states of angular momentum $l=1$ confined by a pure imaginary linear tail. The first two trajectories of the eigenvalues of the Schrödinger equation, for varying intensity of the tail, are shown. The dashed portions of the trajectories are conjectured, as the numerical procedure for obtaining the eigenvalues gets us into trouble for extremely large or very small values of the intensity parameter.

$$k_\infty b = n\pi, \quad n = 1,2,3,\ldots . \tag{6.19}$$

Numerical analysis suggests that for $l \neq 0$, similar to what happens for $l = 0$, the solutions kb of Eq. (6.15) tend to the origin along the line Arg $kb = -\pi/6$ as $|a| \to 0$ and to the zeros of $j_l(kb)$ as $|a| \to \infty$. However, in the absence of an analytical expression for the wave function $\phi(x)$ when $l \neq 0$, we have not found explicit expressions giving the behavior of the solutions kb in the two limits.

VII. CONCLUSIONS

It is clear from the preceding examples that purely imaginary potentials are able to produce confined states. The number of these is infinite when the imaginary part of the potential goes to infinity, and finite when it tends to a constant as the distance increases. The resulting confined states are quasistable and, as it was to be expected, they turn out longer lived when the potential presents a real core besides the imaginary tail.

In what concerns the radial dependence of the potential, the results obtained in the different examples suggest that a steeper increase of the tail with the distance implies more stable confined states. As a general feature, a rapidly increasing imaginary tail has an effect similar to that of an impenetrable wall.

The purely imaginary potentials considered in this paper are, obviously, unrealistic since any potential aiming to represent the quark–antiquark interaction should contain a nonvanishing real part. The addition of an attractive real term to the potentials considered above would lead to more observable confined states, as their reduced width would become smaller. But, importantly enough, the real part of the potential does not need to be of long range, in order to produce confinement, provided an imaginary tail is present.

As a by-product of our study, we have obtained, in Sec. VI, asymptotic expansions for the wave function in a linear potential. Although in our discussion the intensity parameter a was assumed to be pure imaginary, the quoted expressions are equally valid for complex a and, in particular, for real a.

ACKNOWLEDGMENTS

The authors acknowledge the financial aid of Diputacion General de Aragon and Comision Asesora de Investigacion Cientifica y Tecnica. One of us (J.S.) is very grateful to Prof. H. A. Weidenmüller for hospitality and to the Max-Planck-Society for support during his visit in Heidelberg.

[1] A. W. Thomas, "Chiral symmetry and the bag model: A new starting point for nuclear physics," in *Advances in Nuclear Physics*, Vol. 13, edited by V. Negele and E. W. Vogt (Plenum, New York, 1983), pp. 1–137; and A. W. Thomas, "Towards a description of nuclear physics at the quark level," CERN preprint TH.3668, 1983, presented at the International School of Nuclear Physics, 7th course, Mesons, Isobars, Quarks and Nuclear Excitations, Erice, 1983.

[2] A. Martin, J. Phys. Paris **43**, C 3 (1982), presented to the XXI International Conference on High Energy Physics, Paris, 1982; C. Quigg, "Topics in Quarkonium Physics," FNAL preprint FERMILAB-Pub-83/25-THY, 1983, presented at the XXIInd Kraków School of Theoretical Physics, Zakopane, 1982; J. Rosner, "Progress in the description of heavy quark systems," Enrico Fermi Institute preprint EFI-83-17, 1983; E. Eichten, "Quarkonium: From the Bottom Up," FNAL preprint FERMILAB-Conf-83/101-THY (1983), presented at the 1983 SLAC Summer Institute, Stanford, CA, 1983, to appear in the *Proceedings of the SLAC Summer Institute on Particle Physics*, edited by P. McDonough; and K. Gottfried, "Hadronic Spectroscopy" in *Proceedings of the International Europhysics Conference on High Energy Physics*, edited by J. Guy and C. Costain (Rutherford Appleton Lab., Chilton, England, 1983), pp. 743–767.

[3] A. S. Davydov, *Quantum Mechanics* (Pergamon, Oxford, 1965), p. 131.

[4] L. J. Slater, "Confluent hypergeometric functions," in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1965), pp. 503–535.

[5] S. K. Bose, A. Jabs, and H. J. W. Müller-Kirsten, Phys. Rev. D **13**, 1489 (1976). See also references therein.

[6] A. Cruz and J. Sesma, Math. Comp. **35**, 1317 (1980).

[7] A. Cruz and J. Sesma, Math. Comp. **41**, 597 (1983).

[8] F. W. J. Olver, "Bessel functions of integer order," in Ref. 4, pp. 355–433.

[9] H. A. Antosiewicz, "Bessel functions of fractional order", in Ref. 4, pp. 435–478.

[10] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965); R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1966); H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972); A. G. Sitenko, *Lectures in Scattering Theory* (Pergamon, Oxford, 1971); and J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).

[11] H. Buchholz, *The Confluent Hypergeometric Function* (Springer, Berlin, 1969), p. 90.

[12] W. B. Jones and W. J. Thron, *Continued Fractions: Analytic Theory and Applications* (Addison–Wesley, Reading, MA, 1980), p. 212.

[13] S. Flügge, *Practical Quantum Mechanics* (Springer, Berlin, 1974), p. 101; B. J. Harrington, S. Y. Park, and A. Yildiz, Phys. Rev. Lett. **34**, 369 (1975); J. F. Gunion and R. S. Willey, Phys. Rev. D **12**, 174 (1975); and A. F. Antippa and A. J. Phares, J. Math. Phys. **19**, 308 (1978).

[14] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), part 1, p. 551; and F. W. J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974), p. 229.

Penrose diagrams for the Einstein, Eddington–Lemaitre, Eddington–Lemaitre–Bondi, and anti-de Sitter universes

Frank J. Tipler
Department of Mathematics and Department of Physics, Tulane University, New Orleans, Louisiana 70118

(Received 22 March 1985; accepted for publication 20 September 1985)

Penrose diagrams including the c boundary at infinity and the singularities for the Einstein, Eddington–Lemaitre, Eddington–Lemaitre–Bondi, and anti-de Sitter universes are constructed. Penrose diagrams for the Einstein, Eddington–Lemaitre, and anti-de Sitter universes have been published before, but these diagrams are incomplete in that the published diagrams do not contain the c-boundary points of the universes they are supposed to represent.

I. INTRODUCTION

The most widely used method of representing the global structure of a space-time is the Penrose diagram. Although the technique for constructing such diagrams was introduced by Penrose over twenty years ago,[1] no new diagrams of standard cosmologies have been given since Penrose's original paper,[1] as can be seen by comparing the examples in his original paper with the list he gave in his most recent paper, which discusses the global structure of exact solutions.[2] In this most recent list are found 11 diagrams, and of the nine figures representing cosmological solutions, three are incomplete. The Penrose diagrams for the Einstein static universe, the Eddington–Lemaitre universe, and the anti-de Sitter universe are incomplete, since single c-boundary points are not brought into a finite distance by a conformal transformation and included in the diagrams as boundary points of the conformal space-time, but rather are left dangling as disjoint points "at infinity." (This is also the representation of the global structure of these space-times in Penrose's Adams Prize essay,[3] and in Hawking and Ellis.[4]) Penrose actually included the single ideal points of the Einstein universe in a rough diagram in his first paper on the conformal structure of infinity, but this rough diagram (Fig. 9 of Ref. 1) incorrectly represents the true conformal structure of the Einstein universe, as I shall show below.

I shall show how to include the isolated ideal points in the Penrose diagrams for the Einstein, Eddington–Lemaitre, and anti-de Sitter space times. I shall also construct the Penrose diagram for a spatially homogeneous and isotropic cosmology which I term the "Eddington–Lemaitre–Bondi universe": it is a $k = +1$ Friedmann universe with positive cosmological constant that starts at a singularity and asymptotically approaches the Einstein static universe (see Bondi,[5] p. 84 for a discussion of this cosmology). This cosmological model is interesting because its causal structure is identical to a closed universe that begins in a Friedmann singularity, but ends in a Mixmaster singularity, in which the horizons disappear.[6–8] If the actual universe is closed, it is more likely to be represented by a model with a chaotic final singularity, rather than the extremely regular Friedmann final singularity, as Penrose has pointed out.[9,10] It is possible that such a singularity has no horizons, though this point is not established (see Ref. 11 for a discussion). If such is indeed the actual universe's global causal structure, it would be useful to have a simple Penrose diagram which would illustrate it.

II. CALCULATING THE DIAGRAMS

In standard coordinates, the Einstein static universe can be written as

$$ds'^2 = -dt'^2 + R_0^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{1}$$

where R_0 is a constant, $-\infty < t' < +\infty$, $0 \leq \chi \leq \pi$, $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$. We can most simply bring temporal infinity in to a finite distance via the transformation $t' = (R_0/a)\tan t$, where a is a constant to be determined later. In the new time coordinate t, (1) is

$$ds'^2 = R_0^2 a^{-2} \sec^4 t\, ds^2,$$

where

$$ds^2 = -dt^2 + a^2\cos^4 t\,[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{2}$$

with $-\pi/2 < t < +\pi/2$. Metric (2), which is conformal to the Einstein metric (1), is the metric we will use to form the Penrose diagram. It is the metric of a closed Friedmann universe with spatial topology S^3 and with initial and final singularities at $t = -\pi/2$ and $t = \pi/2$, respectively.

As is usual in the construction of Penrose diagrams,[4] I shall use the manifest spherical symmetry of (2) to restrict attention to the t, χ coordinates only; each point in my proposed Penrose diagram will represent a two-sphere if $\chi \neq 0$ or $\chi \neq \pi$. If $\chi = 0$ or $\chi = \pi$, the point in the diagram will represent a point at the origin of coordinates or the antipode of the three-sphere $t = \text{const}$, respectively.

The conventions for the representations of timelike, spacelike, and null lines in Penrose diagrams are the same as those in Minkowski diagrams: the vertical direction is timelike, the horizontal direction is spacelike, and most important, null lines are those at 45°, so that any line with an angle of less than 45° from the vertical is timelike, and any line with an angle of greater than 45° is spacelike. In the construction of the Penrose diagrams for the standard universes,[1–3] this convention is obtained by conformally mapping the universes into a proper subset of the Einstein universe (1) with $R_0 = 1$. For this particular Einstein universe, the metric in the (t', χ) "plane" is $ds^2 = -dt'^2 + d\chi^2$, which manifestly has null lines at 45°, and indeed the coordinates (t', χ) measure proper time and distance, respectively.

We can retain the Minkowski diagram conventions in a two-dimensional pictorial representation of (2) only by using proper time and distance in this case also. The coordinate t

 0022-2488/86/020559-03$02.50 © 1986 American Institute of Physics

already measures proper time in (2); infinitesimal proper distance r in the χ direction is $dr = a\cos^2 t\, d\chi$. Integrating in the χ direction gives

$$r = a\chi\cos^2 t + f(t), \tag{3}$$

where $0 \leqslant \chi \leqslant \pi$. The function $f(t)$ must vanish if r is to measure proper distance. [Alternatively, we can obtain $f(t) = 0$ by imposing the usual Friedmann universe requirement that the ratio of the proper distances between two fundamental observers (those that have a constant coordinate distance) at two different times be a function of proper time t alone. See Ref. 7 for a discussion of this requirement.]

A restriction on the constant a is obtained by requiring that the paths of all the fundamental observers $\chi = \text{const}$ be manifestly timelike in the (t, r) coordinates. This requirement means that we must have $|dr/dt| < 1$ for all fundamental observers. For fundamental observers, $|dr/dt| = a|\sin 2t\,|\chi$, so this inequality will be satisfied for all fundamental observers for all time if $a \leqslant 1/\pi$, since $0 \leqslant \chi \leqslant \pi$ and $-\pi/2 \leqslant t \leqslant \pi/2$.

As suggested by Penrose,[1] the diagram will look more symmetric if the fundamental observer, which is stationary in the diagram, is chosen to be the maximum area two-sphere in a $t = \text{const}$ hypersurface. Thus the paths of the fundamental observers in the (t, r) "plane" will be given by

$$r = a\cos 2t(\chi - \pi/2), \tag{4}$$

with $a < 2/\pi$.

The Penrose diagram for the Einstein static universe is given in Fig. 1: the curved dashed lines are the fundamental observers at the origin $\chi = 0$ and at the antipodal point $\chi = \pi$.

The dotted line inclined at 45° is a light ray passing back and forth between these two points in space. The magnitude of the velocity dr/dt of the origin of coordinates and the antipodal point reaches a maximum at $t = \pm\pi/4$, and decreases to zero at $t = \pm\pi/2$. The decrease to zero allows the universe (2) to slide into the singularity, so that light rays can pass back and forth between the points $\chi = 0$ and $\chi = \pi$ an infinite number of times. The rough diagram of the Einstein universe given by Penrose (Ref. 1, Fig. 9) did not have this sliding property; the velocity of the fundamental observers increased into the singularity, thus making it impossible for a light ray to pass between fundamental observers an infinite number of times. This rough diagram therefore does not represent correctly the causal structure of the Einstein universe.

The singularities at $t = \pm\pi/2$ are the ideal points i^+, i^-. Although these points are attached to the space-time at a finite distance, they are not attached smoothly. This is unavoidable, for the c boundary consists of two points rather than two three-spheres as in the dust Friedmann universe.

The metric (2) regarded as a Friedmann universe is interesting in its own right, for it is a nonstatic, $k = +1$ Friedmann universe that is a simple example of what Budic and Sachs[3] termed a deterministic space-time: a space-time whose future and past are completely determined by data given on the past light cone of any point. The only examples

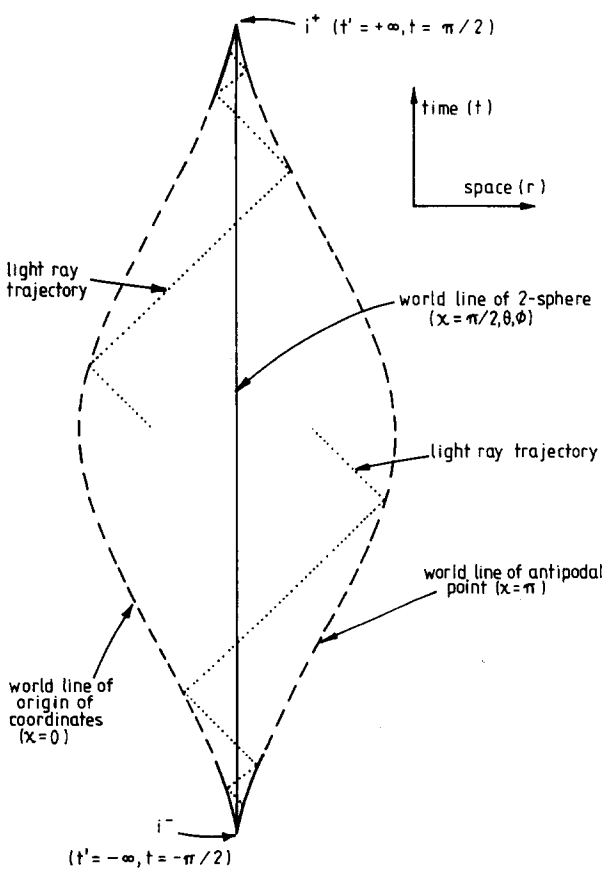


FIG. 1. The Einstein static universe.

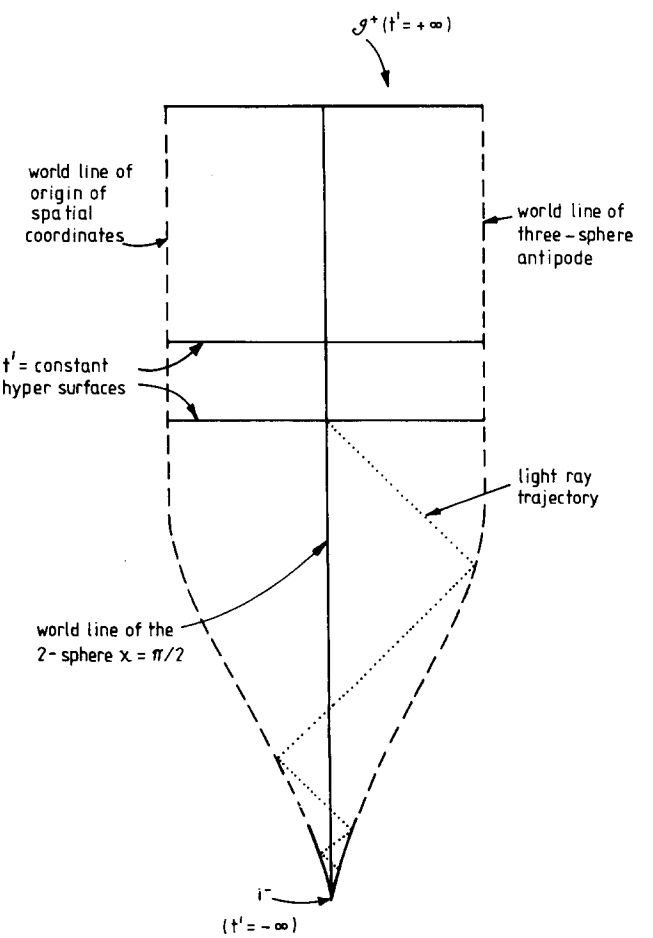


FIG. 2. The Eddington–Lemaitre universe.

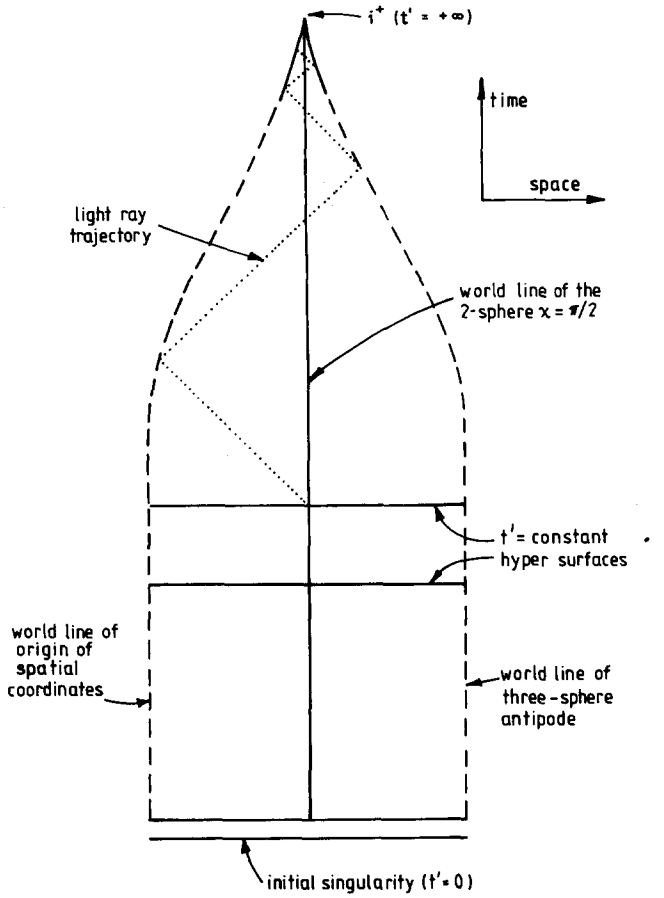


FIG. 3. The Eddington–Lemaitre–Bondi universe.

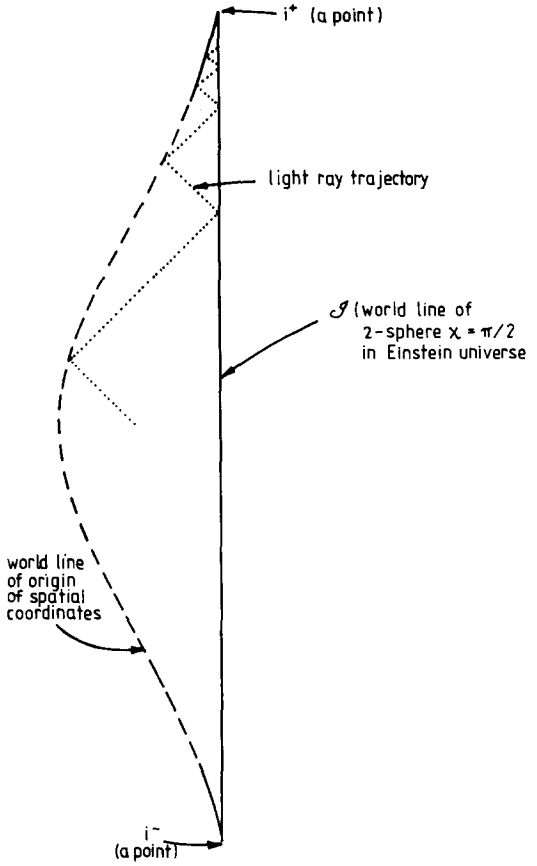


FIG. 4. Anti-de Sitter space.

of deterministic space-times given by Budic and Sachs[12] were obtained by topological identifications. The deterministic nature of (2) comes from the metric.

The Eddington–Lemaitre universe[5] is a $k = +1$ Friedmann universe that asymptotically approaches the Einstein static universe as $t' \to -\infty$, and the de Sitter universe as $t' \to +\infty$. It thus has a conformal structure of the Einstein universe in the past, and the conformal structure of de Sitter space in the future. The conformal structure of de Sitter space is well-known (e.g., Ref. 4), so the Penrose diagram for the Eddington–Lemaitre universe is as pictured in Fig. 2.

The Eddington–Lemaitre–Bondi universe is a $k = +1$ Friedmann universe that begins in an initial singularity (with c-boundary topology S^3) and asymptotically approaches the Einstein universe. The Penrose diagram for such a space-time is given in Fig. 3. I use the conventions of Hawking and Ellis[4]: double lines in the Penrose diagram represent singularities.

Anti-de Sitter space is conformal to the part of the Einstein universe with $-\infty < t' < +\infty$, but with $0 \leqslant \chi < \pi/2$. If we follow the conventions used above for the Einstein universe (the two-sphere labeled by $\chi = \pi/2$ is kept fixed), then the Penrose diagram for anti-de Sitter space including the ideal points i^-, i^+ is correctly given in Fig. 4. The c boundary consists of these points i^+, i^- together with a timelike hypersurface $\mathscr{I}$ which has topology $S^2 \times R^1$ and which begins at the point i^- and ends at the point i^+.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. PHY-84-09672.

[1] R. Penrose, "Conformal treatment of infinity," in *Relativity, Groups, and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964), pp. 563–584.

[2] R. Penrose, "Singularities of spacetime," in *Theoretical Principles in Astrophysics and Relativity*, edited by N. R. Lebovitz, W. H. Reid, and P. O. Vandervoort (Univ. Chicago P., Chicago, 1978), pp. 217–243.

[3] R. Penrose, "Structure of space-time," in *Battelle Recontres: 1967 Lectures in Mathematics and Physics*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968), pp. 121–235.

[4] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).

[5] H. Bondi, *Cosmology* (Cambridge U. P., Cambridge, 1961), 2nd ed.

[6] C. W. Misner, "Mixmaster universe," Phys. Rev. Lett. **22**, 1071 (1969).

[7] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[8] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, NJ, 1975).

[9] R. Penrose, "Singularities in cosmology," in *Confrontation of Cosmological Theories with Observational Data*, edited by M. S. Longair (Reidel, Dordrecht, 1974), pp. 263–272.

[10] R. Penrose, "Singularities and time-asymmetry," in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979), pp. 581–693.

[11] J. D. Barrow, "Chaotic behaviour in general relativity," Phys. Rep. **81**, 1 (1982).

[12] R. Budic and R. K. Sachs, "Deterministic space-times," Gen. Relativ. Gravit. **7**, 21 (1976).

The most general magnetized Kerr–Newman metric

Nora Bretón Baez and Alberto García Díaz
Centro de Investigación y de Estudios Avanzados del IPN, Departamento de Física, Apartado Postal 14-740, 07000 México D. F., Mexico

(Received 21 May 1985; accepted for publication 16 October 1985)

The complete magnetic generalization of the Kerr–Newman (KN) metric obtainable by applying a Harrison transformation to the Ernst potentials determined for an arbitrary linear combination of the two Killing vectors of the seed KN metric is derived.

I. INTRODUCTION

The main purpose of this work is to establish the most general magnetized Kerr–Newman solution, MKN for short, that can be derived by applying the Harrison generating solution technique[1] to the seed Kerr–Newman solution[2] (KN).

Let the KN metric be given as

$$g=\Delta(dr^2/R+d\theta^2)+(A/\Delta)\sin^2\theta\,d\phi^2 - 4a\,[(mr-\nu)/\Delta]\sin^2\theta\,d\phi\,dT - [(R-a^2\sin^2\theta)/\Delta]\,dT^2, \tag{1.1}$$

accompanied by the electromagnetic two-form

$$\omega=\tfrac{1}{2}(f_{\mu\nu}+{}^{*}f_{\mu\nu})dx^{\mu}\wedge dx^{\nu} = d\left\{\frac{(e+ig)}{r+ia\cos\theta}\,[dT-(a-ir\cos\theta)d\phi)]\right\}, \tag{1.2}$$

where

$$\Delta=r^2+a^2\cos^2\theta,\quad R=r^2+a^2-2mr+2\nu,\ 2\nu:=e^2+g^2,\quad A=(r^2+a^2)^2-Ra^2\sin^2\theta \tag{1.3}$$

(m represents the mass, a stands for the rotation parameter, and e and g are the electric and magnetic charges, respectively). The coordinates $\{T,r,\theta,\phi\}$ run the values $-\infty<T<\infty$, $0<r<\infty$, $0\leqslant\theta\leqslant\pi$, $0\leqslant\phi\leqslant2\pi$, respectively.

Choosing a Killing vector as a linear combination of the two Killing directions ∂_ϕ and ∂_T according to

$$K^{\mu}=\alpha\delta^{\mu}_{T}+\beta\delta^{\mu}_{\phi}, \tag{1.4}$$

one defines a function f as

$$-f=K^{\mu}K_{\mu}=(1/\Delta)\{[\alpha a-\beta(r^2+a^2)]^2\sin^2\theta - [\alpha-\beta a\sin^2\theta]^2R\}=:D/\Delta, \tag{1.5}$$

and the Killing one-form as

$$K=K_{\mu}\,dx^{\mu}=f\left[-\frac{(\alpha\,dT+\beta\,d\phi)}{(\alpha^2+a^2\beta^2)} + W(\alpha\,d\phi-\beta\,dT)\right], \tag{1.6}$$

where the function W is

$$W=-[D(\alpha^2+a^2\beta^2)]^{-1}\times\{\alpha\beta[a^2(1-\sin^4\theta)R+r^2\sin^2\theta(r^2+2a^2)] + a(\alpha^2-a^2\beta^2)(R-r^2-a^2)\sin^2\theta\}. \tag{1.7}$$

Notice that the representation of K is not unique. For instance, adding and subtracting in W the term $(\alpha/\beta)\times(\alpha^2+a^2\beta^2)^{-1}$, one arrives at K of the form

$$K=f[-d\phi/\beta+W(\alpha\,d\phi-\beta\,dT)],$$

where now W is given by

$$W=-(1/\beta D)[\alpha(R-a^2\sin^2\theta) + \beta a(r^2+a^2-R)\sin^2\theta].$$

In terms of the quantities defined above, the metric (1.1) can be written as

$$g=f^{-1}\{f\Delta[dr^2/R+d\theta^2] + R\sin^2\theta(\alpha\,d\phi-\beta\,dT)^2-K\otimes K\}. \tag{1.8}$$

The complex Ernst potentials Φ and $\mathscr{E}$ associated to the KN metric and its representation (1.8) can be evaluated from the equations

$$d\Phi=-iK\lrcorner\,\omega,\quad d\mathscr{E}=iK\lrcorner(dK+{}^{*}dK)-2\bar{\Phi}\,d\Phi, \tag{1.9}$$

where $\lrcorner$ denotes the step product and * stands for Hodge's star operation.[3]

The Φ potential amounts to

$$\Phi=[(e+ig)/(r+ia\cos\theta)]\,[\alpha-\beta(a-ir\cos\theta)], \tag{1.10}$$

while the $\mathscr{E}$ potential is

$$\mathscr{E}=f-(2\nu/\Delta)[(\alpha-\beta a)^2+\beta^2r^2\cos^2\theta] + 2(i/\Delta)\cos\theta\{m[a(\alpha-\beta a\sin^2\theta)^2-\beta\Delta(3(\alpha-\beta a) + \beta a\cos^2\theta)]+\beta r[\alpha\Delta+2\nu(\alpha-\beta a\sin^2\theta)]\}. \tag{1.11}$$

For the sake of simplicity we have omitted the additive constants Φ_0 and $\mathscr{E}_0$.

II. MAGNETIZED KERR–NEWMAN CLASS OF SOLUTIONS

According to the magnetizing process (Harrison transformations), the new generated metric assumes the form

$$g=f'^{-1}\{f\Delta[dr^2/R+d\theta^2] + R\sin^2\theta(\alpha\,d\phi-\beta\,dT)^2-K'\otimes K'\}, \tag{2.1}$$

where

$$f'=|\Psi|^{-2}f,\quad K'=f'[-(\alpha\,dT+\beta\,d\phi)/(\alpha^2+a^2\beta^2) + W'(\alpha\,d\phi-\beta\,dT)]. \tag{2.2}$$

 0022-2488/86/020562-02$02.50 © 1986 American Institute of Physics

The complex function Ψ is determined as a function of the "seed" Ernst potentials according to

$$\Psi = 1 - 2(E - iB)\Phi - \delta\mathscr{E},\ \delta := E^2 + B^2, \tag{2.3}$$

where E and B are real constants representing the electric and magnetic field parameters. This function is crucial in the determination of the new Ernst potentials, which are given by

$$\Phi' = \Psi^{-1}(\Phi + (E + iB)\mathscr{E}),\ \mathscr{E}' = \mathscr{E}\Psi^{-1}. \tag{2.4}$$

The function W' ought to fulfill the equation

$$dW' = \Psi\bar{\Psi}\,dW + if^{-1}a\sin\theta \times\{(\Psi\bar{\Psi}_{,\theta} - \bar{\Psi}\Psi_{,\theta})dr - (\Psi\bar{\Psi}_{,r} - \bar{\Psi}\Psi_{,r})d\theta\}, \tag{2.5}$$

which has a general solution

$$\begin{aligned}
DW' = {} & D(W + W_0) - 4(Ee + Bg)[\alpha a - \beta(a^2 + r^2)]r\sin^2\theta - 4(Eg - Be)(\alpha - \beta a\sin^2\theta)R\cos\theta \\
& + 6(\delta/a)\,\nu[(\alpha a - \beta(a^2 + r^2))(\alpha a - \beta(a^2 - r^2))\sin^2\theta + (\alpha - \beta a\sin^2\theta)(\alpha - \beta a(1 + \cos^2\theta))R] \\
& - 4(\delta/a)(Ee + Bg)\{m(\alpha - \beta a)(\alpha - \beta a\sin^2\theta)(\alpha - \beta a - 3\beta a\cos^2\theta)R \\
& + m(\alpha - \beta a)[\alpha a - \beta(a^2 + r^2)][\alpha a - \beta(a^2 - r^2)]\sin^2\theta \\
& - ar[\alpha a - \beta(a^2 + r^2)][\beta^2 r^2 + (\alpha - \beta a)^2]\sin^2\theta \\
& - \beta arR\sin^2\theta[2(\alpha - \beta a)(\alpha - \beta a\sin^2\theta) + \beta(\alpha a - \beta(a^2 + r^2))\cos^2\theta] \\
& \times 4\delta(Eg - Be)R\cos\theta\{(\alpha - \beta a\sin^2\theta)[\beta(a^2 + 2\nu)\cos^2\theta - (\alpha - \beta a)^2] \\
& + \beta\sin^2\theta[\beta r^2(\alpha - \beta a\sin^2\theta) - 2(\alpha - \beta a)(\alpha a - \beta(r^2 + a^2))]\} \\
& + \delta^2\{(\alpha - \beta a)^4 a^{-1}[R(2m^2 + a^2\cos^2\theta) + a^2\sin^2\theta(2m^2 - 4mr + r^2)] \\
& - (\alpha - \beta a)^3\beta[R(8m^2 - 4a^2 - 4\nu + a^2\cos^2\theta)\cos^2\theta + 2mrR\sin^2\theta - r^2(r^2 + 4a^2 + 4\nu)\sin^2\theta] \\
& + (\alpha - \beta a)^2\beta^2[2aR\cos^2\theta((a^2 + 2\nu)(2 - \cos^2\theta) - 3m^2\cos^2\theta) \\
& + 2r^2\sin^2\theta(a(r^2 - 6mr - 2a^2 - 4\nu) - r^2a^{-1}m^2) - 12mraR\sin^2\theta\cos^2\theta] \\
& - (\alpha - \beta a)\beta^3[R\cos^4\theta((a^2 + 2\nu)(2\nu + a^2\cos^2\theta) - 4m^2a^2\cos^2\theta) + (2\nu - 6mr + r^2)r^4\sin^2\theta \\
& + 3arR\sin^2\theta\cos^2\theta(r^3 + ra^2\cos^2\theta + 4\nu r - 2m(2r^2 + a^2\cos^2\theta))] - \beta^4 a[R(a^2 + 2\nu)^2\cos^6\theta + r^6\sin^2\theta \\
& + r^2R\sin^2\theta\cos^2\theta(3r^2 + (3a^2 + 6\nu - 2mr)\cos^2\theta)]\},
\end{aligned} \tag{2.6}$$

where W is given by Eq. (1.7) and W_0 is an arbitrary constant.

The electromagnetic field coupled to the metric line element (2.1) is described by the two-form

$$F'\omega' = K' \wedge d\Phi' + *\{K' \wedge d\Phi'\}, \tag{2.7}$$

where K' and Φ' are given, correspondingly, by (2.2) and (2.4).

We shall denote this magnetic Kerr–Newman metric structure by MKN(α, β).

It contains as particular branch the MKN(0,1) solution[4,5]; by setting in the solution presented above $\alpha = 0$, $\beta = 1$, and $g = 0$, accompanied by a redefinition of the coordinate ϕ, $\phi \to a^2\phi$, one arrives at the solution Ref. 5 in spherical coordinates. [In the expression of f of Ref. 5, formula (2), the factor a^2 in front of f was omitted; the correct definition of f is $a^2 f = -\Delta^{-1}A\,P$.]

In the limit a tending to zero, by choosing the constant W_0 as

$$W_0 = -2\delta a^{-1}[3\nu - 2m\alpha(Ee + Bg) + m^2\alpha^2\delta],$$

one obtains from the MKN(α, β) metric the corresponding magnetic Reissner–Nordstrom solution.

ACKNOWLEDGMENTS

One of us (N.B.B.) acknowledges the Consejo Nacional de Ciencia y Tecnología for financial assistance through a doctoral fellowship. This work was also partially supported by COSNET-SEP.

[1] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[2] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[3] F. J. Ernst and J. F. Plebański, Ann. Phys. **107**, 266 (1977).
[4] F. J. Ernst and W. J. Wild, J. Math. Phys. **17**, 182 (1976).
[5] A. García Díaz, J. Math. Phys. **26**, 155 (1985).

Soliton solutions to the vacuum Einstein equations obtained from a nondiagonal seed solution

Patricio S. Letelier
Departamento de Física, Universidade de Brasília, 70.910, Brasília, D. F., Brazil

(Received 7 September 1984; accepted for publication 3 October 1985)

The one- and two-soliton solutions obtained by using the Belinsky–Zakharov inverse scattering method in the particular case where the seed solution is taken as the one due to van Stockum are studied. For the above-mentioned particular case the inverse scattering problem is solved in terms of a single quadrature.

I. INTRODUCTION

Recently we studied solutions to the vacuum Einstein field equations obtained by using the Belinsky–Zakharov[1,2] inverse scattering method (ISM) with a diagonal seed solution.[3–9] We found that in this case the inverse scattering problem reduces to the computation of a single quadrature.[3,4] This fact was used to study some known solutions as well as to generate some new ones.[3] The ISM and other similar solution-generating techniques[10,11] seldom have been applied to nondiagonal seed solutions. An example of solutions obtained from a nondiagonal seed is the Belinsky–Francaviglia[12] solution that uses a Bianchi II vacuum cosmological model as a seed solution.

The aim of this paper is to study the application of the ISM to the van Stockum[13] solution, which is a nondiagonal solution depending on a single arbitrary function. We find that in this case the inverse scattering problem can also be solved in an explicit way, similar to the diagonal seed solution case. We have two main reasons to study the inverse scattering problem associated with the above-mentioned solution. First, to obtain new nontrivial exact solutions to the vacuum Einstein equations, and second, to better understand the mechanism of the ISM, since one of the shortcomings of most of the new-solution-generating algorithms like the ISM, Bäcklund transformations, etc.,[10,11] is that they do not give significant information about the physical or geometric meaning of the generated solutions. Among the solutions generated using the ISM with a very simple diagonal seed solution,[3] we were able to identify some already known solutions. But, even in these simple cases the meaning of the parameters introduced by the ISM was not easy to understand.

Even though the van Stockum solution has highly unphysical properties, it is known to be related to important solutions like the Kerr solution.[14] Also, recently it has been used to generate an interesting family of new solutions to the vacuum Einstein equations.[15]

In Sec. II we study the Einstein equations, together with the ISM, for a metric with a general signature and two commuting Killing vectors, i.e., we present a unified treatment of the Riemannian case, the axially symmetric case, the cylindrically symmetric case, and the zero signature case. In Sec. III we solve the inverse scattering problem for the van Stockum solution. Finally, in Sec. IV we study the one- and two-soliton solutions associated with the previously mentioned particular solution.

II. THE EINSTEIN EQUATIONS AND THE ISM

The metric that we shall consider in this paper is

$$ds^2 = e^{\omega}\sigma_{AB}\,dx^A\,dx^B + \gamma_{ab}\,dx^a\,dx^b, \tag{2.1}$$

where $(x^A, x^a) = (u, v, x, y)$, the indices A, B, etc., run from 1 to 2, and the indices a, b, etc., run from 3 to 4. The functions ω and γ_{ab} depend only on x^A; (σ_{AB}) and (γ_{ab}) are symmetric 2×2 matrices characterized by

$$\sigma = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.2}$$

$$\det\gamma = \eta\alpha^2, \quad \alpha^2 > 0, \tag{2.3}$$

where ϵ and η are two sign functions. Depending on the value of these functions, the line element (2.1) can be used to represent the following spaces: (a) cylindrically symmetric space-time ($\epsilon = -1$; $\eta = 1$), (b) axially symmetric space-time ($\epsilon = 1$; $\eta = -1$), (c) Riemannian space with axial symmetry ($\epsilon = 1$; $\eta = 1$), and (d) zero signature space ($\epsilon = -1$; $\eta = -1$).

The vacuum Einstein equations for the metric (2.1) reduce to the following system of equations:

$$\sigma^{AB}(\alpha\gamma_{ab,A}\gamma^{bc})_{,B} = 0, \tag{2.4}$$

$$(\ln\alpha)_{,2}\omega_{,1} + (\ln\alpha)_{,1}\omega_{,2} = 2(\ln\alpha)_{,12} - \tfrac{1}{2}\gamma_{ab,1}\gamma^{ab}_{,2}, \tag{2.5a}$$

$$(\ln\alpha)_{,1}\omega_{,1} + (\ln\alpha)_{,2}\omega_{,2} = (\ln\alpha)_{,11} - \epsilon(\ln\alpha)_{,22} + \tfrac{1}{4}\gamma_{ab,1}\gamma^{ab}_{,1} + (\epsilon/4)\gamma_{ab,2}\gamma^{ab}_{,2}, \tag{2.5b}$$

where $(\,)_{,A}$, etc., indicate partial derivation with respect to the coordinate x^A. Note that in Eqs. (2.4) and (2.5) the indicator η does not appear explicitly. It can be easily shown[16] that there is no loss of generality if we choose $\alpha = u$. In this particular "gauge," Eqs. (2.4) and (2.5) can be written in the more appealing form[17]

$$\sigma^{AB}(u\gamma_{,A}\gamma^{-1})_{,B} = 0, \tag{2.6}$$

$$\omega_{,1} = -(1/u) - \tfrac{1}{4}u\,\mathrm{Tr}(\gamma_{,1}\gamma_{,1}^{-1} - \epsilon\gamma_{,2}\gamma_{,2}^{-1}), \tag{2.7a}$$

$$\omega_{,2} = -\tfrac{1}{2}u\,\mathrm{Tr}(\gamma_{,1}\gamma_{,2}^{-1}), \tag{2.7b}$$

where $\gamma_{,A}^{-1} \equiv (\gamma^{-1})_{,A}$. Equation (2.6) is the integrability condition for the system of equations (2.7).

The ISM used to solve (2.6) is based on the fact that (2.6) is the integrability condition for the overdetermined system of equations

$$D_u\psi = [(uU + \lambda V)/(u^2 + \epsilon\lambda^2)]\psi, \tag{2.8a}$$

$$D_v\psi = [(uV - \epsilon\lambda U)/(u^2 + \epsilon\lambda^2)]\psi, \tag{2.8b}$$

 0022-2488/86/020564-04$02.50 © 1986 American Institute of Physics

$$D_u \equiv \partial_u + [2\lambda u/(u^2 + \epsilon\lambda^2)]\partial_\lambda, \tag{2.9a}$$

$$D_v \equiv \partial_v - [2\epsilon\lambda^2/(u^2 + \epsilon\lambda^2)]\partial_\lambda, \tag{2.9b}$$

where ψ is a two by two matrix function of u, v, and the spectral parameter λ, $U \equiv u\gamma_{,1}\gamma^{-1}$, and $V \equiv u\gamma_{,2}\gamma^{-1}$. The inverse scattering problem associated with n simple poles of the scattering matrix can be explicitly solved yielding the n-soliton solution[1,2]

$$(\gamma_n)_{ab} = (\gamma_0)_{ab} - \sum_{k,l}^{n} \frac{(\Gamma^{-1})_{lk} N_a^{(l)} N_b^{(k)}}{\mu_l \mu_k}, \tag{2.10}$$

$$\Gamma_{lk} \equiv m_a^{(l)}(\gamma_0)_{ab} m_b^{(k)}/(\mu_k \mu_l + \epsilon u^2), \tag{2.11a}$$

$$N_a^{(k)} \equiv m_b^{(k)}(\gamma_0)_{ba}, \quad m_a^{(k)} \equiv m_{0b}^{(k)} M_{ba}^{(k)}, \tag{2.11b}$$

$$M^{(k)} \equiv \psi_0^{-1}\Big|_{\lambda=\mu_k}, \tag{2.11c}$$

$$\mu_{k,1} = 2u\mu_k/(u^2 + \epsilon\mu_k^2),$$

$$\mu_{k,2} = -2\epsilon\mu_k^2/(u^2 + \epsilon\mu_k^2), \tag{2.12}$$

where the $m_{0b}^{(k)}$ are arbitrary constants; γ_0 and ψ_0 refer to a known solution to (2.6) and its associated "wave function" solution to (2.8). The known solution γ_0 is called the "seed" or background solution.

The poles' trajectories can be computed using Eqs. (2.12); we find

$$\mu_k = \alpha_k - v + \epsilon_k |[(\alpha_k - v)^2 + \epsilon u^2]^{1/2}|, \tag{2.13}$$

where $\epsilon_k = \pm 1$. Also, the determinant associated with (2.10) can be explicitly computed:

$$\det \gamma_n = \eta(-\epsilon)^n u^{2(n+1)} \prod_{k=1}^{\eta} \mu_k^{-2}. \tag{2.14}$$

The expression (2.10) is a solution of (2.6) but does not satisfy (2.3) (with $\alpha = u$). To remedy this problem we can define the physical quantity[1,2]

$$\gamma^{\mathrm{Ph}} = u\gamma/(|\det\gamma|)^{1/2} \tag{2.15}$$

that satisfies both Eq. (2.3) with $\alpha = u$ and (2.6).

The metric coefficient ω can also be explicitly computed, yielding

$$\omega_n = \omega_0 + \ln\left[u^{-n^2/2}\left(\prod_{k=1}^{n}\mu_k\right)^{n+1} \times \prod_{\substack{k,l=1 \\ k>l}}^{n} (\mu_k - \mu_l)^{-2} \det\Gamma\right], \tag{2.16}$$

where ω_0 is the function ω associated with the seed solution γ_0, and the factor

$$\prod_{\substack{k,l=1 \\ k>l}}^{n} (\mu_l - \mu_k)^{-2}$$

should be set equal to 1 for $n = 1$.

To close this section we want to point out that the ISM solution to (2.6) does not depend on the value of η. A unified treatement of the elliptic case $\epsilon = 1$ and the hyperbolic case $\epsilon = -1$ can be also found in Ref. 18. Note that in Ref. 18 it is assumed that $\epsilon = -\eta$. Thus the present results are slightly more general since we assume that ϵ and η are not related. Riemannian metrics can be used to generate axially symmetric odd-number-soliton space-times.[3] They also can be used to generate even-number-soliton solutions associated with Euclidean SU(2) gauge fields.[19]

III. THE SEED SOLUTION AND THE FUNCTION ψ_0

The seed solution that we shall study in this section is the solution of Eqs. (2.6) and (2.7) given by

$$\gamma = u\begin{pmatrix} 0 & 1 \\ 1 & \phi \end{pmatrix}, \tag{3.1}$$

$$\omega_0 = \ln(C_0/u^{1/2}), \tag{3.2}$$

where C_0 is an integration constant and ϕ satisfies

$$\phi_{,11} + \phi_{,1}/u + \epsilon\phi_{,22} = 0. \tag{3.3}$$

In this case, $\eta = -1$. The van Stockum solution is obtained by putting $\epsilon = 1$ in (3.3). Note that there does not exist a hyperbolic version of this metric with signature ± 2.

A convenient change of function to solve (2.8) is

$$\Lambda = (u^2 - 2\epsilon\lambda v - \epsilon\lambda^2)^{1/2}\psi. \tag{3.4}$$

From (3.4), (2.8), and (2.9) we get the equivalent system of equations to (2.8):

$$D_u\Lambda = \{[u(U-I) + \lambda V]/(u^2 + \epsilon\lambda^2)\}\Lambda, \tag{3.5a}$$

$$D_v\Lambda = \{[uV - \epsilon\lambda(U-I)]/(u^2 + \epsilon\lambda^2)\}\Lambda. \tag{3.5b}$$

By putting $\lambda = 0$ in (3.5) we get the "initial condition"[1,2]

$$\Lambda|_{\lambda=0} = \gamma/u. \tag{3.6}$$

Equations (3.5) for the seed solution (3.1) reduce to

$$D_u\Lambda = [(u^2\phi_{,1} + \lambda u\phi_{,2})/(u^2 + \epsilon\lambda^2)]\mathscr{E}\Lambda, \tag{3.7a}$$

$$D_v\Lambda = [(u^2\phi_{,2} - \epsilon\lambda u\phi_{,1})/(u^2 + \epsilon\lambda^2)]\mathscr{E}\Lambda, \tag{3.7b}$$

where

$$\mathscr{E} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{3.8}$$

It is a matter of simple verification to show that

$$\Lambda = \begin{pmatrix} 0 & 1 \\ 1 & F \end{pmatrix} \tag{3.9}$$

satisfies (3.6) and (3.7) as long as

$$D_uF = (u^2\phi_{,1} + \lambda u\phi_{,2})/(u^2 + \epsilon\lambda^2), \tag{3.10a}$$

$$D_vF = (u^2\phi_{,2} - \epsilon\lambda\phi_{,2})/(u^2 + \epsilon\lambda^2), \tag{3.10b}$$

$$F|_{\lambda=0} = \phi. \tag{3.10c}$$

Solutions to (3.10) for different functions ϕ are studied in Refs. 3–9. Since in the solution (2.10)–(2.16) ψ_0 only enters evaluated along the poles' trajectories, to compute the soliton solutions we only need

$$F_k \equiv F|_{\lambda=\mu_k}. \tag{3.11}$$

From (2.12) we find that (3.10a) and (3.10b) along the poles' trajectories reduce to

$$\partial_u F_k = (u/2\mu_k)(\mu_{k,1}\phi_{,1} - \epsilon\mu_{k,2}\phi_{,2}), \tag{3.12a}$$

$$\partial_v F_k = (u/2\mu_k)(\mu_{k,1}\phi_{,2} + \mu_{k,2}\phi_{,1}). \tag{3.12b}$$

Thus

$$F_k = \int \frac{u}{2\mu_k}[(\mu_{k,1}\phi_{,1} - \epsilon\mu_{k,2}\phi_{,2})du + (\mu_{k,1}\phi_{,2} + \mu_{k,2}\phi_{,1})dv]. \tag{3.13}$$

The existence of F_k is guaranteed by Eq. (3.3) and by the fact that $\ln \mu_k$ also satisfies this same equation. The condition (3.10c) is also satisfied, since

$$\left.\frac{\mu_{k,1}}{\mu_k}\right|_{\mu_k\to 0} = \frac{2}{u}, \quad \left.\frac{\mu_{k,2}}{\mu_k}\right|_{\mu_k\to 0} = 0. \tag{3.14}$$

The expression (3.13) unifies the corresponding results presented in Refs. 3 and 4.

IV. ONE- AND TWO-SOLITON SOLUTIONS

From (3.4), (3.9), (3.11), and (2.10)–(2.16) we find that the one-soliton solution associated with the seed solution (3.1) and (3.2) can be written as

$$\gamma_{33}^{\mathrm{Ph}} = \frac{-p_1(\mu_1 + \epsilon u^2/\mu_1)}{2q_1 + P_1(\phi - 2F_1)}, \tag{4.1a}$$

$$\gamma_{34}^{\mathrm{Ph}} = \frac{(\mu_1 - \epsilon u^2/\mu_1)(q_1 - p_1F_1) - \epsilon p_1(u^2/\mu_1)\phi}{2q_1 + p_1(\phi - 2F_1)}, \tag{4.1b}$$

$$\gamma_{44}^{\mathrm{Ph}} = -\frac{(\mu_1 + \epsilon u^2/\mu_1)(q_1 - p_1F_1)^2 + \epsilon p_1(u^2/\mu_1)[2q_1 + p_1(\phi - 2F_1)]\phi}{p_1[2q_1 + p_1(\phi - 2F_1)]}, \tag{4.1c}$$

$$\omega_1 = \ln\left[C_0\frac{[2q_1 + p_1(\phi - 2F_1)]\mu_1}{\mu_1^2 + \epsilon u^2}\right], \tag{4.2}$$

where we have introduced the notation $p_k \equiv m_{03}^{(k)}$ and $q_k \equiv m_{04}^{(k)}$. Also, we have denoted the "renormalized" integration constant by the same symbol used in (3.2), a practice that we shall follow in this paper. First we want to point out that (4.1) and (4.2) is not a solution of physical interest, since for either value of ϵ we have that the metric (2.1) constructed with (4.1) and (4.2) has the wrong signature. In the $\epsilon = -1$ case we have that the seed solution has the signature $(-\ +\ +\ -)$ and the one-soliton ISM produces either $(-\ +\ +\ -)$ or $(-\ +\ -\ +)$ as Eq. (2.14) with $\eta = 1$ indicates. Also, in the case $\epsilon = +1$ we have that the seed solution has signature $(+\ +\ +\ -)$ that changes to either $(+\ +\ +\ +)$ or $(+\ +\ -\ -)$. An interesting feature of this solution is that the particular case obtained by letting $F_1 = \phi = 0$ in (4.1) and (4.2) is a flat solution; this fact is unexpected, since we start applying the ISM to a nonflat background solution. As a matter of fact (3.1) and (3.2) with $\phi = 0$ represent a Taub-like solution.[20]

In a similar way we find that the two-soliton solution can be written as

$$\gamma_{33}^{\mathrm{Ph}} = \frac{\epsilon p_1p_2u(\mu_2 - \mu_1)}{\mu_1\mu_2\Delta}\left\{\frac{p_1f_2}{\mu_2}[2q_2 + p_2(\phi - 2F_2)] - \frac{p_2f_1}{\mu_1}[2q_1 + p_1(\phi - 2F_1)]\right\}, \tag{4.3a}$$

$$\begin{aligned}\gamma_{34}^{\mathrm{Ph}} = {}& \frac{\mu_1\mu_2}{u} + \frac{\mu_1\mu_2 + \epsilon u^2}{\mu\mu_1\mu_2\Delta}\{p_1p_2f_2[2q_2 + p_2(\phi - 2F_2)] \\ & \times [q_1 + p_1(\phi - F_1)] - f_3[p_1q_1 + p_2q_1 \\ & + p_1p_2(\phi - F_1 - F_2)]\ [p_1q_2 + p_2q_1 \\ & + p_1p_2(2\phi - F_1 - F_2)] + p_1p_2f_1[2q_1 \\ & + p_1(\phi - 2F_1)]\ [q_2 + p_2(\phi - F_2)]\},\end{aligned} \tag{4.3b}$$

$$\begin{aligned}\gamma_{44}^{\mathrm{Ph}} = {}& \frac{\mu_1\mu_2}{u}\phi + \frac{\mu_1\mu_2 + \epsilon u^2}{u\mu_1\mu_2\Delta}\{p_2f_2[2q_2 + p_2(\phi - 2F_2)] \\ & \times [q_1 + p_1(\phi - F_1)]^2 - 2f_3[(p_1q_2 + p_2q_1 \\ & + p_1p_2(\phi - F_1 - F_2)]\ [q_1 + p_1(\phi - F_1)] \\ & \times [q_2 + p_2(\phi - F_2)] + p_1f_1[2q_1 + p_1(\phi - 2F_1)] \\ & \times [q_2 + p_2(\phi - F_2)]^2\},\end{aligned} \tag{4.3c}$$

$$\omega_2 = \ln\left[\frac{C_0\Delta}{u^{1/2}(\mu_1^2 + \epsilon u^2)(\mu_2^2 + \epsilon u^2)}\right], \tag{4.4}$$

where

$$f_1 = \mu_1^2(\mu_2^2 + \epsilon u^2)(\mu_1\mu_2 + \epsilon u^2), \tag{4.5a}$$

$$f_2 = \mu_2^2(\mu_1^2 + \epsilon u^2)(\mu_1\mu_2 + \epsilon u^2), \tag{4.5b}$$

$$f_3 = \mu_1\mu_2(\mu_1^2 + \epsilon u^2)(\mu_2^2 + \epsilon u^2), \tag{4.5c}$$

and

$$\begin{aligned}\Delta = {}& \epsilon u^2(\mu_2 - \mu_1)^2[p_1q_2 + p_2q_1 + p_1p_2(\phi - F_1 - F_2)]^2 \\ & + (\mu_1\mu_2 + u^2)^2[p_1q_2 - p_2q_1 - p_2p_1(F_2 - F_1)]^2.\end{aligned} \tag{4.6}$$

First we notice that the two-soliton solution (4.3) and (4.4) with $\epsilon = 1$ has the right signature, since a two-soliton transformation does not change the sign of the determinant of the seed solution, as Eq. (2.14) indicates. An interesting subcase is obtained by putting $p_2 = \phi = F_2 = 0$ in (4.3) and (4.4), i.e.,

$$\gamma_{33}^{\mathrm{Ph}} = 0, \quad \gamma_{34}^{\mathrm{Ph}} = -\epsilon u, \tag{4.7a}$$

$$\gamma_{44}^{\mathrm{Ph}} = -2\epsilon\epsilon_1q_1(\alpha_2 - \alpha_1)u/p_1\sqrt{(\alpha_1 - v)^2 + \epsilon u^2}, \tag{4.7b}$$

$$\omega_2 = \omega_0. \tag{4.8}$$

In deriving (4.7b) we have made use of the identities

$$\epsilon u^2(\mu_2 - \mu_1)^2 + (\mu_1\mu_2 + \epsilon u^2)^2 = (\mu_1^2 + \epsilon u^2)(\mu_2^2 + \epsilon u^2), \tag{4.9}$$

$$(\mu_2 - \mu_1)(\mu_1\mu_2 + \epsilon u^2)/\mu_1\mu_2 = 2(\alpha_2 - \alpha_1). \tag{4.10}$$

By taking $p_1 = q_2 = 0$ in (4.3) we get that $\gamma_{33}^{\mathrm{Ph}}$ and $\gamma_{34}^{\mathrm{Ph}}$, and ω are given by (4.7a) and (4.8), respectively, and that

$$\begin{aligned}\gamma_{44}^{\mathrm{Ph}} = {}& \frac{\mu_1\mu_2}{u}\phi + \frac{\mu_1\mu_2 + \epsilon u^2}{uf_3} \\ & \times [(f_2 - 2f_3)\phi + 2(f_3 - f_2)F_2].\end{aligned} \tag{4.11}$$

The complementary case, i.e., $p_2 = q_1 = 0$, is obtained by doing $f_2 \to f_1$ and $F_2 \to F_1$ in (4.11).

In the axially symmetric case we can interpret the two-soliton solution obtained from the Weyl solution as arising from the superposition of a potential ϕ and the potentials associated with two semi-infinite wires.[3] The particular case (4.7) and (4.8) tells us that the "flat" van Stockum solution is related through the ISM to the van Stockum solution associated with a point mass. The potential associated with (4.7), i.e., $\phi' \equiv \gamma_{44}^{\mathrm{Ph}}/u$, is obtained only in the special case

$\phi = F_k = p_2 = 0$. Note that the potential associated with the two-soliton generated from the Weyl solution is obtained from restrictions the constants p_k and q_k only.

The potential ϕ' associated with (4.11) is a solution of the *linear equation* (3.3), since the solution (4.11) has the same form as the van Stockum solution (modulo an invertion in the $\epsilon = 1$ case). This is an example of the nonlinear superposition principle[21] with a connecting "function" ϕ'. The solutions used in the superposition are ϕ, $\ln u$, $\ln \mu_1$, and $\ln \mu_2$. Really, in this case we have a generalized nonlinear superposition principle, since in (4.11) appears F_2, which is a functional of the above-mentioned particular solutions, i.e., in this case we have a connecting functional.

ACKNOWLEDGMENT

I want to thank Prof. M. F. Letelier (Universidad de Santiago) for the discussion we had about the nonlinear superposition principle.

[1]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)].
[2]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].
[3]P. S. Letelier, J. Math. Phys. **26**, 467 (1985).
[4]P. S. Letelier, J. Math. Phys. **25**, 2675 (1984).
[5]P. S. Letelier, J. Math. Phys. **26**, 326 (1985).
[6]P. S. Letelier, Phys. Rev. D **26**, 3728 (1982).
[7]P. S. Letelier, Rev. Bras. Fis. **13**, 698 (1983).
[8]P. S. Letelier, Rev. Bras. Fis. **14**, 371 (1984).
[9]P. S. Letelier, Class. Quantum Gravit. **2**, 419 (1985).
[10]C. M. Cosgrove, J. Math. Phys. **21**, 2417 (1980).
[11]C. M. Cosgrove, J. Math. Phys. **23**, 615 (1982), and references therein.
[12]V. Belinsky and M. Francaviglia, Gen. Relativ. Gravit. **14**, 213 (1982).
[13]W. J. van Stockum, Proc. R. Soc. Edinburgh **57**, 135 (1937).
[14]E. Herlt, Gen. Relativ. Gravit. **9**, 711 (1978).
[15]Y. Nakamura, J. Math. Phys. **24**, 606 (1983).
[16]See, for instance, P. S. Letelier, Nuovo Cimento B **69**, 145 (1982); T. Lewis, Proc. R. Soc. London Ser. A **136** (1932).
[17]Equations (2.6) and (2.7) are equivalent to those given in Refs. 1 and 2.
[18]G. A. Alekseev, Dokl. Akad. Nauk SSSR, Ser. Mat. Fiz. **256**, 827 (1981) [Sov. Phys. Dokl. **26**, 158 (1981)].
[19]P. S. Letelier, J. Math. Phys. **23**, 1175 (1982), and references therein.
[20]A discussion of this point can be found in Ref. 5.
[21]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1972), Vol. II, pp. 2 and 3.

Some exact solutions of Einstein–Dirac–Maxwell fields and a massive neutrino

Amar Chandra Patra
Department of Mathematics, Taki Government College, Taki, 24 Parganas, West Bengal, India
Dipankar Ray
Department of Mathematics, University of Jadavpur, Calcutta-700032, India

(Received 25 July 1984; accepted for publication 8 February 1985)

The field equations obtained by Krori, Chaudhury, and Bhattacharjee for Einstein–Dirac–Maxwell fields for several different metrics with a time-independent and a time-dependent Dirac field for a massless neutrino and for a massive neutrino are completely integrated. It is also noted that one of the cases, the particular solutions given by Krori, Chaudhury, and Bhattacharjee, is not correct. All other particular solutions are the special cases of the solutions obtained here.

I. INTRODUCTION

The field equations of Einstein–Dirac–Maxwell fields are

$$R_{jk} - \tfrac{1}{2}g_{jk}R = -8\pi G E_{jk} + 8\pi G T_{jk}, \tag{1.1}$$

$$\gamma^j \psi_{;j} + m\psi = 0, \tag{1.2}$$

$$F^{jk}{}_{,k} = 0, \tag{1.3}$$

$$F_{jk,l} + F_{kl,j} + F_{ij,k} = 0, \tag{1.4}$$

with

$$T_{jk} = \tfrac{1}{4}(\psi^\dagger \gamma_j \psi_{;k} - \psi^\dagger{}_{;k}\gamma_j \psi + \psi^\dagger \gamma_k \psi_{;j} - \psi^\dagger{}_{;j}\gamma_k \psi), \tag{1.5}$$

$$E_{jk} = -F^l_k F_{jl} + \tfrac{1}{4} g_{jk} F^{lm} F_{lm}. \tag{1.6}$$

We use units in which $h = c = 1$. We adopt the conventions of Jauch and Rohrlich[1] for Dirac γ matrices and the notations of Brill and Wheeler[2] with regard to $\psi^\dagger$, ψ^*, and $\nabla_\mu \psi$.

In a recent paper Krori, Chaudhury, and Bhattacharjee[3] sought solutions of coupled Einstein–Dirac–Maxwell equations for the following cases.

(i) A masseless neutrino with the following metrics where Dirac field ψ is time independent.

(a) The Weyl metric given by

$$ds^2 = e^{2u}dt^2 - e^{2K-2u}(dr^2 + dz^2) - r^2 e^{-2u} d\phi^2, \tag{1.7}$$

where u and K are functions of r alone.

(b) The static plane symmetric metric given by

$$ds^2 = e^{2u}(dx^2 - dt^2) + e^{2v}(dy^2 + dz^2), \tag{1.8}$$

where u and v are functions of x alone.

(ii) A massless neutrino with the following metrics where the Dirac field ψ is time dependent.

(a) The Einstein–Rosen[4] metric given by

$$ds^2 = e^{2(\alpha-\beta)}(dt^2 - dr^2) - r^2 e^{-2\beta} d\phi^2 - e^{2\beta} dz^2, \tag{1.9}$$

where α and β are functions of r alone.

(b) The static plane symmetric metric given by (1.8).

(iii) A massive neutrino with the following metrics.

(a) Marder's metric[5] given by

$$ds^2 = e^{2(\alpha-\beta)}(dt^2 - dr^2) - r^2 e^{-2\beta} d\phi^2 - e^{2(\beta+\nu)} dz^2, \tag{1.10}$$

where α, β, ν are functions of r alone.

(b) The plane symmetric metric given by (1.8)

In all these cases, the authors obtained the field equations and gave some particular solutions.

In the present work we obtain the complete set of solutions for these field equations.

It is also noted that for the case of massive neutrino with Marder's metric the solution given by Krori, Chaudhury, and Bhattacharjee[3] is not correct.

II. MASSLESS NEUTRINO WITH TIME-INDEPENDENT DIRAC FIELD

A. Equations

Krori, Chaudhury, and Bhattacharjee[3] obtained the following Einstein–Dirac–Maxwell field equations for the massless neutrino with time-independent Dirac field using the Weyl metric given by (1.7):

$$u_{,11} + u_{,1}/r = C^2 e^{-2u}, \tag{2.1}$$

$$k_{,11} - k_{,1}/r - u_{,11} - u_{,1}/r + 2u_{,1}{}^2 = -C^2 e^{-2u}, \tag{2.2}$$

$$k_{,11} + k_{,1}/r - u_{,11} - u_{,1}/r = C^2 e^{-2u}. \tag{2.3}$$

The Dirac field ψ, which is, in this case, a function of r only, is given by

$$\psi = (1/\sqrt{r}) e^{(u-k)/2}\psi_0, \tag{2.4}$$

ψ_0 being an arbitrary constant spinor, and the only nonvanishing components of energy-momentum tensor and electromagnetic field tensor are, respectively, given by

$$T_{20} = \tfrac{1}{4}e^{u-k}(2u_{,1} - k_{,1})\psi^\dagger \gamma' \gamma^2 \gamma^0 \psi, \tag{2.5}$$

$$T_{30} = \tfrac{1}{4}e^{u-k}(2u_{,1} - 1/r)\psi^\dagger \gamma' \gamma^3 \gamma^0 \psi, \tag{2.6}$$

and

$$F_{02} = C_1 e^{-k} \quad F_{31} = C_2 e^{-k}, \tag{2.7}$$

where C_1 and C_2 are constants of integration and $C^2 = 4\pi G(C_1{}^2 + C_2{}^2)$.

Krori, Chaudhury, and Bhattacharjee[3] presented some particular solutions of Eqs. (21)–(2.3); we try to give the general solutions of the same equations.

 0022-2488/86/020568-05$02.50 © 1986 American Institute of Physics

B. Solutions

Adding Eqs. (2.1) and (2.2) one gets

$$k_{,11} - k_{,1}/r + 2u_{,1}{}^2 = 0. \tag{2.8}$$

Again, adding (2.1) and (2.3) and using (2.1) one obtains

$$k_{,11} + k_{,1}/r = 2u_{,11} + 2u_{,1}/r. \tag{2.9}$$

Assuming $\ln r = x$ and $u - x = v$, Eq. (2.1) reduces to

$$v_{xx} = C^2 e^{-2v}. \tag{2.10}$$

Integrating (2.9) twice one gets

$$k = 2u + A \ln r + \ln B, \tag{2.11}$$

where A and B are constants of integration. Integrating Eq. (2.10) twice and simplifying one can obtain

$$e^{2u} = [Cr^{1+D}/2DE^D + (CE^D/2D)r^{1-D}]^2 \tag{2.12}$$

and from (2.11),

$$e^{2k} = [Cr^{1+D}/2DE^D + (CE^D/2D)r^{1-D}]^4 B^2 r^{2A}, \tag{2.13}$$

where D and E are constants of integration.

Putting (2.12) and (2.13) into Eqs. (2.1)–(2.3) one can check that all the equations are satisfied. Thus the complete set of solutions of Eqs. (2.1)–(2.3) are given by (2.12) and (2.13).[6]

If one puts $B = 1, A = 0$ and $CE^D = \sqrt{2}\,D$ in (2.12) and (2.13) one can get the particular solutions of Eqs. (2.1)–(2.3) as obtained by Krori, Chaudhury, and Bhattacharjee.[3]

III. MASSLESS NEUTRINO WITH TIME-INDEPENDENT DIRAC FIELD

A. Equations

The Einstein–Dirac–Maxwell field equations derived by Krori, Chaudhury, and Bhattacharjee[3] for a massless neutrino with time-independent Dirac field using the plane symmetric metric given by (1.8) are given by

$$e^{-2u}(u_{,11} + 2u_{,1}v_{,1}) = (K/2)(C_1^2 + C_2^2)e^{-4v}, \tag{3.1}$$

$$e^{-2u}(u_{,11} - 2v_{,11} - 2u_{,1}v_{,1} + 2v_{,1}{}^2) = (K/2)(C_1^2 + C_2^2)e^{-4b}, \tag{3.2}$$

$$e^{-2u}(v_{,11} + 2v_{,1}{}^2) = -(K/2)(C_1^2 + C_2^2)e^{-4v}. \tag{3.3}$$

Here the Dirac field ψ, which is in this case a function of x only, is given by

$$\psi = e^{-(v+u/2)}\psi_0, \tag{3.4}$$

ψ_0 being an arbitrary constant spinor, and the only nonvanishing components of the energy momentum tensor and the electromagnetic field tensor are, respectively, given by

$$T_{20} = \tfrac{1}{4}e^{-u}(v_{,1} - u_{,1})\psi^\dagger\gamma'\gamma^2\gamma^0\psi, \tag{3.5}$$

$$T_{30} = \tfrac{1}{4}e^{-u}(v_{,1} - u_{,1})\psi^\dagger\gamma'\gamma^3\gamma^0\psi, \tag{3.6}$$

and

$$F_{20} = C_1 e^{-2v}, \quad F_{23} = C_2 e^{-2v}, \tag{3.7}$$

where C_1 and C_2 are constants of integration and $K = 8\pi G$.

Krori, Chaudhury, and Bhattacharjee[3] obtained a particular solution of Eqs. (3.1)–(3.3)

We seek a complete set of solutions of the same equations.

B. Solutions

Subtracting Eq. (3.1) from Eq. (3.2),

$$v_{,11} - 2u_{,1}v_{,1} + v_{,1}{}^2 = 0. \tag{3.8}$$

Adding Eq. (3.2) and Eq. (3.3),

$$u_{,11} + 3v_{,11} - 2u_{,1}v_{,1} + 4_{v,1}{}^2 = 0. \tag{3.9}$$

Let $u = u(v)$, then Eqs. (3.8) and (3.9) reduce to

$$v_{,11} + (1 - 2u_v)v_{,1}{}^2 = 0, \tag{3.10}$$

and

$$(u_v + 3)v_{,11} + (u_{vv} - 2u_v + 4) + v_{,1}{}^2 = 0. \tag{3.11}$$

Eliminating $v_{,11}/v_{,1}{}^2$ between (3.10) and (3.11),

$$u_{vv} + 2u_v{}^2 + 3u_v + 1 = 0. \tag{3.12}$$

In view of Eq. (3.12) we note that if $u_{vv} = 0$ then $u_v = -1$ or $-\frac{1}{2}$. One can consider the following three cases.

Case 1: If $u_v = -1$, then Eq. (3.10) gives on integration

$$e^{2v} = (3C_1x + C_2)^{2/3}, \tag{3.13}$$

$$\therefore e^{2u} = C_3(3C_1x + C_2)^{-2/3}, \tag{3.14}$$

where C_1, C_2, C_3 are constants of integration. The solutions given by Eqs. (3.13) and (3.14) are obtained by Krori, Chaudhury, and Bhattacharjee.[3]

Case 2: If $u_v = -\frac{1}{2}$ then the solutions of Eqs. (3.1)–(3.3) are given by

$$e^{2v} = 2C_4x + C_5, \tag{3.15}$$

$$e^{4u} = C_6(2C_4x + C_5)^{-1}, \tag{3.16}$$

where C_4, C_5, and C_6 are constants of integration.

Case 3: Let $u_{vv} \neq 0$, then integrating Eq. (3.12) twice, one can obtain

$$2u = -v + \ln(1 - e^{m-v}) + m_1, \tag{3.17}$$

where m and m_1 are constants of integration.

Inserting the value of u from Eq. (3.17) in Eq. (3.10) and integrating twice, one can get

$$m_2x + m_3 = \tfrac{1}{2}(e^v - e^m)^2 + 2e^m(e^v - e^m) + e^{2m}\ln(e^v - e^m), \tag{3.18}$$

where m_2 and m_3 are constants of integration.

Putting Eqs. (3.17) and (3.18) into Eqs. (3.1)–(3.3) one can check that all the equations are satisfied. Hence in this case the complete set of solutions of Eqs. (3.1)–(3.3) are given by (3.17) and (3.18).

IV. MASSLESS NEUTRINO WITH TIME-DEPENDENT DIRAC FIELD

A. Equations

Krori, Chaudhury, and Bhattacharjee[3] obtained the following Einstein–Dirac–Maxwell field equations for a massless neutrino with the time-dependent Dirac field using the Einstein–Rosen metric given by (1.9):

$$e^{2\beta-2\alpha}(\alpha_{,11} - \beta_{,11} - \beta_{,1}/r + \alpha_{,1}/r) = C^2e^{-2\alpha} - 8\pi GT_{00}, \tag{4.1}$$

$$e^{2\beta-2\alpha}(\alpha_{,11} - \beta_{,11} + 2\beta_{,1}{}^2 - \beta_{,1}/r - \alpha_{,1}/r) = -C^2e^{-2\alpha} + 8\pi GT_{11}, \tag{4.2}$$

$$e^{2\beta-2\alpha}(\beta_{,11}+\beta_{,1}/r)=C^2e^{-2\alpha}. \tag{4.3}$$

The Dirac field ψ, which is, in this case, a function of r and t, is given by

$$\psi=(1/\sqrt{r})e^{(\beta-\alpha)/2}e^{i\omega(\gamma'\gamma^0-t)}\psi_c, \tag{4.4}$$

ψ_c being an aribtrary constant spinor, and the nonvanishing components of energy-momentum tensor and electromagnetic field tensor are, respectively, given by

$$T_{00}=T_{11}=-(e^{2\beta-2\alpha}/r)\omega\psi_c{}^*\psi_c, \tag{4.5}$$

$$T_{01}=T_{10}=\tfrac{1}{4}e^{\beta-\alpha}\psi^\dagger(4i\omega\gamma')\psi, \tag{4.6}$$

$$T_{02}=T_{20}=\tfrac{1}{4}e^{\beta-\alpha}[\psi^\dagger(2i\omega\gamma^2)\psi + (\alpha_{,1}-2\beta_{,1})\psi^\dagger\gamma'\gamma^2\gamma^0\psi], \tag{4.7}$$

$$T_{03}=T_{30}=\tfrac{1}{4}e^{\beta-\alpha}[\psi^\dagger(2i\omega\gamma^3)\psi + (\alpha_{,1}-1/r)\psi^\dagger\gamma'\gamma^3\gamma^0\psi], \tag{4.8}$$

and

$$F_{02}=C_1e^{-2\beta},\quad F_{31}=C_2e^{-2\beta}, \tag{4.9}$$

where C_1, C_2 are constants of integration, $C=4\pi G(C_1{}^2+C_2{}^2)$, and ω is a positive real number.

Krori, Chaudhury and Bhattacharjee[3] obtain some particular solutions of the equations. We try to seek general solutions of the same equations.

B. Solutions

Adding Eqs. (4.1) and (4.2),

$$\alpha_{,11}-\beta_{,11}+\beta_{,1}{}^2-\beta_{,1}/r=0. \tag{4.10}$$

We note that Eq. (4.3) is identical to Eq. (2.1) and Eq. (4.10) is identical to the result obtained by adding Eqs. (2.2) and (2.3). Thus the solutions for α and β will be similar to those for k and u, respectively, where u and k, are, respectively, given by (2.12) and (2.13).

V. MASSLESS NEUTRINO WITH TIME-DEPENDENT DIRAC FIELD

A. Equations

Using the plane symmetric metric given by (1.8), Krori, Chaudhury, and Bhattacharjee[3] obtained the following Einstein–Dirac–Maxwell field equations for a massless neutrino with time-dependent Dirac field:

$$e^{-2u}(u_{,11}+2u_{,1}v_{,1})=(K/2)(C_1{}^2+C_2{}^2)e^{-4v}+KT_{00}, \tag{5.1}$$

$$e^{-2u}(u_{,11}+2v_{,11}-2u_{,1}v_{,1}+2v_{,1}{}^2) =(K/2)(C_1{}^2+C_2{}^2)e^{-4v}-KT_{11}, \tag{5.2}$$

$$e^{-2u}(v_{,11}+2v_{,1}{}^2)=-(K/2)(C_1{}^2+C_2{}^2)e^{-4v}. \tag{5.3}$$

The Dirac field $\psi(x,t)$ is given by

$$\psi=e^{-(v+u/2)}e^{i\omega(\gamma'\gamma^0x-t)}\psi_{v_0}, \tag{5.4}$$

where

$$\psi_{v_0}=\alpha_0\begin{bmatrix}1\\ \pm 1\\ i\\ \pm i\end{bmatrix},$$

α_0 being an arbitrary constant.

Krori, Chaudhury, and Bhattacharjee obtained some particular solutions of Eqs. (5.1)–(5.3). In the present paper, we shall try to seek the complete set of solutions.

B. Solutions

Adding (5.1) and (5.2),

$$e^{-2u}(u_{,11}+v_{,11}+v_{,1}{}^2)=(K/2)(C_1{}^2+C_2{}^2)e^{-4v}. \tag{5.5}$$

Equation (5.5) is identical to the equation obtained by adding (3.1) and (3.2) and Eq. (5.3) is identical to Eq. (3.3). Hence the solutions of Eqs. (5.1)–(5.3) are the solutions of Eqs. (3.1)–(3.3).

VI. MASSIVE NEUTRINO

A. Equations

Using the Marder metric given by (1.10), Krori, Chaudhury, and Bhattacharjee[3] obtained the following field equations for a massive neutrino:

$$\alpha_{,1}\nu_{,1}+\alpha_{,1}/r+\nu_{,1}/r-\beta_{,1}{}^2-2\beta_{,1}\nu_{,1} =(8\pi Gm\lambda/r)e^{\beta-\alpha-\nu}, \tag{6.1}$$

$$\nu_{,11}-\alpha_{,1}\nu_{,1}+2\beta_{,1}\nu_{,1}+\beta_{,1}{}^2+\nu_{,1}{}^2-\alpha_{,1}/r+\nu_{,1}/r=0, \tag{6.2}$$

$$\alpha_{,11}-2\beta_{,11}+\beta_{,1}{}^2-2\beta_{,1}/r=0, \tag{6.3}$$

$$\alpha_{,11}+\nu_{,11}+2\beta_{,1}\nu_{,1}+\beta_{,1}{}^2+\nu_{,1}{}^2=0, \tag{6.4}$$

with

$$\psi=\frac{1}{\sqrt{r}}\exp\left(-m\gamma'\int e^{\alpha-\beta}\,dr\right)e^{(\beta-\alpha-\nu)/2}, \tag{6.5}$$

and the nonvanishing components of energy-momentum tensors are

$$T_{11}=(m\lambda/r)e^{\beta-\alpha-\nu}, \tag{6.6}$$

where

$$\lambda=\psi_c{}^\dagger\psi_c,$$

$$T_{20}=\tfrac{1}{4}e^{\beta-\alpha}(\alpha_{,1}-2\beta_{,1}-\nu_{,1})\psi^\dagger\gamma'\gamma^2\gamma^0\psi, \tag{6.7}$$

$$T_{30}=\tfrac{1}{4}e^{\beta-\alpha}(\alpha_{,1}-l/r)\psi^\dagger\gamma'\gamma^3\gamma^0\psi. \tag{6.8}$$

It is to be noted that the solutions obtained by Krori, Chaudhury, and Bhattacharjee[3] are not correct. We try to present a complete set of solutions of the same equations.

B. Solutions

Equations (6.1)–(6.4) can be easily reduced to

$$u_{,1}w_{,1}-2v_{,1}w_{,1}-v_{,1}{}^2=Me^{v-u-w}, \tag{6.9}$$

$$w_{,11}-u_{,1}w_{,1}+(v_{,1}+w_{,1})^2=0, \tag{6.10}$$

$$u_{,11}-2v_{,11}+v_{,1}{}^2=0, \tag{6.11}$$

$$u_{,11}+w_{,1}+(v_{,1}+w_{,1})^2=0, \tag{6.12}$$

where

$$u=\alpha-\ln r,\quad v=\beta-\ln r,$$

$$w=\nu+\ln r,\quad M=8\pi Gm\lambda,$$

say. Subtracting (6.10) from (6.12) and integrating, we get

$$e^w u_{,1}=A, \tag{6.13}$$

where A is a constant of integration.

Adding (6.9) and (6.10) and using (6.13), one can obtain

$$(A/u_{,1})_{,11} = Me^{v-u}. \tag{6.14}$$

Using (6.11), (6.13), and (6.14) in Eq. (6.9), one can get, after some calculations,

$$u_{,11}/u_{,1}^2 - 2v_{,1}/u_{,1} = B_1, \tag{6.15}$$

where B_1 is a constant.

We can also obtain from (6.9) using (6.13) that

$$\frac{u_{,1} - 2v_{,1}}{u_{,1}^2} u_{,11} + \frac{v_{,1}^2}{u_{,1}} = -\frac{M}{A} e^{v-u}. \tag{6.16}$$

Let $v = v(u)$, then Eqs. (6.11), (6.15), and (6.16) can be written as

$$(1 - 2v_u)u_{,11} + (v_u^2 - 2v_{uu})u_{,1}^2 = 0, \tag{6.17}$$

$$u_{,11} - (2v + B_1)u_{,1}^2 = 0, \tag{6.18}$$

$$(1 - 2v_u)(u_{,11}/u_{,1}) + v_u^2 u_{,1} = -(M/A)e^{v-u}. \tag{6.19}$$

Eliminating $u_{,11}/u_{,1}^2$ between (6.17) and (6.18), one gets

$$2v_{uu} + 3v_u^2 + 2(B_1 - 1)v_u - B_1 = 0. \tag{6.20}$$

Case 1: Let $v_{uu} = 0$, then $v_u = \text{const}$, 1 say. Then one can easily obtain the following solutions of Eqs. (6.1)–(6.4):

$$\alpha = \ln r + (1/B_3)\ln(B_4 r + B_5), \tag{6.21}$$

$$\beta = \ln r + (1/B_3)\ln(B_4 r + B_5) + B_2, \tag{6.22}$$

and

$$\nu = -\ln r + \ln(B_4 r + B_5) + \ln B_6, \tag{6.23}$$

where B_2, B_3, B_4, B_5, and B_6 are all constants of integration.

Case 2: Let $v_{uu} \neq 0$. Then (6.20) gives, on integration twice,

$$2v = Au + B\ln(p - e^{qu}) + F, \tag{6.24}$$

where A, B, P, q and F are all constants.

Using (6.24) in (6.18) and integrating, we get

$$\gamma = \int \frac{e^{gu}\,du}{(p - e^{qu})^f}, \tag{6.25}$$

where f, g are constants. Also, from (6.13),

$$e^w = A/u_{,1}. \tag{6.26}$$

Putting the values of u, v, w from (6.24)–(6.26) into Eqs. (6.9)–(6.12) one can check that all the equations are satisfied. Thus the solutions of Eqs. (6.9)–(6.12) are given by (6.24)–(6.26). Hence the solution of Eqs. (6.1)–(6.4) are given by

$$\alpha = u + \ln r, \tag{6.27}$$

$$\beta = v + \ln r, \tag{6.28}$$

$$\nu = w - \ln r, \tag{6.29}$$

where u, v, w are given by (6.24)–(6.26).

VII. MASSIVE NEUTRINO

A. Equations

Using the plane symmetric metric given by (1.8), Krori, Chaudhury, and Bhattacharjee[3] derived the following field equations for a massive neutrino with time-dependent Dirac field:

$$2u_{,1}v_{,1} + v_{,1} = -8\pi\psi_0^\dagger\psi_0 m e^{u-2v}, \tag{7.1}$$

$$u_{,11} + v_{,11} + v_{,1}^2 = 0, \tag{7.2}$$

$$2v_{,11} - 2u_{,1}v_{,1} + 3v_{,1}^2 = 0, \tag{7.3}$$

with Dirac field ψ given by

$$\psi = e^{-(v + u/2)}\psi_0 \tag{7.4}$$

and the only nonvanishing components of the electromagnetic field tensor are given by

$$T_{11} = \psi_0^\dagger\psi_0 m e^{-u-2v}. \tag{7.5}$$

B. Solutions

Let $u = u(v)$, then Eqs. (7.1)–(7.3) can be written as

$$(u_v + 1)v_{,11} + (u_{vv} + 1)v_{,1}^2 = 0, \tag{7.6}$$

$$2v_{,11} + (3 - 2u_v)v_{,1}^2 = 0. \tag{7.7}$$

Eliminating $v_{,11}/v_{,1}^2$ between (7.6) and (7.7) one can obtain

$$2u_{vv} + 2u_v^2 - u_v - 1 = 0. \tag{7.8}$$

In view of Eq. (7.8), we note that if $u_{vv} = 0$ then $u_v = 1$ or $-\frac{1}{2}$. One can consider the following three cases.

Case 1: Let $u_v = 1$, then Eq. (7.7) gives on integration twice,

$$v = 2\ln(p_2 x + p_3), \tag{7.9}$$

$$\therefore\ u = 2\ln(p_2 x + p_3) + p_1, \tag{7.10}$$

where p_1, p_2, p_3 are constants of integration.

Putting (7.9) and (7.10) into Eqs. (7.1)–(7.3) one can check that all the equations are satisfied. Hence the complete set of solutions of Eqs. (7.1)–(7.3) are given by (7.9) and (7.10).

Putting $p_1 = 0, p_3 = A$ and $p_2 = (\lambda/2)\sqrt{m}$ in (7.9) and (7.10),

$$v = u = 2\ln(A + (\lambda/2)\sqrt{m}x), \tag{7.11}$$

where $\lambda^2 = -\frac{8}{3}\pi\psi_0^\dagger\psi_0$. The particular solution (7.11) was obtained by Krori, Chaudhury, and Bhattacharjee[3].

Case 2: Let $u_v = -\frac{1}{2}$. Then Eq. (7.7) gives, on integration twice,

$$e^{2v} = p_4 x + p_5, \tag{7.12}$$

$$e^{4u} = p_6(p_4 x + p_5)^{-1}, \tag{7.13}$$

where p_4, p_5, and p_6 are constants of integration. In this case, (7.12) and (7.13) are the complete set of solutions of Eqs. (7.1)–(7.3).

Case 3: Let $u_{vv} \neq 0$. Equation (7.8) gives, on integration twice,

$$u = -(v/2) + \ln(e^{(3/2)q} - e^{(3/2)v}) + q_1, \tag{7.14}$$

where q and q_1 are constants of integration. Using (7.14) in (7.7) and integrating twice, one can obtain

$$q_2 x + q_3 = \int \frac{e^{2v}\,dv}{e^{(3/2)q} - e^{(3/2)v}}, \tag{7.15}$$

where q_2 and q_3 are constants of integration.

Putting (7.14) and (7.15) into Eqs. (7.1)–(7.3), one can check that all the equations are satisfied. Hence in this case, the complete set of equations of (7.1)–(7.3) are given by (7.14) and (7.15).

VIII. CONCLUSION

In summary the solutions of Einstein–Dirac–Maxwell field equations obtained in this paper are as follows.

(1) A massless neutrino with time-independent Dirac field.

(1a) For the Weyl metric (1.7) the solutions are given by (2.12) and (2.13).

(1b) For the plane symmetric metric given by (1.8), one gets three classes of solutions. The first is given by (3.13) and (3.14). The second is given by (3.15) and (3.16). The third is given by (3.17) and (3.18).

(2) A massless neutrino with time-dependent Dirac field.

(2a) For the Einstein–Rosen metric (1.9), the solutions for α and β will be similar to those for k and u, respectively, where u and k are given by (2.12) and (2.13), respectively.

(2b) For the plane symmetric metric (1.8), the solutions are similar to those discussed earlier in (1b).

(3) A massive neutrino.

(3a) For the Marder metric (1.10), one gets two classes of solutions. The first one is given by (6.21), (6.22), and (6.23). The second is given by

$$\alpha = u + \ln r, \quad \beta = v + \ln r, \quad \nu = w - \ln r,$$

where u, v and w are given by (6.24)–(6.26). It is to be noted that the solutions given by Krori, Chaudhury, and Bhattacharjee[3] are not correct.

(3b) For the plane symmetric metric (1.8), one obtains three classes of solutions. The first one is given by (7.9) and (7.10). The second is given by (7.12) and (7.13). The third is given by (7.14) and (7.15).

[1] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976), Appendix A_2.

[2] D. Brill and J. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).

[3] K. D. Krori, T. Chaudhury, and R. Bhattacharjee, Phys. Rev. D **25**, 1492 (1982).

[4] A. Einstein and N. J. Rosen, J. Franklin Inst. **223**, 43 (1937).

[5] L. Marder, Proc. R. Soc. London, Ser. A **244**, 524 (1958).

[6] Equations (2.12) and (2.13) can also be obtained from Eqs. (2.1) and (2.3) only. Thus Eq. (2.2) is really superfluous.

Symmetry groups of state vectors in canonical quantum gravity

Donald M. Witt
Department of Physics, University of Wisconsin, Milwaukee, Wisconsin 53201

(Received 25 February 1985; accepted for publication 20 September 1985)

In canonical quantum gravity, the diffeomorphisms of an asymptotically flat hypersurface S, not connected to the identity, but trivial at infinity, can act nontrivially on the quantum state space. Because state vectors are invariant under diffeomorphisms that are connected to the identity, the group of inequivalent diffeomorphisms is a symmetry group of states associated with S. This group is the zeroth homotopy group of the group of diffeomorphisms fixing a frame of infinity on S. It is calculated for all hypersurfaces of the form $S = S^3/G$-point, where the removed point is thought of as infinity on S and the symmetry group S is the zeroth homotopy group of the group of diffeomorphisms of S^3/G fixing a point and frame, denoted $\pi_0 \operatorname{Diff}_F(S^3/G)$. Before calculating $\pi_0 \operatorname{Diff}_F(S^3/G)$, it is necessary to find π_0 of the group of diffeomorphisms. Once $\pi_0 \operatorname{Diff}(S^3/G)$ is known, $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ is calculated using a fiber bundle involving $\operatorname{Diff}(S^3/G)$, $\operatorname{Diff}_{x_0}(S^3/G)$, and S^3/G. Finally, a fiber bundle involving $\operatorname{Diff}_F(S^3/G)$, $\operatorname{Diff}(S^3/G)$, and the bundle of frames over S^3/G is used along with $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ to calculate $\pi_0 \operatorname{Diff}_F(S^3/G)$. The groups $\pi_0 \operatorname{Diff}_F(S^3/G)$ are comprised of $SU(2)$ coverings of $SO(3)$ crystallographic groups, the product of these with a cyclic group, cyclic groups, and the product of two cyclic groups.

I. INTRODUCTION

In canonically quantized Yang–Mills theory, the state space is invariant only under asymptotically trivial gauge transformations. Moreover, the classes of inequivalent gauge transformations on a hypersurface with trivial topology are isomorphic to the group of additive integers $\mathbf{Z}$ (see Refs. 1 and 2). Since these groups are Abelian, they only have one-dimensional irreducible representations, each of which is fixed by a single θ:

$$g_n \Psi = e^{in\theta} \Psi ,$$

where Ψ is a quantum state vector and g_n a gauge transformation of degree n. These are Yang–Mills θ-states. Even in the case of nontrivial topologies these groups are Abelian.[3]

A counterpart of θ-states in canonical quantum gravity are irreducible representations of the groups of inequivalent diffeomorphism.[4–6] This is easy to demonstrate using the momentum constraint $D_b \pi^{ab} = 0$ in the Schrödinger picture, where $\hat{\pi}^{ab} = (\hbar/i)(\delta/\delta g_{ab})$. Multiplying the constraint $D_b \hat{\pi}^{ab} \Psi = 0$ by an arbitrary test vector field ξ with compact support and integrating it over an open hypersurface S yields[7]

$$\int_S \xi_a D_b \hat{\pi}^{ab} \Psi \, d^3x = 0 .$$

Integrating by parts and adding the resulting integral to itself yields

$$0 = \int_S D_b \xi_a \hat{\pi}^{ab} \Psi \, d^3x + \int_S D_a \xi_b \hat{\pi}^{ab} \Psi \, d^3x$$

$$= \int_S D_{(a} \xi_{b)} \hat{\pi}^{ab} \Psi \, d^3x = \frac{\hbar}{i} \frac{d\Psi}{d\lambda} (g_{ab} + \lambda D_{(a} \xi_{b)}) \big|_{\lambda = 0} .$$

Recalling the properties of Lie derivatives, it follows that

$$0 = \frac{d\Psi}{d\lambda} (g_{ab} + \lambda D_{(a} \xi_{b)}) \big|_{\lambda = 0} = \frac{d\Psi}{d\lambda} (T_\lambda^* g_{ab}) \big|_{\lambda = 0} ,$$

where T_λ^* is the induced action of any path of diffeomorphisms T_λ trivial in a neighborhood of infinity and having ξ as a tangent vector at $\lambda = 0$. Hence the constraint is equivalent to $(d\Psi/d\lambda) \circ T_\lambda^* |_{\lambda = 0} = 0$, for T_λ as above. Since this is true for all such diffeomorphisms, it follows that $(d\Psi/d\lambda) \circ T_\lambda^* = 0$. The above set of diffeomorphisms is the identity component of the group of diffeomorphisms trivial in neighborhood of infinity, $\operatorname{Diff}_{N^\infty \mathrm{id}}(S)$. Therefore, $\Psi \circ T^* = \Psi$ for all T in $\operatorname{Diff}_{N^\infty \mathrm{id}}(S)$.[8] If S is closed,[9] then no boundary term appears in the above integration by parts so all vector fields can be used, and $\Psi \circ T^* = \Psi$ for all T in the identity component of the full diffeomorphism group, $\operatorname{Diff}_{\mathrm{id}}(S)$. The groups, which can act nontrivially on the quantum state space, are $\operatorname{Diff}_{N^\infty}(S)/\operatorname{Diff}_{N^\infty \mathrm{id}}(S)$ or $\operatorname{Diff}(S)/\operatorname{Diff}_{\mathrm{id}}(S)$, for S open or closed, respectively. These groups are written $\pi_0 \operatorname{Diff}_{N^\infty}(S)$ and $\pi_0 \operatorname{Diff}(S)$. The group $\pi_0 \operatorname{Diff}_{N^\infty}(S)$ is isomorphic to the zeroth homotopy group of $\operatorname{Diff}_{F^\infty}(S)$ (diffeomorphisms fixing a frame at infinity). Therefore, gravity's "θ-states" are given by irreducible representations of $\pi_0 \operatorname{Diff}_{F^\infty}(S)$ and $\pi_0 \operatorname{Diff}(S)$ depending on S. If the open hypersurface S is obtained from a closed three-manifold M^3 by removing a point x_0 as in the case of an asymptotically flat hypersurface, then $\pi_0 \operatorname{Diff}_{F^\infty}(S)$ is equal to $\pi_0 \operatorname{Diff}_{Fx_0}(M^3)$ (the zeroth homotopy of the group of diffeomorphisms of M^3 fixing a frame F at the point x_0). [From this point on $\pi_0 \operatorname{Diff}_{Fx_0}(M^3)$ will be denoted $\pi_0 \operatorname{Diff}_F(M^3)$.]

Even if the hypersurface S is closed, the frame fixing diffeomorphisms are of interest. Suppose S is closed and $S = M \sharp N$, where N is thought of as a generic environment. Then the group $\pi_0 \operatorname{Diff}(S)$ includes as a subgroup $\pi_0 \operatorname{Diff}_F(M)$. A generic "environment" N around M prevents one from deforming to the identity in $M \sharp N$ any diffeomorphism that cannot be deformed to the identity in $\operatorname{Diff}_F(M)$. In representing an isolated system by an asymptotically flat space-time one imposes the conditions that M is far from the interface between M and N, and M is small enough to neglect the large-scale curvature of the back-

 0022-2488/86/020573-20$02.50 © 1986 American Institute of Physics

ground. If these conditions fail, then symmetries of the environment cannot be ignored and the full symmetry group is $\pi_0 \operatorname{Diff}(M\sharp N)$, but $\pi_0 \operatorname{Diff}_F(M)$ is a subgroup for N generic.

One major difference between the Yang–Mills theory and gravity is that the symmetry groups $\pi_0 \operatorname{Diff}_F(M^3)$ are not necessarily Abelian, but for example can be SU(2) coverings of SO(3) crystal groups which are non-Abelian and therefore have irreducible representations of dimension greater than 1. Some of the results were given in an earlier paper[10] along with an outline of techniques. In the present paper, calculations and results are given for $M^3 = S^3/G$, where G is a finite group acting freely on S^3, along with a correction to a table appearing in the earlier paper.

The isometry group of S^3/G with the standard metric is calculated using elementary group theory. The calculations give the explicit action of $\operatorname{Isom}(S^3/G)$ on $\pi_1(S^3/G,x_0)$, which is used in later calculations. The technique used to calculate $\operatorname{Isom}(S^3/G)$ uses the facts that S^3 covers S^3/G, $\operatorname{Isom}(S^3) = \mathrm{O}(4)$, and $\mathrm{SU}(2)\times\mathrm{SU}(2)$ covers SO(4). After the group $\operatorname{Isom}(S^3/G)$ is calculated, $\pi_0 \operatorname{Isom}(S^3/G)$ and the topology of $\operatorname{Isom}(S^3/G)$ are given. These results can be found in Tables I and II.

The groups $\pi_0 \operatorname{Diff}(S^3/G)$ are found using the recent work of Hatcher, and the calculations of $\pi_0 \operatorname{Isom}(S^3/G)$. Hatcher has proven the Smale conjecture,[11] the statement that O(4) and $\operatorname{Diff}(S^3)$ have the same homotopy type.[12] He also proposes a generalization of the Smale conjecture that in general $\operatorname{Isom}(S^3/G)$ and $\operatorname{Diff}(S^3/G)$ have the same homotopy type.[13] This has been proven in several cases,[14] and the weaker result that $\pi_0 \operatorname{Isom}(S^3/G) = \pi_0 \operatorname{Diff}(S^3/G)$ has been proven in almost all cases.[15–17] Assuming Hatcher's conjecture, the groups $\pi_0 \operatorname{Diff}(S^3/G)$ are found.

Let $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ be the zeroth homotopy group of the group of diffeomorphisms fixing the point x_0 on S^3/G. Unlike $\pi_0 \operatorname{Diff}(S^3/G)$ not all elements of $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ for G noncyclic can be realized by isometries. A fiber bundle having total space $\operatorname{Diff}(S^3/G)$, base space S^3/G, and fiber $\operatorname{Diff}_{x_0}(S^3/G)$ is used to obtain an exact sequence involving $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ and $\pi_0 \operatorname{Diff}(S^3/G)$. The group $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ is found using the exact sequence combined with the action of $\pi_0 \operatorname{Diff}_{x_0}(S^3/G)$ on $\pi_1(S^3/G,x_0)$.

A fiber bundle with fiber $\operatorname{Diff}_F(S^3/G)$, total space $\operatorname{Diff}^+(S^3/G)$ (orientation preserving diffeomorphisms of S^3/G), and base space $F^+(S^3/G)$ (the bundle of oriented frames over S^3/G) is used to obtain an exact sequence involving $\pi_0 \operatorname{Diff}_F(S^3/G)$ and $\pi_0 \operatorname{Diff}^+(S^3/G)$. The group $\pi_0 \operatorname{Diff}_F(S^3/G)$ is found using this exact sequence and the group $\pi_0 \operatorname{Diff}^+_{x_0}(S^3/G)$. For G noncyclic, no elements of $\pi_0 \operatorname{Diff}_F(S^3/G)$ can be realized by isometries. Hence the structure of $\pi_0 \operatorname{Diff}_F(S^3/G)$ is very different compared to $\pi_0 \operatorname{Diff}^+(S^3/G)$ (see Table IV). For G noncyclic, the group $\pi_0 \operatorname{Diff}_F(S^3/G)$ is an SU(2) double cover of $\pi_0 \operatorname{Diff}^+_{x_0}(S^3/G)$. This is related to the fact that the 2π rotation parallel to a two-sphere[18] is nontrivial in $\pi_0 \operatorname{Diff}_F(S^3/G)$ for G a noncyclic group (see Sec. VI). If G is a cyclic group, the 2π rotation is trivial and the groups $\pi_0 \operatorname{Diff}^+(S^3/G)$, $\pi_0 \operatorname{Diff}^+_{x_0}(S^3/G)$, and $\pi_0 \operatorname{Diff}_F(S^3/G)$ are isomorphic to each other.

II. PRELIMINARIES

The first theorem is a combination of the standard Noether isomorphism theorems. It characterizes the subgroups of a quotient group, and is very useful when calculating the quotient group of two quotient groups.

Theorem 2.1: If K is a normal subgroup of G, $K \trianglelefteq G$, then all subgroups of G/K are of the form H/K where $K<H<G$. Moreover, $N/K \trianglelefteq G/K$ if and only if $N \trianglelefteq G$, also $(G/K)/(N/K) \approx G/N$.

Proof: See any standard text on groups, for example, Hungerford.[19]

The finite subgroups of SO(3) and SU(2) will be encountered in our calculations. The finite subgroups of SO(3) are cyclic, dihedral, tetrahedral, octahedral, and iscosahedral groups. Their presentation in terms of generators and relations are

$$\mathbf{Z}_n = \langle x: x^n = 1\rangle ,$$
$$D_{2n} = \langle x,y: x^2 = y^n = (xy)^n = 1\rangle ,$$
$$T = \langle x,y: x^2 = (xy)^3 = y^3 = 1\rangle ,$$
$$O = \langle x,y: x^2 = (xy)^3 = y^4 = 1\rangle ,$$
$$I = \langle x,y: x^2 = (xy)^3 = y^5 = 1\rangle .$$

The finite subgroups SU(2) come from the double covering of SO(3) by SU(2). These groups are the cyclic, binary dihedral, and binary polyhedral groups. Their presentations are

$$\mathbf{Z}_n = \langle x: x^n = 1\rangle ,$$
$$D^*_{4n} = \langle x,y: x^2 = (xy)^2 = y^n\rangle ,$$
$$T^* = \langle x,y: x^2 = (xy)^3 = y^3, x^4 = 1\rangle ,$$
$$O^* = \langle x,y: x^2 = (xy)^3 = y^4, x^4 = 1\rangle ,$$
$$I^* = \langle x,y: x^2 = (xy)^3 = y^5, x^4 = 1\rangle .$$

A useful property of finite subgroups of both SO(3) and SU(2) is the following: Given H_1 and H_2 finite subgroups of $G = \mathrm{SO}(3)$ or SU(2) with H_1 isomorphic to H_2, then H_1 is conjugate to H_2.

The other finite groups encountered are the finite subgroups of SO(4). Here SO(4) is doubled covered by the group $\mathrm{SU}(2)\times\mathrm{SU}(2)$, where the projection map P is given by $P(q_1,q_2)q = q_1 q q_2^*$, where $q\in\mathbf{H}$ (see Ref. 20) and $(q_1,q_2)\in\mathrm{SU}(2)\times\mathrm{SU}(2)$ (see Ref. 21). The kernel of P is $\mathbf{Z}_2 = \{\pm(1,1)\}$. It follows that $\mathrm{SO}(4) = \mathrm{SU}(2)\times\mathrm{SU}(2)/\mathbf{Z}_2$. Hence the subgroups of SO(4) can be found using Theorem 1.1.

The following theorem gives us all the finite groups that act freely on S^3.

Theorem 2.2: The only finite groups that can act freely on S^3 are $\mathbf{Z}_n$, D^*_{4n}, $D'_{2^k(2n+1)}$, T^*, $T'_{8\cdot 3^k}$, O^*, I^*, and the direct product of any of these with a cyclic group of relatively prime order.

Proof: See Thomas.[22]

The groups $D'_{2^k(2n+1)}$ and $T'_{8\cdot 3^k}$ have presentations

$$D'_{2^k(2n+1)} = \langle x,y: x^{2^k} = 1,\ y^{2n+1} = 1,\ xyx^{-1} = y^{-1}\rangle,$$
$$k>3,\quad n>1$$
$$T'_{8\cdot 3^k} = \langle x,y,z: x^2 = (xy^2 = y^2,\ zxz^{-1} = y,$$
$$zyz^{-1} = xy,\ z^{3^k} = 1\rangle,\quad k>2 .$$

The groups in Theorem 2.2 are all isomorphic to subgroups of SO(4). A remaining question is whether S^3/G is a spherical space form[23] for every free action of a finite group G. Thomas has proven that the orbit spaces are at least homotopic to spherical space forms, and Rubinstein has proven, in most cases, that any free action gives an orbit space diffeomorphic to a spherical space form.[22] We will therefore only consider spherical space forms.

The following theorem gives the main tool for calculating the homotopy groups of fiber bundles.

Theorem 2.3: Let p: $E \to B$ be a principal bundle with total space E, a connected base space B, and fiber G. The sequence

$$\cdots \to \pi_i(G,g_0) \xrightarrow{i_{\#i}} \pi_i(E,e_0) \xrightarrow{p_{\#i}} \pi_i(B,b_0) \xrightarrow{\partial_{\#i}} \pi_{i-1}(G,g_0)$$

$$\cdots \to \pi_1(B,b_0) \xrightarrow{\partial_{\#1}} \pi_0(G,g_0) \xrightarrow{j_{\#0}} \pi_0(E,e_0) \to 1$$

is exact. Further, it is an exact sequence of groups if $\pi_0(E,e_0)$ is a group.

Proof: See Steenrod,[24] pages 91 and 94.

The last theorem we need is the lifting theorem which is used to prove which maps lift to the covering space.

Theorem 2.4: Let p: $(\tilde{X},\tilde{x}_0) \to (X,x_0)$ be a fibration with unique path lifting. Let Y be a connected space. A necessary and sufficient condition that a map f: $(Y,y_0) \to (X,x_0)$ have a lifting $(Y,y_0) \to (\tilde{X},x_0)$ is that in $\pi_1(X,x_0)$, $f_{\#1}\pi_1(Y,y_0) \leqslant p_{\#1}\pi_1(\tilde{X},\tilde{x}_0)$.

Proof: See Hu.[25]

III. ISOMETRY GROUPS

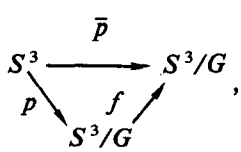
Let S^3/G be a spherical space form with fundamental group G. Since S^3/G is covered by S^3 the isometries of S^3/G lift to elements of O(4). In particular, we will show that $\mathrm{Isom}(S^3/G) = N_{O(4)}(G)/G$, where we use the following definition.

Definition 2.5: Given G and H groups with $H \leqslant G$, the *normalizer* of H in G, denoted by $N_G(H)$, is defined as $\{g \in G \mid gHg^{-1} = H\}$.

When H is a finite group the above definition can be restated in the following form.

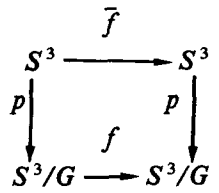
Lemma 2.6: If H is finite and $H \leqslant G$, then $N_G(H) = \{g \in G \mid ghg^{-1} \in H, \ \forall h \in H\}$. Further, $N_G(H) = \{g \in G \mid ghg^{-1} \in H$, for all generators h of $H\}$. (This alternate form of definition is not always true for infinite groups.)

Now let us look at the isometries of S^3/G. Let p: $S^3 \to S^3/G$ be the covering map corresponding to the orthogonal action of G, and let f: $S^3/G \to S^3/G$ be an isometry. If we define $\bar{p}$ by $\bar{p} = fp$,



then $\bar{p}$ is also a covering map.

Since the universal covering space is unique there exists an isometry, $\bar{f}$: $S^3 \to S^3$ such that the diagram



commutes. This means every isometry f of S^3/G lifts to an isometry of S^3.

However, an isometry of S^3 is the lift of an isometry of S^3/G only if a well-defined f is provided by the relation $f([x]) = [\bar{f}(x)]$, that is only when $[\bar{f}(x')] = [\bar{f}(x)]$ for all $x' \in [x]$. Then f is well defined if and only if for all $h \in G$ there exists $h' \in G$ such that $h'\bar{f}(x) = \bar{f}(h(x))$, that is $\bar{f} \in N_{\mathrm{Isom}(S^3)}(G)$. Observe from the definition of f, that $\bar{f}$ induces the same isometry as $\bar{f}'$ if and only if $\bar{f}' = \bar{f}h$ for some $h \in G$. Hence

$$\mathrm{Isom}(S^3/G) = N_{\mathrm{Isom}(S^3)}(G)/G = N_{O(4)}(G)/G,$$

as claimed.

Finally, the group of orientation preserving isometries of S^3 is SO(4), so the group of orientation preserving isometries of S^3/G is $\mathrm{Isom}^+(S^3/G) = N_{SO(4)}(G)/G$, because the lift of an orientation preserving isometry is orientation preserving. (See the proof of Theorem 4.9.)

Let us look at the example $S^3/\mathbf{Z}_2$. If we regard O(4) as 4×4 orthogonal matrices the group $\mathbf{Z}_2$ consists of the matrices ± 1. Because these two diagonal elements commute with every element in O(4), $N_{O(4)}(\mathbf{Z}_2) = O(4)$ and $N_{SO(4)}(\mathbf{Z}_2) = SO(4)$. Hence $\mathrm{Isom}(S^3/\mathbf{Z}_2) = O(4)/\mathbf{Z}_2$ and

$$\mathrm{Isom}^+(S^3/\mathbf{Z}_2) = SO(4)/\mathbf{Z}_2 = SO(3)\times SO(3).$$

The calculations in the above example are very easy to perform since we know $\mathbf{Z}_2$ is normal in both SO(4) and O(4). In general the group will not be normal in O(4) or SO(4), so representing O(4) and SO(4) as 4×4 matrices is not the best way of doing calculations. By using the fact that $SO(4) = SU(2)\times SU(2)/\mathbf{Z}_2$, the calculations of $N_{SO(4)}(G)$ are simplified. This does not help with the $N_{O(4)}(G)$ calculations but, fortunately for us, the only spherical space forms with orientation reversing isometries are lens spaces (see Theorem 4.9). Therefore, $\mathrm{Isom}(S^3/G) = \mathrm{Isom}^+(S^3/G)$ for G noncyclic.

The main and most useful properties of normalizers used in our calculations are summarized in the following theorem. The first half of Theorem 3.1 tells us the normalizer $N_G(H)$ is the maximal subgroup of G containing H as a normal subgroup. The last half gives us the primary tools for doing calculations, namely, the normalizer of products of groups and quotient groups.

Theorem 3.1: Let $N_G(H)$ denote the normalizer of $H \leqslant G$. Then we have the following.

(a) $N_G(H)$ is group.
(b) $H \trianglelefteq N_G(H)$.
(c) $H \trianglelefteq G$ iff $N_G(H) = G$.
(d) If $H \trianglelefteq M \leqslant G$, then $M \leqslant N_G(H)$.
(e) $gN_G(H)g^{-1} = N_G(gHg^{-1})$, $\forall g \in G$.
(f) If $H_i \leqslant G_i$ for $i = 1,2$, then $N_{G_1\times G_2}(H_1\times H_2) = N_{G_1}(H_1)\times N_{G_2}(H_2)$.

(g) If $K \lhd H \lhd G$ and $K \trianglelefteq G$, then $N_G(H)/K = N_{G/K}(H/K)$.

(h) If H is finite, $N_G(H)$ is finite if and only if $C_G(H)$ is finite, where $C_G(H) \equiv \{g \in G \mid gh = hg, \forall h \in H\}$.

Proof: (a) through (d) follow easily from the definitions.

(e) Let $\tilde{g} \in gN_G(H)g^{-1}$. Then $\tilde{g} = gg_1g^{-1}$, where $g_1Hg_1^{-1} = H$. So $\tilde{g}gHg^{-1}\tilde{g}^{-1} = gg_1Hg_1^{-1}g^{-1} = gHg^{-1}$. This implies $\tilde{g} \in N_G(gHg^{-1})$. Hence $gN_G(H)g^{-1} \lhd N_G(gHg^{-1})$. Now, let $g_1 \in N_G(gHg^{-1})$. Then $g_1gHg^{-1}g_1^{-1} = gHg^{-1}$. Since $N_G(H)$ is a group, $g_1^{-1} \in N_G(gHg^{-1})$. It follows that $g_1^{-1}gHg^{-1}g_1 = gHg^{-1}$. Let $\hat{g} = g^{-1}g_1g$. Then

$$\hat{g}H\hat{g}^{-1} = (g^{-1}g_1g)H(g^{-1}g_1g)^{-1} = g^{-1}g_1(gHg^{-1})g_1^{-1}g$$
$$= g^{-1}g_1(g_1^{-1}gHg^{-1}g_1)g_1^{-1}g = g^{-1}gHg^{-1}g = H.$$

Hence $\hat{g} \in N_G(H)$. Since $g_1 = g\hat{g}g^{-1}$ and $\hat{g} \in N_G(H)$ we have $g_1 \in gN_G(H)g^{-1}$. Hence $N_G(gHg^{-1}) \lhd gN_G(H)g^{-1}$.

(f) $(g_1,g_2) \in N_{G_1 \times G_2}(H_1 \times H_2)$
$$\Leftrightarrow (g_1,g_2)(H_1,H_2)(g_1,g_2)^{-1} = (H_1,H_2)$$
$$\Leftrightarrow (g_1H_1g_1^{-1},\ g_2H_2g_2^{-1}) = (H_1,H_2)$$
$$\Leftrightarrow (g_1,g_2) \in N_{G_1}(H_1) \times N_{G_2}(H_2).$$

(g) Let $K \lhd H \lhd G$ and $K \trianglelefteq G$. Clearly, $H/K \lhd G/K$ and $H/K \lhd N_{G/K}(H/K)$. Applying isomorphism Theorem 2.1, $N_{G/K}(H/K) = N/K$, where $K \lhd N \lhd G$; further $H \trianglelefteq N$. Applying Theorem 3.1(d), $N \lhd N_G(H)$, clearly, $N/K \lhd N_G(H)/K$. Since $H \trianglelefteq N_G(H)$, Theorem 2.1 implies $H/K \trianglelefteq N_G(H)/K$. Theorem 3.1(d) implies $N_G(H)/K \lhd N/K$. Therefore $N/K = N_G(H)/K$.

(h) The homomorphism $\tau: N_G(H) \to \mathrm{Aut}(H)$, given by $\tau(g)h = ghg^{-1}$, has $\ker \tau = C_G(H)$. Then $N_G(H)/C_G(H) = N_G(H)/\ker \tau \approx \mathrm{Im}\,\tau \lhd \mathrm{Aut}(H)$ implies $|N_G(H)| \leqslant |\mathrm{Aut}(H)|\,|C_G(H)|$. If H is finite, then $\mathrm{Aut}(H)$ is finite. Therefore, if H is finite, $N_G(H)$ is finite if and only if $C_G(H)$ is. Q.E.D.

The first normalizer we calculate is $N_{\mathrm{SU}(2)}(S^1)$, where $S^1 = \{e^{i\theta} \mid \theta \in \mathbf{R}\}$. It arises when we look at cyclic groups and subgroups of SO(4) not contained in SU(2). The technique used in finding it is similar to that used for all other normalizer calculations.

Lemma 3.2: $N_{\mathrm{SU}(2)}(S^1) = \{q \in \mathrm{SU}(2) \mid q = e^{i\theta} \text{ or } je^{i\theta}, \theta \in \mathbf{R}\}$.

Proof: Let $q \in \mathrm{SU}(2)$. By definition, this means for all θ there exists a ψ such that $qe^{i\theta}q^{-1} = e^{i\psi}$. Let $q = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}$. Matrix multiplication gives

$$qe^{i\theta}q^{-1} = \begin{bmatrix} aa^*e^{i\theta} + bb^*e^{-i\theta} & -abe^{i\theta} + abe^{-i\theta} \\ a^*b^*e^{-i\theta} - a^*b^*e^{i\theta} & aa^*e^{-i\theta} + bb^*e^{i\theta} \end{bmatrix}. \tag{3.1}$$

We now find for what values the right-hand side of (3.1) equals $e^{i\psi}$. That $qe^{i\theta}q^{-1} = e^{i\psi}$ implies

$$(1)\ aa^*e^{i\theta} + bb^*e^{-i\theta} = e^{i\psi}, \tag{3.2}$$

$$(2)\ -abe^{i\theta} + abe^{-i\theta} = 0. \tag{3.3}$$

Assume $ab \neq 0$. Then Eq. (3.3) implies $e^{i\theta} = e^{-i\theta}$ for all θ, a contradiction. Therefore $ab = 0$. Since $q \in \mathrm{SU}(2)$ implies a and b cannot be zero simultaneously, either $a = 0$ or $b = 0$.

Let $a = 0$ in Eq. (3.2), then $e^{-i\theta} = e^{i\psi}$. If $b = 0$, then $e^{i\theta} = e^{i\psi}$. Hence Eqs. (3.2) and (3.3) are always satisfied for $\tilde{q} \in \{q \in \mathrm{SU}(2) \mid q = e^{i\theta} \text{ or } q = je^{i\theta}\}$. Q. E. D.

The next lemma gives us the normalizers of all finite subgroups of SU(2); combining it with Theorem 3.1 allows us to calculate the isometry groups. The proof of Lemma 3.3 below is independent of the embedding of the subgroup, since any two isomorphic finite subgroups of SU(2) are conjugate in SU(2) and Theorem 3.1(e) implies their normalizers are conjugate.

Lemma 3.3:

(a) $N_{\mathrm{SU}(2)}(\mathbf{Z}_n) = \begin{cases} N_{\mathrm{SU}(2)}(S^1), & \text{if } n > 2, \\ \mathrm{SU}(2), & \text{if } n = 1,2, \end{cases}$

(b) $N_{\mathrm{SU}(2)}(D^*_{4n}) = \begin{cases} D^*_{8n}, & \text{if } n > 2, \\ O^*, & \text{if } n = 2, \end{cases}$

(c) $N_{\mathrm{SU}(2)}(T^*) = O^*$,

(d) $N_{\mathrm{SU}(2)}(O^*) = O^*$,

(e) $N_{\mathrm{SU}(2)}(I^*) = I^*$.

Proof: (a) Clear, if $n = 1$ or 2. If $n > 2$, then $N_{\mathrm{SU}(2)}(\mathbf{Z}_n) = N_{\mathrm{SU}(2)}(S^1)$. This follows immediately from the proof of Lemma 3.2 with $\theta = 2\pi/n$.

(b) For $n > 2$, D^*_{4n} is generated by $A = e^{i\pi/n}$ and $B = j$. All elements of D^*_{4n} are of the form A^s and BA^r. Let $q \in N_{\mathrm{SU}(2)}(D^*_{4n})$. Then qAq^{-1} must equal A^s because qAq^{-1} must be an element of D^*_{4n} with the same order as A. Therefore $N_{\mathrm{SU}(2)}(D^*_{4n}) \lhd N_{\mathrm{SU}(2)}(S^1)$, n > 2, which implies $q = e^{i\theta}$ or $je^{i\theta}$. It follows that $qBq^{-1} = je^{-2i\theta}$ or $je^{2i\theta}$, which are elements of D^*_{4n} only when $\theta = \pi m/2n$. Therefore $N_{\mathrm{SU}(2)}(D^*_{4n}) = D^*_{8n}$ for $n > 2$.

If $n = 2$, then D^*_8 is generated by $A = i$ and $B = j$, and $C_{\mathrm{SU}(2)}(D^*_8) = \mathbf{Z}_2$, which means $N_{\mathrm{SU}(2)}(D^*_8)$ is finite [see Theorem 3.1(h)]. Moreover, $|N_{\mathrm{SU}(2)}(D^*_8)| = 8m$ because D^*_8 is normal in $N_{\mathrm{SU}(2)}(D^*_8)$. Now $D^*_8 \trianglelefteq O^*$. Hence $O^* \lhd N_{\mathrm{SU}(2)}(D^*_8)$, [see Theorem 3.1(d)]. Since $|N_{\mathrm{SU}(2)}(D^*_8)| = 8m$ and the only finite subgroups of SU(2) are binary dihedral, cyclic, and binary polyhedral groups, it follows that $O^* \neq N_{\mathrm{SU}(2)}(D^*_8)$ implies $O^* \lhd N_{\mathrm{SU}(2)}(D^*_8) = D^*_{8m}$. The statement $O^* \lhd D^*_{8m}$ is a contradiction because the only subgroups of binary dihedral groups are cyclic or binary dihedral groups. Therefore $N_{\mathrm{SU}(2)}(D^*_8) = O^*$.

(c) $C_{\mathrm{SU}(2)}(T^*) \lhd C_{\mathrm{SU}(2)}(D^*_8)$ because $D^*_8 \lhd T^*$ and in general if $K \lhd H \lhd G$, then $C_G(H) \lhd C_G(K)$. It follows that $N_{\mathrm{SU}(2)}(T^*)$ is finite. Now $T^* \trianglelefteq O^*$. Hence $O^* \lhd N_{\mathrm{SU}(2)}(T^*)$. If $O^* \neq N_{\mathrm{SU}(2)}(T^*)$, then $N_{\mathrm{SU}(2)}(T^*) = D^*_{24k}$ which means $O^* \lhd D^*_{24k}$, a contradiction. Therefore $N_{\mathrm{SU}(2)}(T^*) = O^*$.

(d) $N_{\mathrm{SU}(2)}(O^*)$ is finite because $D^*_8 \lhd O^*$. Now $O^* \lhd N_{\mathrm{SU}(2)}(O^*)$. If $O^* \neq N_{\mathrm{SU}(2)}(O^*)$, then $O^* \lhd N_{\mathrm{SU}(2)}(O^*) = D^*_{48l}$, a contradiction. Therefore $N_{\mathrm{SU}(2)}(O^*) = O^*$.

(e) $D^*_8 \lhd I^*$, which implies $N_{\mathrm{SU}(2)}(I^*)$ is finite. Now $I^* \lhd N_{\mathrm{SU}(2)}(I^*)$. Hence $I^* \neq N_{\mathrm{SU}(2)}(I^*)$ implies $I^* \lhd N_{\mathrm{SU}(2)}(I^*) = D^*_{120n}$, a contradiction. Q.E.D.

We now use Lemma 3.3 to calculate the SO(4) normalizers of the finite noncyclic SU(2) subgroups and products of these groups with cyclic groups. For a noncyclic subgroup of SO(4), any orthogonal free action gives the same mani-

fold (see Ref. 26, page 115); hence for a noncyclic group the normalizer of the group modulo the group is independent of the embedding provided the action is free.

Theorem 3.2: Let $G \subset SU(2)$ with G finite noncyclic, then

$$\mathrm{Isom}\left(\frac{S^3}{G}\right) = \begin{cases} \mathbf{Z}_2 \times SO(3), & \text{if } G = T^* \text{ or } D^*_{4n}, \quad n < 2, \\ P_3 \times SO(3), & \text{if } G = D^*_8, \\ SO(3), & \text{if } G = O^* \text{ or } I^*, \end{cases}$$

where P_3 is the permutation group of three objects. Moreover

$$\mathrm{Isom}\left(\frac{S^3}{G \times \mathbf{Z}_k}\right) = \begin{cases} \mathbf{Z}_2 \times N_{SU(2)}(S^1), & \text{if } G = T^* \\ & \text{or } D^*_{4n}, \text{ for } n > 2, \\ P_3 \times N_{SU(2)}(S^1), & \text{if } G = D^*_8, \\ N_{SU(2)}(S^1), & \text{if } G = O^* \text{ or } I^*. \end{cases}$$

Proof: If $G \subset SU(2)$, then an embedding of G in $SO(4)$ is given by $\tilde{G}/\mathbf{Z}_2$ where $\tilde{G} = G \times \{\pm 1\} \subset SU(2) \times SU(2)$ and $\mathbf{Z}_2 = \{\pm (1,1)\}$. Theorem 3.1(f) implies

$$N_{SU(2)\times SU(2)}(G \times \{\pm 1\}) = N_{SU(2)}(G) \times N_{SU(2)}(\{\pm 1\}).$$

Lemma 3.3 implies $N_{SU(2)}(\{\pm 1\}) = SU(2)$. Hence

$$N_{SU(2)\times SU(2)}(G \times \{\pm 1\}) = N_{SU(2)}(G) \times SU(2).$$

Theorem 3.1(g) implies

$$\frac{N_{SU(2)\times SU(2)}(G \times \{\pm 1\})}{\mathbf{Z}_2} = N_{SU(2)\times SU(2)/\mathbf{Z}_2}\left(\frac{G \times \{\pm 1\}}{\mathbf{Z}_2}\right) = N_{SO(4)}(G),$$

$$\mathrm{Isom}\left(\frac{S^3}{G}\right) = \frac{N_{SO(4)}(G)}{G} = \frac{N_{SU(2)\times SU(2)}(G \times \{\pm 1\})/\mathbf{Z}_2}{(G \times \{\pm 1\})/\mathbf{Z}_2}.$$

The isomorphism Theorem 2.1 implies

$$\mathrm{Isom}\left(\frac{S^2}{G}\right) \approx \frac{N_{SU(2)\times SU(2)}(G \times \{\pm 1\})}{G \times \{\pm 1\}}.$$

It follows that[27]

$$\mathrm{Isom}\left(\frac{S^3}{G}\right) \approx \frac{N_{SU(2)}G}{G} \times \frac{SU(2)}{\{\pm 1\}} = \frac{N_{SU(2)}(G)}{G} \times SO(3).$$

From Lemma 3.3 we have

$$\mathrm{Isom}\left(\frac{S^3}{G}\right) = \begin{cases} O^*/D^*_8 \times SO(3), & \text{if } G = D^*_8, \\ D^*_{8n}/D^*_{4n} \times SO(3), & \text{if } G = D^*_{4n}, \quad n > 2, \\ O^*/T^* \times SO(3), & \text{if } G = T^*, \\ SO(3), & \text{if } G = O^* \text{ or } I^*. \end{cases}$$

Hence

$$\mathrm{Isom}\left(\frac{S^3}{G}\right) = \begin{cases} P_3 \times SO(3), & \text{if } G = D^*_8, \\ \mathbf{Z}_2 \times SO(3), & \text{if } G = T^* \text{ or } D^*_{4n}, \quad n > 2, \\ SO(3), & \text{if } G = O^* \text{ or } I^*. \end{cases}$$

If the fundamental group is $G \times \mathbf{Z}_k$, then an $SO(4)$ embedding is given by $\tilde{G}/\mathbf{Z}_2$, where $\tilde{G} = G \times \mathbf{Z}_{2k}$ and $\mathbf{Z}_{2k} = \{\pm e^{i2\pi n/k}\}$:

$$N_{SU(2)\times SU(2)}(G \times \mathbf{Z}_{2k}) = N_{SU(2)}(G) \times N_{SU(2)}(\mathbf{Z}_{2k}),$$
$$N_{SU(2)}(\mathbf{Z}_{2k}) = N_{SU(2)}(S^1).$$

Using an argument similar to the one above we have

$$\mathrm{Isom}\left(\frac{S^3}{G \times \mathbf{Z}_k}\right) \approx \frac{N_{SU(2)}(G)}{G} \times \frac{N_{SU(2)}(S^1)}{\mathbf{Z}_{2k}}.$$

We can show

$$N_{SU(2)}(S^1)/\mathbf{Z}_{2k} \approx N_{SU(2)}(S^1) \qquad \text{Q.E.D.}$$

There are two remaining families of noncyclic finite subgroups of $SO(4)$ not discussed in Theorem 3.2, namely $D'_{2^k(2n+1)} \times \mathbf{Z}_r$ and $T'_{8\cdot 3^k} \times \mathbf{Z}_r$. Again we only need to do the calculations for a particular embedding of these groups.

A particular embedding of $D'_{2^k(2n+1)} \times \mathbf{Z}_r$ in $SO(4)$, which acts freely on S^3 for $k > 2$, is $\tilde{D}/\mathbf{Z}_2$, where $\mathbf{Z}_2 = \{\pm (1,1)\}$ and $\tilde{D}$ is generated by

$$\tilde{a} = (e^{i\pi/2n+1}, -1), \quad \tilde{b} = (j, e^{i\pi/2^{k-1}}), \quad \tilde{c} = (1, e^{i\pi/r}).$$

Theorem 3.3:

$$\mathrm{Isom}\left(\frac{S^3}{D'_{2^k(2n+1)} \times \mathbf{Z}_r}\right) = \frac{D^*_{8(2n+1)} \times N_{SU(2)}(S^1)}{\tilde{D}}.$$

Proof: Let $(h_1,h_2) \in N_{SU(2)\times SU(2)}(\tilde{D})$. Then $(h_1,h_2)(g_1,g_2)(h_1^{-1},h_2^{-1}) \in \tilde{D}$. This means $h_1 g h_1^{-1} \in D^*_{4(2n+1)}$ and $h_2 g_2 h_2^{-1} \in \mathbf{Z}_{2k_r}$. Hence $h_1 \in N_{SU(2)}(D^*_{4(2n+1)})$ and $h_2 \in N_{SU(2)}(\mathbf{Z}_{2k_r})$. Therefore $N_{SU(2)\times SU(2)}(\tilde{D}) \subset D^*_{8(2n+1)} \times N_{SU(2)}(S^1)$.

Let $h \in N_{SU(2)}(S^1)$. Case (i) $h = e^{i\theta}$. Then $(1,h)\tilde{a}(1,h^{-1}) = \tilde{a}$, $(1,h)\tilde{b}(1,h^{-1}) = \tilde{b}$, and $(1,h)\tilde{c}(1,h^{-1}) = \tilde{c}$. Hence $(1,e^{i\theta}) \in N_{SU(2)\times SU(2)}(\tilde{D})$. Case (ii) $h = je^{i\theta}$. Then

$$(1,h)\tilde{a}(1,h^{-1}) = \tilde{a},$$
$$(1,h)\tilde{b}(1,h^{-1}) = (1, je^{i\theta})(j, e^{i\pi/2^{k-1}})(1, -e^{-i\theta}j) = (j, e^{-i\pi/2^{k-1}}) = \tilde{a}^{2n+1}\tilde{b}^{2^{k-1}}\tilde{b}^{-1},$$

and

$$(1,h)\tilde{c}(1,h^{-1}) = (1, je^{i\theta})(1, e^{i\pi/r})(1, -e^{-i\theta}j) = (1, e^{-i\pi/r}) = \tilde{c}^{-1}.$$

Hence $(1, je^{i\theta}) \in N_{SU(2)\times SU(2)}(\tilde{D})$. Therefore $\{1\} \times N_{SU(2)}(S^1) \subset N_{SU(2)\times SU(2)}(\tilde{D})$.

Now, we look at the action of $(e^{i\pi/2(2n+1)},1)$ and $(j,1)$ on the generators of $\tilde{D}$:

$$(e^{i\pi/2(2n+1)},1)\tilde{a}(e^{-i\pi/2(2n+1)},1) = \tilde{a},$$
$$(e^{i\pi/2(2n+1)},1)\tilde{b}(e^{-i\pi/2(2n+1)},1) = (e^{i\pi/2(2n+1)} j e^{-i\pi/2(2n+1)}, e^{i\pi/2^{k-1}}) = (j e^{-i\pi/2n+1}, e^{i\pi/2^{k-1}}) = \tilde{b}\tilde{a}^{-1}\tilde{b}^{2^{k-1}},$$

and

$$(e^{i\pi/2(2n+1)},1)\tilde{c}(e^{-i\pi/2(2n+1)},1) = \tilde{c}.$$

Hence $(e^{i\pi/2(2n+1)},1)\in N_{SU(2)\times SU(2)}(\tilde{D})$. We now have

$$(j,1)\tilde{a}(-j,1)=\tilde{a}^{-1},$$

$$(j,1)\tilde{b}(-j,1)=\tilde{b}^{-1},$$

$$(j,1)\tilde{c}(-j,1)=\tilde{c}.$$

Hence $(j,1)\in N_{SU(2)\times SU(2)}(\tilde{D})$. Now $D^*_{8(2n+1)}$ is generated by $e^{i\pi/2(2n+1)}$ and j, therefore $D^*_{8(2n+1)}\times\{1\}\leqslant N_{SU(2)\times SU(2)}(\tilde{D})$. It follows that $N_{SU(2)\times SU(2)}(\tilde{D})=D^*_{8(2n+1)}\times_{SU(2)}(S^1)$,

$$N_{SO(4)}(D'_{2^k(2n+1)}\times\mathbb{Z}_r)$$
$$=N_{SU(2)\times SU(2)/\mathbb{Z}_2}\left(\frac{\tilde{D}}{\mathbb{Z}_2}\right)$$
$$=\frac{N_{SU(2)\times SU(2)}(\tilde{D})}{\mathbb{Z}_2},$$

$$\mathrm{Isom}\left(\frac{S^3}{D'_{2^k(2n+1)}\times\mathbb{Z}_r}\right)$$
$$=\frac{N_{SO(4)}(D'_{2^k(2n+1)}\times\mathbb{Z}_r)}{D'_{2^k(2n+1)}\times\mathbb{Z}_r}$$
$$=\frac{N_{SU(2)\times SU(2)}(\tilde{D})/\mathbb{Z}_2}{\tilde{D}/\mathbb{Z}_2}$$
$$=\frac{N_{SU(2)\times SU(2)}(\tilde{D})}{\tilde{D}}.$$

Q.E.D.

A particular embedding of $T'_{8\cdot 3^k}\times\mathbb{Z}_r$ in SO(4) is $\tilde{T}/\mathbb{Z}_2$, where $\tilde{T}\leqslant SU(2)\times SU(2)$, is generated by

$\tilde{a}=\exp[(\pi/3)(i+j+k)/\sqrt{3}],\exp[i(\pi/3^k)]$,

$\tilde{b}=(j,1)$, $\tilde{c}=(k,1)$, and $\tilde{d}=(1,e^{i\pi/r})$, for $k\geqslant 2$.

Theorem 3.4:

$$\mathrm{Isom}\left(\frac{S^3}{T'_{8\cdot 3^k}\times\mathbb{Z}_r}\right)=\frac{(T^*\times S^1)\cup(O^*\setminus T^*\times jS^1)}{\tilde{T}}.$$

Proof: Clearly, $\tilde{T}\leqslant T^*\times\mathbb{Z}_{2\cdot 3^k r}$. Let $(h_1,h_2)\in N_{SU(2)\times SU(2)}(\tilde{T})$. Then $(h_1,h_2)(g_1,g_2)(h_1^{-1},h_2^{-1})\in\tilde{T}$ implies $h_1g_1h_1^{-1}\in T^*$ and $h_2g_2h_2^{-1}\in\mathbb{Z}_{2\cdot 3^k r}$. Therefore $N_{SU(2)\times SU(2)}(\tilde{T})\leqslant O^*\times N_{SU(2)}(S^1)$.

Suppose $(1,je^{i\theta})\in N_{SU(2)\times SU(2)}(\tilde{T})$. Then

$$(1,je^{i\theta})\tilde{a}(1,-e^{-i\theta}j)=(\exp[(\pi/3)(i+j+k)/\sqrt{3}],$$
$$\exp[-i(\pi/3^k)])\in\tilde{T},$$

which implies

$$(\exp[(2\pi/3)(i+j+k)/\sqrt{3}],1)\in\tilde{T}.$$

Now T^* is generated by $\exp[(2\pi/3)(i+j+k)/\sqrt{3}]$ and j. Hence it follows that $T^*\times\{1\}\leqslant\tilde{T}$, which implies that $\tilde{T}$ is a direct product of T^* and a cyclic group, a contradiction. Therefore $(1,je^{i\theta})\notin N_{SU(2)\times SU(2)}(\tilde{T})$.

Suppose $(t,je^{i\theta})\in N_{SU(2)\times SU(2)}(\tilde{T})$, where $(t,e^{i\phi})\in N_{SU(2)\times SU(2)}(\tilde{T})$. Then $(t^{-1},e^{-i\phi})\in N_{SU(2)\times SU(2)}(\tilde{T})$. This implies $(1,je^{i(\theta-\phi)})\in N_{SU(2)\times SU(2)}(\tilde{T})$, a contradiction. Therefore $N_{SU(2)\times SU(2)}(\tilde{T})$ cannot contain elements of the form $(t,je^{i\theta})$, where $(t,e^{i\phi})$ is in the normalizer.

Now, we look at the action of $(1,e^{i\theta})$ and $(e^{i\pi/4},je^{i\theta})$ on the generators: $(1,e^{i\theta})\tilde{a}(1,e^{-i\theta})=\tilde{a}$, $(1,e^{i\theta})\tilde{b}(1,e^{-i\theta})=\tilde{b}$, $(1,e^{i\theta})\tilde{c}(1,e^{-i\theta})=\tilde{c}$, $(1,e^{i\theta})\tilde{d}(1,e^{-i\theta})=\tilde{d}$. Thus

$$(e^{i\pi/4},je^{i\theta})\tilde{a}(e^{-i\pi/4},-e^{i\theta}j)=\tilde{a}^{-1}\tilde{b}\tilde{c},$$
$$(e^{i\pi/4},je^{i\theta})\tilde{b}(e^{-i\pi/4},-e^{-i\theta}j)=\tilde{c},$$
$$(e^{i\pi/4},je^{i\theta})\tilde{c}(e^{-i\pi/4},-e^{-i\theta}j)=\tilde{b}^3,$$
$$(e^{i\pi/4},je^{i\theta})\tilde{d}(e^{-i\pi/4},-e^{-i\theta}j)=\tilde{d}^{-1}.$$

Hence $(1,e^{i\theta})$, $(e^{i\pi/4},je^{i\theta})\in N_{SU(2)\times SU(2)}(\tilde{T})$, and $\tilde{T}\leqslant N_{SU(2)\times SU(2)}(\tilde{T})$. Therefore $T^*\times S^1\leqslant N_{SU(2)\times SU(2)}(\tilde{T})$. Now O^* is generated by T^* and $e^{i\pi/4}$. Therefore elements of $O^*\times N_{SU(2)}(S^1)$ with the form $(t,je^{i\phi})$, where $t\in O^*\setminus T^*$ are in $N_{SU(2)\times SU(2)}(\tilde{T})$. It follows that

$$\mathrm{Isom}\left(\frac{S^3}{T'_{8\cdot 3^k}\times\mathbb{Z}_r}\right)=\frac{(T^*\times S^1)\cup(O^*\setminus T^*\times jS^1)}{\tilde{T}}.$$

Q.E.D.

The lens spaces $L(p,q)$ are the spaces $S^3/\mathbb{Z}_p$, where the $\mathbb{Z}_p$ action is generated by $h(z_0,z_1)=(e^{2\pi i/p}z_0,e^{2\pi iq/p}z_1)$. The positive integer q tells how $\mathbb{Z}_p$ is embedded in SO(4), and satisfies $0<q<p$ and $(p,q)=1$. Clearly, this action is equivalent to the action on S^3 where

$$(z_0,z_1)=\begin{pmatrix} z_0 & z_1\\ -z_1^* & z_0^*\end{pmatrix}$$

and

$$\left(\begin{pmatrix} e^{\pi i(q+1)/p} & 0\\ 0 & e^{-\pi i(q+1)/p}\end{pmatrix},\begin{pmatrix} e^{\pi i(q-1)/p} & 0\\ 0 & e^{-\pi i(q-1)/p}\end{pmatrix}\right)$$

is an element of $SU(2)\times SU(2)$. Every space $S^3/\mathbb{Z}_p$ with $\mathbb{Z}_p$ acting freely as a subgroup of SO(4) is equivalent to a lens space. The isometry groups of lens spaces are calculated below.

Theorem 3.5:

$$\mathrm{Isom}(L(p,q))=\begin{cases} N_{SU(2)}(S^1)\times N_{SU(2)}(S^1), & \text{if } q^2\equiv 1 \bmod p \text{ and } q\not\equiv\pm 1 \bmod p,\\ N_{SU(2)}(S^1)\times SU(2)/\mathbb{Z}_p, & \text{if } q\equiv\pm 1 \bmod p,\\ O(4)/\mathbb{Z}_2, & \text{if } p=2,\\ (S^1\times S^1)\cup j(S^1\times S^1)\cup[(S^1\times jS^1)\cup(jS^1\times S^1)]T/\mathbb{Z}_p, & \text{if } p>2 \text{ and } q^2\equiv -1 \bmod p,\\ (S^1\times S^1)\cup j(S^1\times S^1)/\mathbb{Z}_p, & \text{otherwise.}\end{cases}$$

Proof: Let $S^3=\{(z_0,z_1)\in\mathbb{C}^2|z_0z_0^*+z_1z_1^*=1\}$. Then $\gamma(z_0,z_1)=(e^{2\pi i/p}z_0,\ e^{2\pi iq/p}z_1)$, $g\in N_{SO(4)}(\mathbb{Z}_p)$ if and only if $g\gamma g^{-1}=\gamma^n$ for some n. This is equivalent to $\tilde{g}\gamma=\gamma^n\tilde{g}$, where

$$\tilde{g}=\left(\begin{pmatrix} a & b\\ -b^* & a^*\end{pmatrix},\begin{pmatrix} c & d\\ -d^* & c^*\end{pmatrix}\right)\in SU(2)\times SU(2)$$

with action given by

$$g(z_0, z_1) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}\begin{pmatrix} z_0 & z_1 \\ -z_1^* & z_0^* \end{pmatrix}\begin{pmatrix} c^* & -d \\ d^* & c \end{pmatrix},$$

$$\tilde{g}\gamma(z_0,z_1) = \tilde{g}(e^{2\pi i/p} z_0, e^{2\pi iq/p} z_1) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}\begin{pmatrix} e^{2\pi i/p} z_0 & e^{2\pi iq/p} z_1 \\ -e^{-2\pi iq/p} z_1^* & e^{-2\pi i/p} z_0^* \end{pmatrix}\begin{pmatrix} c^* & -d \\ d^* & c \end{pmatrix},$$

$$\tilde{g}\gamma(z_0,z_1) = \begin{pmatrix} ac^* e^{2\pi i/p} z_0 - bc^* e^{-2\pi iq/p} z_1^* + ad^* e^{2\pi iq/p} z_1 + bd^* e^{-2\pi i/p} z_0^* \\ ade^{2\pi i/p} z_0 + bde^{-2\pi iq/p} z_1^* + ace^{2\pi iq/p} z_1 + bce^{-2\pi i/p} z_0^* \end{pmatrix}, \tag{3.4}$$

$$\gamma^n \tilde{g}(z_0,z_1) = \begin{pmatrix} e^{2\pi in/p}(ac^* z_0 - bc^* z_1 + ad^* z_1 + bd^* z_0^*) \\ e^{2\pi iqn/p}(-adz_0 + bdz_1^* + acz_1 + bcz_0^*) \end{pmatrix}. \tag{3.5}$$

Now, find for what values of a, b, c, and d there exists an n such that $\tilde{g}\gamma = \gamma^n \tilde{g}$. This means (3.4) and (3.5) are equal, which implies

$$ac^* e^{2\pi i/p} z_0 - bc^* e^{-2\pi iq/p} z_1^* + ad^* e^{2\pi iq/p} z_1 + bd^* e^{-2\pi q/p} z_0^* = e^{-2\pi in/p}(ac^* z_0 - bc^* z_1^* + ad^* z_1 + bd^* z_0^*),$$
$$-ade^{2\pi i/p} z_0 - bde^{-2\pi iq/p} z_1^* + ace^{2\pi iq/p} z_1 + bce^{-2\pi iq/p} z_0^* = e^{2\pi iqn/p}(-adz_0 - bdz_1^* + acz_1 + bcz_0^*).$$

These two equations must hold for all $(z_0, z_1) \in S^3$. In particular they must be true for (1,0), (i,0), (0,1), and (0,i). These values of (z_0, z_1) give the following eight equations:

$$ace^{2\pi i/p} + bd^* e^{-2\pi i/p} = e^{2\pi in/p}(ac^* + bd^*), \tag{3.6}$$
$$-ade^{2\pi i/p} + bce^{-2\pi i/p} = e^{2\pi iqn/p}(-ad + bc), \tag{3.7}$$
$$ac^* e^{2\pi i/p} - bd^* e^{-2\pi i/p} = e^{2\pi in/p}(ac^* - bd^*), \tag{3.8}$$
$$-ade^{2\pi i/p} - bce^{-2\pi i/p} = e^{2\pi iqn/p}(-ad - bd^*), \tag{3.9}$$
$$ad^* e^{2\pi iq/p} - bc^* e^{-2\pi iq/p} = e^{2\pi in/p}(ad^* - bc^*), \tag{3.10}$$
$$ace^{2\pi iq/p} + bd^* e^{-2\pi iq/p} = e^{2\pi iqn/p}(ac + bd^*), \tag{3.11}$$
$$ad^* e^{2\pi iq/p} + bc^* e^{-2\pi iq/p} = e^{2\pi in/p}(ad^* + bc^*), \tag{3.12}$$
$$ace^{2\pi iq/p} - bd^* e^{-2\pi iq/p} = e^{2\pi iqn/p}(ac - bd^*). \tag{3.13}$$

Adding equation (3.6) to (3.7), (3.8) to (3.9), (3.10) to (3.12), and (3.11) to (3.13) implies the following set of conditions:

$$e^{2\pi i/p} = e^{2\pi in/p}, \quad \text{if } ac^* \neq 0,$$
$$e^{2\pi i/p} = e^{2\pi iqn/p}, \quad \text{if } ad \neq 0,$$
$$e^{2\pi iq/p} = e^{2\pi in/p}, \quad \text{if } ad^* \neq 0,$$
$$e^{2\pi iq/p} = e^{2\pi iqn/p}, \quad \text{if } ac \neq 0.$$

Likewise subtracting equations (3.8) from (3.6), (3.9) from (3.7), (3.12) from (3.10), and (3.13) from (3.11), gives the following conditions:

$$e^{-2\pi i/p} = e^{2\pi in/p}, \quad \text{if } bd^* \neq 0,$$
$$e^{-2\pi i/p} = e^{2\pi iqn/p}, \quad \text{if } bc \neq 0,$$
$$e^{-2\pi iq/p} = e^{2\pi in/p}, \quad \text{if } bc^* \neq 0,$$
$$e^{-2\pi iq/p} = e^{2\pi iqn/p}, \quad \text{if } bd^* \neq 0.$$

Combining the above conditions we have the following.

If $ac \neq 0$, then

$$e^{2\pi i(1-n)/p} = 1 \quad \text{and} \quad e^{2\pi iq(1-n)/p} = 1.$$

If $ad \neq 0$, then

$$e^{2\pi i(1-qn)/p} = 1 \quad \text{and} \quad e^{2\pi i(q-n)/p} = 1.$$

If $bd \neq 0$, then

$$e^{2\pi i(n+1)/p} = 1 \quad \text{and} \quad e^{2\pi iq(n+1)/p} = 1.$$

If $bc \neq 0$, then

$$e^{2\pi i(qn+1)/p} = 1 \quad \text{and} \quad e^{2\pi iq(n+q)/pp} = 1.$$

Suppose $ac \neq 0$ and $b \neq 0$, then it follows from the above that $e^{2\pi i(1-n)/p} = 1$, $e^{2\pi iq(1-n)/p} = 1$, $e^{2\pi i(qn+1)/p} = 1$, and $e^{2\pi iq(1-n)/p} = 1$, which implies $n = 1 - ps$, $q(1-n) = pr$, $qn + 1 = pk$, and $q(1-n) = pl$. Now $n = 1 - ps$ implies $qps = pr$, $q(1-ps) + 1 = pk$, and $qps = pl$. These equations imply $q = p(s+k) - 1$, $q \equiv -1 \bmod p$. Suppose $d \neq 0$ was also nonzero, then $e^{2\pi i(n+1)/p} = 1$. So $n + 1 = pj$, which is only true for $p = 2$. Therefore $ac \neq 0$ and $b \neq 0$ implies $q \equiv -1 \bmod p$ and $d = 0$ unless $p = 2$.

Suppose $ac \neq 0$ and $d \neq 0$, then $e^{2\pi i(1-n)/p} = 1$, $e^{2\pi iq(1-n)/p} = 1$, $e^{2\pi i(1-qn)/p} = 1$, and $e^{2\pi i(q-n)/p} = 1$. Hence $n = 1 - ps$, $q(1-n) = pr$, $1 - qn = pk$, and $q - n = pl$, which implies $q \equiv 1 \bmod p$. Again, $a,b,c,d \neq 0$ implies $p = 2$. Therefore, $ac \neq 0$ and $d \neq 0$ implies $q \equiv 1 \bmod p$ unless $p = 2$.

If we apply the above conditions $ac \neq 0$ and $b \neq 0$ to (3.4) and (3.5) we obtain

$$\tilde{g}\gamma(z_0,z_1) = \begin{pmatrix} ac^* e^{2\pi i/p} z_0 - bc^* e^{-2\pi iq/p} z_1 \\ ace^{2\pi iq/p} z_1 + bce^{-2\pi i/p} z_0^* \end{pmatrix},$$

$$\gamma^n \tilde{g}(z_0,z_1) = \begin{pmatrix} e^{2\pi in/p}(ac^* z_0 - bc^* z_1) \\ e^{2\pi iqn/p}(acz_1 + bcz_0^*) \end{pmatrix},$$

which are equal for all $(z_0,z_1) \in S^3$, when $n = 1$ and $q \equiv -1 \bmod p$. Likewise, if we apply $ac \neq 0$ and $d \neq 0$, we obtain

$$\tilde{g}\gamma(z_0,z_1) = \begin{pmatrix} ac^* e^{2\pi i/p} z_0 + ad^* e^{2\pi iq/p} z_1 \\ ace^{2\pi iq/p} z_1 - ade^{2\pi i/p} z_0 \end{pmatrix},$$

$$\gamma^n \tilde{g}(z_0,z_1) = \begin{pmatrix} e^{2\pi in/p}(ac^* z_0 + ad^* z_1) \\ e^{2\pi iqn/p}(ac^* z_1 - adz_0) \end{pmatrix},$$

which are equal for all $(z_0, z_1) \in S^3$, when $n = 1$ and $q \equiv 1 \bmod p$. The above results imply $N_{SU(2)\times SU(2)}(\mathbf{Z}_p) = N_{SU(2)}(S^1) \times SU(2)$, for $q \equiv 1 \bmod p$, $N_{SU(2)\times SU(2)}(\mathbf{Z}_p) = SU(2) \times N_{SU(2)}(S^1)$, for $q \equiv -1 \bmod p$, and $N_{SU(2)\times SU(2)}(\mathbf{Z}_2) = SU(2) \times SU(2)$.

Suppose $ad \neq 0$. Now look at Eqs. (3.4) and (3.5). If $b = c = 0$ we have

$$\tilde{g}\gamma(z_0,z_1) = (ad^* e^{2\pi iq/p} z_1, ade^{2\pi i/p} z_0),$$

and

$$\gamma^n \tilde{g}(z_0,z_1) = (e^{2\pi in/p} ad^* z_1, e^{2\pi iqn/p} adz_0),$$

which are equal for all $(z_0, z_1) \in S^3$ if and only if $n = q$ and $q^2 \equiv 1 \bmod p$.

Suppose $bd \neq 0$ and $a = c = 0$, then (3.4) and (3.5) become

$$\tilde{g}\gamma(z_0,z_1) = (bd^* e^{-2\pi i/p} z_0^*, bd^* e^{-2\pi iq/p} z_1^*),$$

and

$$\gamma^n \tilde{g}(z_0,z_1) = (e^{2\pi in/p} bd^* z_0^*, e^{2\pi iqn/p} bdz_1^*),$$

which are equal for all $(z_0,z_1) \in S^3$, when $n = -1$.

It follows from the above that

$$N_{SU(2)\times SU(2)}(\mathbf{Z}_p) = \begin{cases} N_{SU(2)}(S^1) \times N_{SU(2)}(S^1), \\ \quad \text{if } q^2 \equiv 1 \bmod p \\ \quad \text{and } q \not\equiv \pm 1 \bmod p, \\ (S^1 \times S^1) \cup j(S^1 \times S^1), \\ \quad \text{if } q^2 \not\equiv 1 \bmod p. \end{cases}$$

The above calculations give the orientation preserving isometries of $L(p,q)$. If $p = 2$, then $L(2,1) = S^3/\mathbf{Z}_2$ and

$$\mathbf{Z}_2 = \left\{ \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{pmatrix}, \begin{pmatrix} -1 & & & 0 \\ & -1 & & \\ & & -1 & \\ 0 & & & -1 \end{pmatrix} \right\},$$

a normal subgroup of $O(4)$. Hence $\mathrm{Isom}(L(2,1)) = O(4)/\mathbf{Z}_2$, which implies $L(2,1)$ has orientation reversing isometries. Suppose $p > 2$, we would like to find the orientation reversing isometries of $L(p,q)$. Clearly, every element of $O(4) \setminus SO(4)$ has the action $gT(z_0,z_1)$, where $T(z_0,z_1) = (z_0^*, z_1)$ and $g \in SO(4)$. The definition of normalizer implies that $gT \in N_{O(4)}(\mathbf{Z}_p)$ if and only if $gT\gamma T^{-1} g^{-1} = \gamma^n$, which is equivalent to $\tilde{g}T\gamma(z_0,z_1) = \gamma^n \tilde{g}T(z_0,z_1)$, for all $(z_0,z_1) \in S^3$ and $\tilde{g} \in SU(2) \times SU(2)$.

Now, we express both $\tilde{g}T\gamma(z_0,z_1)$ and $\gamma^n \tilde{g}T(z_0,z_1)$ in terms of

$$\tilde{g} = \left(\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \begin{pmatrix} c & d \\ -d^* & c^* \end{pmatrix} \right).$$

We have

$$\tilde{g}T\gamma(z_0,z_1) = \begin{pmatrix} ac^* e^{-2\pi i/p} z_0^* - bc^* e^{-2\pi iq/p} z_1^* + ad^* e^{2\pi iq/p} z_1 + bd^* e^{2\pi i/p} z_0 \\ -ade^{-2\pi i/p} z_0^* + bde^{-2\pi iq/p} z_1^* + ace^{2\pi iq/p} z_1 + bce^{2\pi i/p} z_0 \end{pmatrix}, \tag{3.14}$$

and

$$\gamma^n \tilde{g}T(z_0,z_1) = \begin{pmatrix} e^{2\pi in/p}(ac^* z_0^* - bc^* z_1^* + ad^* z_1 + bd^* z_0) \\ e^{2\pi iqn/p}(-adz_0^* + bdz_1^* + acz_1 + bcz_0) \end{pmatrix}. \tag{3.15}$$

Evaluating (3.14) and (3.15) at (1,0), $(i,0)$, (0,1), and $(0,i)$, we have the following eight equations:

$$ac^* e^{-2\pi i/p} + bd^* e^{2\pi i/p} = e^{2\pi in/p}(ac^* + bd^*), \tag{3.16}$$

$$-ade^{-2\pi i/p} + bce^{2\pi i/p} = e^{2\pi iqn/p}(-ad + bc), \tag{3.17}$$

$$-ac^* e^{-2\pi i/p} + bd^* e^{2\pi i/p} = e^{2\pi in/p}(-ac^* + bd^*), \tag{3.18}$$

$$ade^{-2\pi i/p} + bce^{2\pi i/p} = e^{2\pi iqn/p}(ad + bc), \tag{3.19}$$

$$-bc^* e^{-2\pi iq/p} + ad^* e^{2\pi iq/p} = e^{2\pi in/p}(-bc^* + ad^*), \tag{3.20}$$

$$bde^{-2\pi iq/p} + ace^{2\pi iq/p} = e^{2\pi iqn/p}(bd + ac), \tag{3.21}$$

$$bc^* e^{-2\pi iq/p} + ad^* e^{2\pi iq/p} = e^{2\pi in/p}(bc^* + ad^*), \tag{3.22}$$

$$-bde^{-2\pi iq/p} + ace^{2\pi iq/p} = e^{2\pi iqn/p}(ac - bd). \tag{3.23}$$

Adding equation (3.16) to (3.18), (3.17) to (3.19), (3.20) to (3.22), and (3.21) to (3.23), we have

$$bd^* e^{2\pi i/p} = bd^* e^{2\pi in/p}, \tag{3.24}$$

$$bce^{2\pi i/p} = bce^{2\pi iqn/p}, \tag{3.25}$$

$$ad^* e^{2\pi iq/p} = ad^* e^{2\pi in/p}, \tag{3.26}$$

$$ace^{2\pi iq/p} = ace^{2\pi iqn/p}, \tag{3.27}$$

Subtracting equation (3.16) from (3.18), (3.17) from (3.19), (3.20) from (3.22), and (3.21) from (3.11) we have

$$ac^* e^{-2\pi i/p} = ac^* e^{2\pi in/p}, \tag{3.28}$$

$$ade^{-2\pi i/p} = ade^{2\pi iqn/p}, \tag{3.29}$$

$$bc^* e^{-2\pi iq/p} = bc^* e^{2\pi in/p}, \tag{3.30}$$

$$bde^{-2\pi iq/p} = bde^{2\pi iqn/p}. \tag{3.31}$$

Now, suppose $bd^* \neq 0$, then (3.24) implies $n = 1$. Equation (3.31) gives $e^{-2\pi iq/p} = e^{2\pi iq/p}$, a contradiction unless $p = 2$, because $(p,q) = 1$. Likewise, if $ac \neq 0$, (3.28) implies $n = -1$ and (3.27) implies $e^{2\pi iq/p} = e^{-2\pi iq/p}$, a contradiction. Therefore, $bd^* = 0$ and $ac = 0$.

Suppose $bc \neq 0$, then (3.30) implies $n = -q$. Equation (3.25) implies $e^{2\pi i/p} = e^{-2\pi iq^2/p}$, which is true if and only if $q^2 = -1 + sp$. Likewise, if $ad \neq 0$, then (3.26) implies $n = q$. Equation (3.29) implies $e^{-2\pi i/p} = e^{2\pi iq^2/p}$. Hence $q^2 = -1 + sp$. Therefore $\tilde{g}T \in N_{O(4)}(\mathbf{Z}_p)$ if and only if $q^2 \equiv -1 \bmod p$. Further, either $ad \neq 0$ and $bc = 0$, or $bc \neq 0$ and $ad = 0$ unless $p = 2$. The action of these orientation reversing isometries on S^3 are given by

$$g_1 T(z_0,z_1) = (ad^* e^{2\pi iq/p} z_1, -ade^{-2\pi i/p} z_0^*)$$

or

$$g_2 T(z_0,z_1) = (-bc^* e^{-2\pi iq/p} z_1^*, bce^{2\pi i/p} z_0).$$

Therefore, $\mathrm{Isom}(L(p,q)) = \mathrm{Isom}^+(L(p,q))$ for $q^2 \not\equiv -1 \bmod p$ and

$$\mathrm{Isom}(L(p,q)) = (S^1 \times S^1) \cup j(S^1 \times S^1) \cup [(jS^1 \times S^1) \cup (S^1 \times jS^1)]T/\mathbf{Z}_p$$

for $q^2 \equiv -1 \bmod p$. Q.E.D.

Table I summarizes the results of this section. The groups $\tilde{D}$, $\tilde{T}$, and $(jS^1 \times S^1)T$ are defined in Theorems 3.3, 3.4, and 3.5, respectively.

TABLE I. The isometry groups of spherical spaces $M = S^3/G$.

$\pi_1 M$	Isom (M)
D^*_{4n}, $n>3$	$\mathbf{Z}_2\times SO(3)$
D^*_8	$P_3\times SO(3)$
T^*	$\mathbf{Z}_2\times SO(3)$
O^*	$SO(3)$
I^*	$SO(3)$
$D^*_{4n}\times\mathbf{Z}_p$, $n>3$	$\mathbf{Z}_2\times N_{SU(2)}(S^1)$
$D^*_8\times\mathbf{Z}_p$	$P_3\times N_{SU(2)}(S^1)$
$T^*\times\mathbf{Z}_p$	$\mathbf{Z}_2\times N_{SU(2)}(S^1)$
$O^*\times\mathbf{Z}_p$	$N_{SU(2)}(S^1)$
$I^*\times\mathbf{Z}_p$	$N_{SU(2)}(S^1)$
$D'_{2^k(2n+1)}\times\mathbf{Z}_p$, $p=1$ or $(2^k(2n+1),p)=1$	$D^*_{8(2n+1)}\times N_{SU(2)}(S^1)/\tilde{D}$
$T'_{8\cdot 3^k}\times\mathbf{Z}_p$, $p=1$ or $(8\cdot 3^k,p)=1$	$(T^*\times S^1)\cup(O^*\setminus T^*\times jS^1)/\tilde{T}$
$\mathbf{Z}_2$	$O(4)/\mathbf{Z}_2$
$\mathbf{Z}_p$, $q^2\equiv 1 \bmod p$ with $q\not\equiv\pm 1 \bmod p$	$N_{SU(2)}(S^1)\times N_{SU(2)}(S^1)/\mathbf{Z}_p$
$\mathbf{Z}_p$, $q\equiv\pm 1 \bmod p$ with $p>2$	$N_{SU(2)}(S^1)\times SU(2)/\mathbf{Z}_p$
$\mathbf{Z}_p$, $q^2\equiv -1 \bmod p$ with $p>2$	$(S^1\times S^1)\cup j(S^1\times S^1)\cup[(S^1\times jS^1)\cup(jS^1\times S^1)]T/\mathbf{Z}_p$
$\mathbf{Z}_p$, remaining cases	$(S^1\times S^1)\cup j(S^1\times S^1)/\mathbf{Z}_p$

IV. π_0 Isom(S^3/G) AND π_0 Diff(S^3/G)

The groups π_0 Isom(S^3/G) are found using the definition of π_0. The groups π_0 Diff(S^3/G) are given as results of the weak form of Hatcher's conjecture.

Strong Conjecture[13]: Isom$(S^3/G)\simeq$Diff(S^3/G). The motivation for the conjecture is the following. Suppose $G=1$, then the spherical space is just S^3, Isom$(S^3)=O(4)$, and Hatcher[11] has shown that $O(4)\simeq$Diff(S^3). Therefore the conjecture holds for $G=1$. If $G\neq 1$, then S^3 is the universal covering space of S^3/G and isometries of S^3/G lift to isometries of S^3; likewise diffeomorphisms of S^3/G lift to S^3. Because the diffeomorphisms and isometries lift and $O(4)\simeq$Diff(S^3), it seems feasible that Isom$(S^3/G)\simeq$Diff(S^3/G). So far the conjecture has been proven for S^3, $S^3/\mathbf{Z}_2$, $S^3/D'_{2^k(2n+1)}$, $S^3/D'_{2^k(2n+1)}\times\mathbf{Z}_p$, and for $S^3/D^*_{4m}\times\mathbf{Z}_p$, $m\neq 2$, by Ivanov.[14] It turns out for the calculations we do all that is needed is the following weaker conjecture.

Weak Conjecture: (i) π_0 Diff$(S^3/G)=\pi_0$ Isom(S^3/G), (ii) π_1 Diff(S^3/G) has the same number of generators as π_1 Isom(S^3/G).

After each theorem giving π_0 Isom(S^3/G), there is a corollary stating part (i) of the weak conjecture in cases for which π_0 Diff(S^3/G) has been calculated by Rubinstein. The last theorem in this section (Theorem 4.6) uses lens spaces to prove that the only spaces with orientation reversing diffeomorphisms are lens spaces. Moreover, this result is independent of the conjecture. Table II summarizes the results.

TABLE II. Topological structure of the isometry groups of spherical spaces $M=S^3/G$, and indication of which conjecture is known for each space.

$\pi_1 M$	Topology of Isom (M)	π_0 Isom (M)	Conjecture	Orientation reversing diffeomorphisms
D^*_{4n}, $n>3$	$\mathbf{Z}_2\times\mathbf{R}P^3$	$\mathbf{Z}_2$	W	No
D^*_8	$P_3\times\mathbf{R}P^3$	P_3	W	No
T^*	$\mathbf{Z}_2\times\mathbf{R}P^3$	$\mathbf{Z}_2$	Open	No
O^*	$\mathbf{R}P^3$	1	W	No
I^*	$\mathbf{R}P^3$	1	Open	No
$D^*_{4n}\times\mathbf{Z}_p$, $n>3$	$\mathbf{Z}_2\times\mathbf{Z}_2\times S^1$	$\mathbf{Z}_2\times\mathbf{Z}_2$	S	No
$D^*_8\times\mathbf{Z}_p$	$P_3\times\mathbf{Z}_2\times S^1$	$P_3\times\mathbf{Z}_2$	W	No
$T^*\times\mathbf{Z}_p$	$\mathbf{Z}_2\times\mathbf{Z}_2\times S^1$	$\mathbf{Z}_2\times\mathbf{Z}_2$	Open	No
$O^*\times\mathbf{Z}_p$	$\mathbf{Z}^2\times S^1$	$\mathbf{Z}_2$	W	No
$I^*\times\mathbf{Z}_p$	$\mathbf{Z}_2\times S^1$	$\mathbf{Z}_2$	Open	No
$D'_{2^k(2n+1)}\times\mathbf{Z}_p$, $p>1$	$\mathbf{Z}_2\times\mathbf{Z}_2\times S^1$	$\mathbf{Z}_2\times\mathbf{Z}_2$	S	No
$T'_{8\cdot 3^k}\times\mathbf{Z}_p$, $p>1$	$\mathbf{Z}_2\times S^1$	$\mathbf{Z}_2$	Open	No
$\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{R}P^3\times\mathbf{R}P^3$	$\mathbf{Z}_2$	S	Yes
$\mathbf{Z}_p$, $q^2\equiv 1 \bmod p$ with $q\not\equiv\pm 1 \bmod p$	$\mathbf{Z}_2\times\mathbf{Z}_2\times S^1\times S^1$	$\mathbf{Z}_2\times\mathbf{Z}_2$	W	No
$\mathbf{Z}_p$, $q\equiv\pm 1 \bmod p$ with $p>2$	$\mathbf{Z}_2\times S^1\times\mathbf{R}P^3$	$\mathbf{Z}_2$	W	No
$\mathbf{Z}_p$, $q^2\equiv -1 \bmod p$ with $p>2$	$\mathbf{Z}_4\times S^1\times S^1$	$\mathbf{Z}_4$	W	Yes
$\mathbf{Z}_p$, remaining cases	$\mathbf{Z}_2\times S^1\times S^1$	$\mathbf{Z}_2$	W	No
1	$\mathbf{Z}_2\times S^3\times\mathbf{R}P^3$	$\mathbf{Z}_2$	S	Yes

As an example of calculating $\pi_0\,\mathrm{Isom}(S^3/G)$. Look at $G=\mathbf{Z}_2$. Then $\mathrm{Isom}(S^3/\mathbf{Z}_2)=\mathrm{O}(4)/\mathbf{Z}_2$. The identity component of $\mathrm{O}(4)/\mathbf{Z}_2$ is just $\mathrm{SO}(4)/\mathbf{Z}_2$; hence

$$\pi_0\,\mathrm{Isom}(S^3/\mathbf{Z}_2)=\mathrm{O}(4)/\mathbf{Z}_2/\mathrm{SO}(4)/\mathbf{Z}_2$$
$$=\mathrm{O}(4)/\mathrm{SO}(4)=\mathbf{Z}_2\,.$$

Looking at Table I, most of the calculations will be this easy, and the following theorems give the results.

Theorem 4.1:

$$\pi_0\,\mathrm{Isom}(S^3/G)$$
$$=\begin{cases}P_3 & \text{if } G=D^*_8,\\ P_3\times\mathbf{Z}_2, & \text{if } G=D^*_8\times\mathbf{Z}_p \text{ with } p>1,\\ \mathbf{Z}_2\times\mathbf{Z}_2, & \text{if } G=D^*_{4n}\times\mathbf{Z}_p,\ n\geqslant 3,\\ & \quad \text{or } T^*\times\mathbf{Z}_p, \text{ with } p>1,\\ \mathbf{Z}_2, & \text{if } G=D^*_{4n},\ n\geqslant 3, T^*, O^*\times\mathbf{Z}_p,\\ & \quad \text{or } I^*\times\mathbf{Z}_p \text{ with } p>1,\\ 1, & \text{if } G=O^*, \text{or } I^*.\end{cases}$$

Proof: If $G\leqslant\mathrm{SU}(2)$ and noncyclic, then Theorem 3.2 implies

$$\mathrm{Isom}(S^3/G)=\begin{cases}\mathbf{Z}_2\times\mathrm{SO}(3), & \text{if } G=T^*, \text{or } D^*_{4n},\ n\geqslant 3,\\ P_3\times\mathrm{SO}(3), & \text{if } G=D^*_8,\\ \mathrm{SO}(3), & \text{if } G=O^* \text{ or } I^*.\end{cases}$$

Since $\mathrm{SO}(3)$ is connected, the only disconnected part is due to the finite groups P_3 or $\mathbf{Z}_2$. Therefore

$$\pi_0\,\mathrm{Isom}(S^3/G)=\begin{cases}\mathbf{Z}_2, & \text{if } G=T^*, \text{or } D^*_{4n},\ n\geqslant 3,\\ P_3, & \text{if } G=D^*_8,\\ 1, & \text{if } G=O^* \text{ or } I^*.\end{cases}$$

If $G=H\times\mathbf{Z}_p$, where $H\leqslant\mathrm{SU}(2)$, and H is noncyclic, then Theorem 3.2 implies

$$\mathrm{Isom}(S^3/G)=\begin{cases}\mathbf{Z}_2\times N_{\mathrm{su}(2)}(S^1), & \text{if } G=T^*\times\mathbf{Z}_p,\\ & \quad \text{or } D^*_{4n}\times\mathbf{Z}_p,\ n\geqslant 3,\\ P_3\times N_{\mathrm{su}(2)}(S^1), & \text{if } G=D^*_8\times\mathbf{Z}_p,\\ N_{\mathrm{su}(2)}(S^1), & \text{if } G=O^*\times\mathbf{Z}_p\\ & \quad \text{or } I^*\times\mathbf{Z}_p.\end{cases}$$

Clearly, $N_{\mathrm{su}(2)}(S^1)$ has two components. Therefore

$$\pi_0\,\mathrm{Isom}(S^3/G)=\begin{cases}\mathbf{Z}_2\times\mathbf{Z}_2, & \text{if } G=T^*\times\mathbf{Z}_p,\\ & \quad \text{or } D^*_{4n}\times\mathbf{Z}_p, \quad n\geqslant 3,\\ P_3\times\mathbf{Z}_2, & \text{if } G=D^*_8\times\mathbf{Z}_p,\\ \mathbf{Z}_2, & \text{if } G=O^*\times\mathbf{Z}_p \text{ or } I^*\times\mathbf{Z}_p.\end{cases}$$

Q.E.D.

Corollary 4.2: $\pi_0\,\mathrm{Diff}(S^3/G)=\pi_0\,\mathrm{Isom}(S^3/G)$ for $G\neq T^*\times\mathbf{Z}_p$ or $I^*\times\mathbf{Z}_p$.

Proof: See Rubinstein[15] and Theorem 4.1 for $G=D^*_{4n}\times\mathbf{Z}_p$ and $D'_{2^k(2n+1)}\times\mathbf{Z}_p$ with $p\geqslant 3$. See Rubinstein and Birman[16] and Theorem 4.1 for $G=O^*\times\mathbf{Z}_p$ with $p\geqslant 1$. Q.E.D.

The groups $\pi_0\,\mathrm{Isom}(S^3/G)$ above were easy to calculate because the isometry groups were just products. For $D'_{2^k(2n+1)}$ and $T'_{8\cdot 3^k}$ this is not the case and it takes a little more work.

Theorem 4.3: $\pi_0\,\mathrm{Isom}(S^3/D'_{2^k(2n+1)}\times\mathbf{Z}_p)=\mathbf{Z}_2\times\mathbf{Z}_2$.

Proof: Let $a=[(1,j)]$, $b=[(e^{i\pi/2(2n+1)},1]$, and $c=[(e^{i\pi/2(2n+1)},j)]$, where $[(\ )]$ means equivalence classes in $\pi_0\,\mathrm{Isom}(S^3/G)$. Clearly, no two of the above elements of $\pi_0\,\mathrm{Isom}(S^3/G)$ are connected, or connected to the identity.

All elements of $\pi_0\,\mathrm{Isom}(S^3/G)$ are of the form $[(e^{i\pi s/2(2n+1)},e^{i\theta})]$, $[(e^{i\pi s/2(2n+1)},e^{i\theta}j)]$, $[(e^{i\pi s/2(2n+1)}j,e^{i\theta}]$ or $[(e^{i\pi s/2(2n+1)}j,e^{i\theta}j)]$, where $s\in\mathbf{Z}$ and $\theta\in\mathbf{R}$. The classes $[(e^{i\pi s/2(2n+1)},e^{i\theta})]$ are the same because $[(e^{i\pi s/2(2n+1)},e^{i\theta})]=[\tilde{b}\,][(e^{i\pi s/2(2n+1)},e^{i\theta'})]$, where $\tilde{b}=(j,e^{i\pi/2^{k-1}})$. The classes $[(e^{i\pi s/2(2n+1)},e^{i\theta}j)]$ and $[(e^{i\pi s/2(2n+1)}j,e^{i\theta}j)]$ are equivalent for the same reason. So without loss of generality we only need to consider $[(e^{i\pi s/2(2n+1)},e^{i\theta})]$ and $[(e^{i\pi s/2(2n+1)},e^{i\theta}j)]$.

If s is even, then $[(e^{i\pi s/2(2n+1)},e^{i\theta}]$ is connected to the identity. This is true because if we define f: $[0,1]\to\mathrm{Isom}(S^3/G)$ by

$$f(\lambda)=[(e^{i\pi s/2(2n+1)},e^{i[\theta+\lambda(2\pi-\theta)]})],$$
$$f(1)=[(e^{i\pi s/2(2n+1)},1)2]$$
$$=[\tilde{a}^{-s/2}\tilde{b}^{2^{k-1}}][(e^{i\pi s/2(2n+1)},1)]$$
$$=[(1,1)], \quad \text{if } s/2 \text{ is odd},$$
$$=[\tilde{a}^{-s/2}][(e^{i\pi s/2(2n+1)},1)]$$
$$=[(1,1)], \quad \text{if } s/2 \text{ is even},$$
$$f(0)=[(e^{i\pi s/2(2n+1)},e^{i\theta}]\,.$$

This is the path that is connected to the identity.

If s is odd, then $[(e^{i\pi s/2(2n+1)},e^{i\theta})]$ is connected to b. Because, using the f above we have

$$f(1)=[(e^{i\pi s/2(2n+1)},1)]=[\tilde{a}^{1-s/2}][(e^{i\pi s/2(2n+1)},1)]$$
$$=[(e^{i\pi/2(2n+1)},1)]\,.$$

If s is even, then $[(e^{i\pi s/2(2n+1)},e^{i\theta}j)]$ is connected to a. Let g: $[0,1]\to\mathrm{Isom}(S^3/G)$ be defined by

$$g(\lambda)=[(e^{i\pi s/2(2n+1)},e^{i(\theta+\lambda(2\pi-\theta))}j)]\,,$$
$$g(1)=[(e^{i\pi s/2(2n+1)},j)]\,,$$
$$g(1)=[\tilde{a}^{-s/2}][(e^{i\pi s/2(2n+1)},j)]=[(1,j)], \quad \text{if } s/2 \text{ is even},$$
$$g(1)=[\tilde{a}^{-s/2}\tilde{b}^{2^{k-1}}][(e^{i\pi s/2(2n+1)},j)]$$
$$=[(1,j)], \quad \text{if } s/2 \text{ is odd}.$$

If s is odd using g it follows that $[(e^{i\pi s/2(2n+1)},e^{i\theta}j)]$ is connected to c.

Clearly $a^2=b^2=c^2=1$. Therefore $\pi_0\,\mathrm{Isom}(S^3/G)$ has four elements $\{1,a,b,c\}$. It follows that $\pi_0\,\mathrm{Isom}(S^3/G)=\mathbf{Z}_2\times\mathbf{Z}_2$. Q.E.D.

Corollary 4.4:

$$\pi_0\,\mathrm{Diff}\left(\frac{S^3}{D'_{2^k(2n+1)}\times\mathbf{Z}_p}\right)$$
$$=\pi_0\,\mathrm{Isom}\left(\frac{S^3}{D'_{2^k(2n+1)}\times\mathbf{Z}_p}\right), \quad p\geqslant 1\,.$$

Proof: See Rubinstein[15] and Theorem 4.3. Q.E.D.

Theorem 4.5: $\pi_0 \operatorname{Isom}\left(\dfrac{S^3}{T'_{8\cdot 3^k}\times \mathbf{Z}_p}\right) = \mathbf{Z}_2$.

Proof: Let $a = [(e^{i\pi/4}, j)] \in \pi_0 \operatorname{Isom}(S^3/G)$ with $G = T'_{8\cdot 3^k}\times \mathbf{Z}_p$. Then $a^2 = [(i,-1)] = [\tilde{a}^{3^k}\tilde{b}\tilde{c}][(i,-1)] = [(1,1)]$, where $\tilde{a} = (\exp[(\pi/3)(i+j+k)/\sqrt{3}], \exp(i\pi/3^k))$, $\tilde{b} = (j,1)$, and $\tilde{c} = (k,1)$. So a is not connected to the identity. Therefore a has order two in $\pi_0 \operatorname{Isom}(S^3/G)$.

There are two types of elements in $\pi_0 \operatorname{Isom}(S^3/G)$, namely $[(t,e^{i\theta})]$, $t\in T^*$, and $[(t,je^{i\theta})]$, $t\in O^* \setminus T^*$. If they are of the form $[(t,e^{i\theta})]$, $t\in T^*$, then $t = \pm 1, \pm i, \pm j, \pm k, \frac{1}{2}(1+i+j+k)$, $\frac{1}{2}(1-i+j+k)$, or $\frac{1}{2}(-1+i+j+k)$, etc.

If t is of the form i, j, k, etc., then let $f(\lambda) = [(t,e^{i(\theta-\lambda\theta)})]$, $f\colon [0,1]\to \operatorname{Isom}(S^3/G)$, $f(0) = [(t,e^{i\theta})]$, and $f(1) = [(t,1)] = [(1,1)]$.

If t is of the form $\frac{1}{2}(1+i+j+k)$, $\frac{1}{2}(1-i+j+k)$, $\frac{1}{2}(1+i-j+k)$, etc., then let $f(\lambda) = [(t,\exp i(\theta - \lambda(\theta + [\pi/3^k]s)))]$, where $s\in\mathbf{Z}$. We have $f(1) = [(t,e^{i\pi s/3^k})] = [(1,1)]$. This implies all elements of the type $[(t,e^{i\theta})]$, $t\in T^*$ are the identity in $\pi_0 \operatorname{Isom}(S^3/G)$.

If $t\in O^* \setminus T^*$, then $[(t,je^{i\theta})] = [(t',e^{i\pi s/3^k})] \times [(e^{i\pi/4}, je^{i\theta'})]$, where $(t,je^{i\theta})\in \tilde{T}$. Hence all elements of the form $[(t,je^{i\theta})]$ are connected to $[(e^{i\pi/4},j)]$. Therefore $\pi_0 \operatorname{Isom}(S^3/G) = \mathbf{Z}_2$. Q.E.D.

Theorem 4.6:

$$\pi_0 \operatorname{Isom}(L(p,q)) = \begin{cases} \mathbf{Z}_2\times\mathbf{Z}_2, & \text{if } q^2 \equiv 1 \bmod p \text{ with } q \not\equiv \pm 1 \bmod p, \\ \mathbf{Z}_4, & \text{if } q^2 \equiv -1 \bmod p \text{ with } p>2, \\ \mathbf{Z}_2, & \text{otherwise,} \end{cases}$$

and

$$\pi_0 \operatorname{Isom}^+(L(p,q)) = \begin{cases} \mathbf{Z}_2\times\mathbf{Z}_2, & \text{if } q^2 \equiv 1 \bmod p \text{ with } q \not\equiv \pm 1 \bmod p, \\ 1, & \text{if } p = 2, \\ \mathbf{Z}_2, & \text{otherwise.} \end{cases}$$

Proof: Suppose $p=2$, then $\operatorname{Isom}(S^3/\mathbf{Z}_2) = \mathrm{O}(4)/\mathbf{Z}_2$ and $\operatorname{Isom}^+(S^3/\mathbf{Z}_2) = \mathrm{SO}(4)/\mathbf{Z}_2$. Because $\mathrm{SO}(4)$ is connected and $\mathbf{Z}_2$ acts freely on $\mathrm{SO}(4)$, $\mathrm{SO}(4)/\mathbf{Z}_2$ is connected. Since $\mathrm{SO}(4)/\mathbf{Z}_2$ is connected, it follows that the only disconnected part of $\operatorname{Isom}(S^3/\mathbf{Z}_2)$ comes from the orientation reversing isometries. Hence, the identity component of $\operatorname{Isom}(S^3/\mathbf{Z}_2)$ is just $\operatorname{Isom}^+(S^3/\mathbf{Z}_2)$. Therefore,

$$\pi_0 \operatorname{Isom}(S^3/\mathbf{Z}_2) = \frac{\mathrm{O}(4)/\mathbf{Z}_2}{\mathrm{SO}(4)/\mathbf{Z}_2} = \frac{\mathrm{O}(4)}{\mathrm{SO}(4)} = \mathbf{Z}_2,$$

and $\pi_0 \operatorname{Isom}^+(S^3/\mathbf{Z}_2) = 1$.

Now suppose $p>2$ and $q^2 \not\equiv -1 \bmod p$. This means there are no orientation reversing isometries (Theorem 3.5), so $\operatorname{Isom}^+(L(p,q)) = \operatorname{Isom}(L(p,q))$. If $q\equiv \pm 1 \bmod p$, then

$$\operatorname{Isom}(L(p,q)) = N_{\mathrm{su}(2)}(S^1)\times \mathrm{SU}(2)/\mathbf{Z}_p.$$

Clearly, the identity component of $N_{\mathrm{SU}(2)}(S^1)\times\mathrm{SU}(2)$ is $S^1\times\mathrm{SU}(2)$ and $\mathbf{Z}_p \leqslant S^1\times S^1 \leqslant S^1\times\mathrm{SU}(2)$. From this, it follows that $S^1\times\mathrm{SU}(2)/\mathbf{Z}_p$ is the identity component of $\operatorname{Isom}(L(p,q))$. Hence,

$$\pi_0 \operatorname{Isom}(L(p,q)) = \frac{N_{\mathrm{SU}(2)}(S^1)\times\mathrm{SU}(2)/\mathbf{Z}_p}{S^1\times\mathrm{SU}(2)/\mathbf{Z}_p} = \frac{N_{\mathrm{SU}(2)}(S^1)\times\mathrm{SU}(2)}{S^1\times\mathrm{SU}(2)} = \frac{N_{\mathrm{SU}(2)}(S^1)}{S^1}\times\frac{\mathrm{SU}(2)}{\mathrm{SU}(2)} = \mathbf{Z}_2,$$

for $q\equiv \pm 1 \bmod p$ with $p>2$. If $q^2\equiv 1 \bmod p$ with $q\not\equiv \pm 1 \bmod p$, then

$$\operatorname{Isom}(L(p,q)) = \frac{N_{\mathrm{SU}(2)}(S^1)\times N_{\mathrm{SU}(2)}(S^1)}{\mathbf{Z}_p}.$$

Here $S^1\times S^1$ is the identity component of $N_{\mathrm{SU}(2)}(S^1)\times N_{\mathrm{SU}(2)}(S^1)$. It follows that $S^1\times S^1/\mathbf{Z}_p$ is the identity component of $N_{\mathrm{SU}(2)}(S^1)\times N_{\mathrm{SU}(2)}(S^1)/\mathbf{Z}_p$. Therefore,

$$\pi_0 \operatorname{Isom}(L(p,q)) = \frac{(N_{\mathrm{SU}(2)}(S^1)\times N_{\mathrm{SU}(2)}(S^1))/\mathbf{Z}_p}{(S^1\times S^1/\mathbf{Z}_p)} = \frac{N_{\mathrm{SU}(2)}(S^1)\times N_{\mathrm{SU}(2)}(S^1)}{S^1\times S^1} = \frac{N_{\mathrm{SU}(2)}(S^1)}{S^1}\times\frac{N_{\mathrm{SU}(2)}(S^1)}{S^1} = \mathbf{Z}_2\times\mathbf{Z}_2,$$

for $q^2\equiv 1 \bmod p$ with $q\not\equiv \pm 1 \bmod p$. If q does not satisfy either $q^2\equiv \pm 1 \bmod p$ or $q\equiv \pm 1 \bmod p$, then

$$\operatorname{Isom}(L(p,q)) = (S^1\times S^1)\cup j(S^1\times S^1)/\mathbf{Z}_p,$$

where $\mathbf{Z}_p \leqslant S^1\times S^1$, and $S^1\times S^1$ is the identity component of $(S^1\times S^1)\cup j(S^1\times S^1)$. Moreover, $(S^1\times S^1)\cup j(S^1\times S^1)$ has only two components, namely $S^1\times S^1$ and $j(S^1\times S^1)$. It follows from the above that $S^1\times S^1/\mathbf{Z}_2$ is the identity component of $\operatorname{Isom}(L(p,q))$. Therefore, $\pi_0 \operatorname{Isom}(L(p,q)) = \mathbf{Z}_2$, for $q^2 \not\equiv \pm 1 \bmod p$.

Finally, if $p>2$ and $q^2\equiv -1 \bmod p$, then

$$\operatorname{Isom}^+(L(p,q)) = (S^1\times S^1)\cup j(S^1\times S^1)/\mathbf{Z}_p.$$

Hence, $\pi_0 \operatorname{Isom}^+(L(p,q)) = \mathbf{Z}_2$. Recall from the proof of Theorem 3.5, that

$$g_1T(z_0,z_1) = (ad^*e^{2\pi iq/p}z_1, -ade^{-2\pi i/p}z_0^*)$$

and

$$g_2T(z_0,z_1) = (-bc^*e^{-2\pi iq/p}z_1^*, bce^{2\pi i/p}z_0)$$

are the two types of orientation reversing isometries. Taking the product of these two elements gives

$$(g_1T)(g_2T)(z_0,z_1) = (ace^{2\pi iq/p}d^*be^{2\pi i/p}z_0, ace^{2\pi iq/p}db^*e^{-2\pi i/p}z_1),$$

where $(g_1T)(g_2T)$ is equivalent to

$$\begin{pmatrix} ace^{2\pi iq/p} & 0 \\ 0 & a^*c^*e^{-2\pi iq/p} \end{pmatrix}\begin{pmatrix} z_0 & z_1 \\ -z_1^* & z_0^* \end{pmatrix}\begin{pmatrix} d^*be^{2\pi i/p} & 0 \\ 0 & db^*e^{-2\pi i/p} \end{pmatrix}.$$

Clearly,

$$\left(\begin{pmatrix} ace^{2\pi iq/p} & 0 \\ 0 & a^*c^*e^{-2\pi iq/p} \end{pmatrix}, \begin{pmatrix} d^*be^{2\pi i/p} & 0 \\ 0 & bd^*e^{-2\pi i/p} \end{pmatrix}\right)$$

is an element of the identity component $S^1\times S^1/\mathbf{Z}_p$ of $\mathrm{Im}(L(p,q))$. Therefore, g_2T is the inverse of g_1T in

$$\begin{pmatrix} 0 & -e^{-2\pi i/p}ad \\ e^{2\pi i/p}a^*d^* & 0 \end{pmatrix}\begin{pmatrix} z_0 & z_1 \\ -z_1^* & z_0^* \end{pmatrix}\begin{pmatrix} 0 & -e^{-2\pi iq/p}a^*d^* \\ e^{2\pi iq/p}ad^* & 0 \end{pmatrix}.$$

Clearly

$$\left(\begin{pmatrix} 0 & -e^{-2\pi i/p}ad \\ e^{2\pi i/p}a^*d^* & 0 \end{pmatrix}, \begin{pmatrix} 0 & e^{-2\pi iq/p}a^*d \\ -e^{2\pi iq/p}ad^* & 0 \end{pmatrix}\right)$$

is an element of $j(S^1\times S^1)$. It follows from the above that $\pi_0\,\mathrm{Isom}(L(p,q))$ is generated by g_1T. Since g_1T squared is in $j(S^1\times S^1)$, it follows that g_1T has the order of 4 in $\pi_0\,\mathrm{Isom}(L(p,q))$. Therefore, $\pi_0\,\mathrm{Isom}(L(p,q))=\mathbf{Z}_4$ for $q^2\equiv -1 \bmod p$ with $p>2$. Q.E.D.

Corollary 4.7: $\pi_0\,\mathrm{Diff}(L(p,q))=\pi_0\,\mathrm{Isom}(L(p,q))$.

Proof: See Hodgson and Rubinstein[17] and Theorem 4.6. Q.E.D.

Corollary 4.8: The only lens spaces with orientation reversing diffeomorphisms are ones with $q^2\equiv -1 \bmod p$.

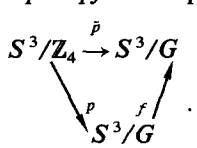
Proof: Corollary 4.7 implies every diffeomorphism is connected to an isometry. Hence $L(p,q)$ has orientation reversing diffeomorphisms if and only if it has orientation reversing isometries, because two diffeomorphisms are connected only if they are both orientation preserving or reversing. The only lens spaces $L(p,q)$ with orientation reversing isometries are ones with $q^2\equiv -1 \bmod p$. Q.E.D.

Corollary 4.9: If G is a noncyclic group which acts freely on S^3, then S^3/G has no orientation reversing diffeomorphisms.

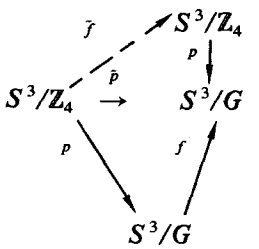
Proof: Let G be a noncyclic group acting freely on S^3. Then Theorem 2.2 implies G is isomorphic to D^*_{4n}, T^*, O^*, I^*, $D'_{2^k(2n+1)}$, $T'_{8\cdot 3^k}$, or the direct product of one of these with a cyclic group of relatively prime order. Each of these groups contains a $\mathbf{Z}_4$ subgroup. We look at the presentations of these groups in Sec. II and observe that a $\mathbf{Z}_4$ subgroup is generated by $x^{2^{k-2}}$ for $D'_{2^k(2n+1)}$ and by x in the other cases. Using the relations given in these group presentations, it can be shown that the $\mathbf{Z}_4$ generated in the above way is conjugate to any other $\mathbf{Z}_4$ contained in these groups.

Given a free action of a noncyclic group G on S^3 there is an induced free $\mathbf{Z}_4$ action on S^3, because $\mathbf{Z}_4\leqslant G$. From the above, it follows that the map $p\colon S^3/\mathbf{Z}_4\to S^3/G$ defined by $p([\chi])=[\chi]_G$ is a well-defined covering map. (Here $[\chi]$ denotes the equivalence class of $\chi\in S^3$ with respect to the $\mathbf{Z}_4$ action and $[\chi]_G$ denotes the equivalence class in S^3/G.)

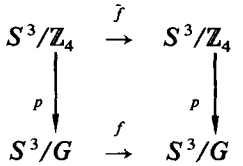
Now, let $f\colon S^3/G\to S^3/G$ be a diffeomorphism of S^3/G. Then $\tilde{p}=pf$ is a map of $S^3/\mathbf{Z}_4$ into S^3/G



From the lifting theorem, Theorem 2.4, $\tilde{p}$ lifts to a map $\tilde{f}$ such that



commutes, because all $\mathbf{Z}_4$ subgroups of G are conjugate to each other in G. Therefore, every diffeomorphism of S^3/Gf lifts to $\tilde{f}$ such that the diagram



commutes. The degree[28] of the composition of maps is given by $\deg(gh)=\deg(g)\deg(h)$ (see Ref. 29, p. 268). Applying this to the above commuting diagram implies $\deg(f)\deg(p)=\deg(p)\deg(\tilde{f})$. Hence $\deg(f)=\deg(\tilde{f})$.

Now, suppose $f\colon S^3/G\to S^3/G$ is an orientation reversing diffeomorphism. Hence $\deg(f)=-1$. Therefore, $\deg(\tilde{f})=-1$, which implies $S^3/\mathbf{Z}_4$ has orientation reversing diffeomorphisms, when $S^3/\mathbf{Z}_4$ is just a lens space $L(4,q)$ (see Rubinstein[15]). The only allowed values of q are 1 and 3. Neither q value satisfies $q^2\equiv -1 \bmod p$. Hence Corollary 4.8 implies $S^3/\mathbf{Z}_4$ has no orientation reversing diffeomorphisms, a contradiction. Therefore, S^3/G cannot have any orientation reversing diffeomorphisms. Q.E.D.

Table II summarizes the results of this section. Under the column labeled conjecture, "W" means the weak conjecture holds, "S" means the strong conjecture holds, and "open" means it is not known at the present time if the conjecture is true. All results given in Table II are independent of the conjecture.

V. $\pi_0\,\mathrm{Diff}_{x_0}(S^3/G)$

Let M be a connected manifold. Define a projection p from $\mathrm{Diff}(M)$ to M by $p(f)=f(x_0)$. This projection produces a fiber bundle with total space $\mathrm{Diff}(M)$, base space M, and fiber $\mathrm{Diff}_{x_0}(M)$. Since it is a fiber bundle, we have the following exact sequence of groups:

$$\cdots\to \pi_i \operatorname{Diff}_{x_0}(M)\to\pi_i \operatorname{Diff}(M)\to\pi_i M$$
$$\to \pi_{i-1} \operatorname{Diff}_{x_0}(M)\to\cdots\to\pi_0 \operatorname{Diff}_{x_0}(M)$$
$$\to \pi_0 \operatorname{Diff}(M)\to 1 .$$

The only part of this sequence we use is

$$\pi_2 M\to\pi_1 \operatorname{Diff}_{x_0}(M)\to\pi_1 \operatorname{Diff}(M)\to\pi_1 M$$
$$\to \pi_0 \operatorname{Diff}_{x_0}(M)\to\pi_0 \operatorname{Diff}(M)\to 1 .$$

Associated with a manifold of the form S^3/G is the fiber bundle $p: S^3\to S^3/G$ and the sequence

$$\cdots\to \pi_i G\to\pi_i S^3\to\pi_i S^3/G\to \pi_{i-1} G$$
$$\to\cdots\to \pi_0 G\to\pi_0 S^3\to \pi_0 S^3/G ,$$

where $\pi_0 G = G$ and $\pi_i G = 1$ for all $i\geqslant 1$, since G is finite. Hence $1\to \pi_i S^3\to \pi_i S^3/G\to 1$, for all $i\geqslant 2$ and $\pi_1 S^3/G = G$. This implies that $\pi_2 S^3/G = \pi_2 S^3 = 1$. The exact sequence we are interested in is

$$1\to \pi_1 \operatorname{Diff}_{x_0}(M)\to \pi_1 \operatorname{Diff}(M)\xrightarrow{p_{\#_1}} G$$
$$\xrightarrow{\partial_{\#_1}} \pi_0 \operatorname{Diff}_{x_0}(M)\xrightarrow{i_{\#_0}} \pi_0 \operatorname{Diff}(M)\to 1 .$$

The above sequence is exact, so $\ker i_{\#_0} = \operatorname{Im}\partial_{\#_1}$, and $\ker\partial_{\#_1} = \operatorname{Im} p_{\#_1}$. Combining this with the fact that the image of a homomorphism is the domain modulo the kernel we have

$$\ker i_{\#_0} = \operatorname{Im}\partial_{\#_1} = G/\ker\partial_{\#_1} = G/\operatorname{Im} p_{\#_1} .$$

The sequence

$$1\to\ker i_{\#_0}\to \pi_0 \operatorname{Diff}_{x_0}(M)\to \pi_0 \operatorname{Diff}(M)\to 1$$

is exact. Since

$$\ker i_{\#_0} = G/\operatorname{Im} p_{\#_1} ,$$

we have

$$1\to G/\operatorname{Im} p_{\#_1}\to \pi_0 \operatorname{Diff}_{x_0}(M)\xrightarrow{i_{\#_0}} \pi_0 \operatorname{Diff}(M)\to 1 \qquad (5.1)$$

is exact. In order to calculate $\operatorname{Im} p_{\#_1}$ and $\pi_0 \operatorname{Diff}_{x_0}(M)$ it is necessary to find out another relation between $\pi_1(M,x_0)$ and $\pi_0 \operatorname{Diff}_{x_0}(M)$.

Given any homeomorphism between a manifold and itself $f: M\to M$ that fixes a point x_0, i.e., $f(x_0) = x_0$, there is an induced automorphism of $\pi_1(M,x_0)$ denoted by $f_{\#_1}$. Further, if two such homeomorphisms are homotopic, then the induced automorphisms are the same. Hence γ: $\pi_0 \operatorname{Diff}_{x_0}(M)$ $\to$ Aut $\pi_1(M,x_0)$, where $\gamma([f]) = f_{\#_1}$ is a well defined homeomorphism. Suppose $\gamma([f]) = \gamma([g])$, where both f and g are orientation preserving and $M = S^3/G$. Clearly, $f_{\#_1} = g_{\#_1}$. This combined with the following lemma implies $f\simeq g$.

Lemma 5.1: If $f, g: S^3/G\to S^3/G$ are orientation preserving and $f_{\#_1} = g_{\#_1}$, then $f\simeq g$.

Proof: S^3/G can be thought of as a CW complex. Let K^1 be its one-skelton, $f|_{K^1}\simeq g|_{K^1}$ because $f_{\#_1} = g_{\#_1}$ (see Ref. 25, page 194). Now $f|_{K^2}\simeq g|_{K^2}$ because the obstruction to deforming f into g restricted to the two-skelton lies in the second cohomology group of S^3/G with local coefficients in $\pi_2(S^3/G) = 0$ (see Ref. 25, pp. 183 and 193). Finally, $f|_{K^3}\simeq g|_{K^3}$ if and only if there is no obstruction to deforming f into g in the third cohomology group of S^3/G with local coefficients in $\pi_3(S^3/G) = \mathbf{Z}$. The obstruction is just the difference of the cohomology classes induced by the two maps f and g. Since both f and g are orientation preserving, their degrees are both equal to $+1$, and they both must determine the same class in $H^3(S^3/G,\mathbf{Z})$. Therefore the obstruction is zero. Since K^3 is S^3/G, f is homotopic to g. Q.E.D.

Homotopy implies isotopy when the weak conjecture holds; therefore $\gamma([f]) = \gamma([g])$ implies $[f] = [g]$ if we look at γ: $\operatorname{Diff}^+_{x_0}(M)\to$ Aut $\pi_1(M,x_0)$. Therefore γ is one to one on $\pi_0 \operatorname{Diff}^+_{x_0}(M)$. The fact that γ is one to one on $\pi_0 \operatorname{Diff}^+_{x_0}(M)$ means Im γ is a subgroup of Aut $\pi_1(M,x_0)$.

Given $[f]\in\operatorname{Diff}^+(M)$ there is an induced homomorphism $f_{\#_1}$: $\pi_1(M,x_0)\to\pi_1(M,f(x_0))$. Since in general an element of $\operatorname{Diff}^+ M$ does not fix x_0, $f_{\#_1}$ is not necessarily an automorphism of $\pi_1(M,x_0)$. By moving the base point $f(x_0)$ to x_0 via conjugation by an element of fundamental group, i.e., an inner automorphism, $\pi_1(M,f(x_0))$ can be mapped to $\pi_1(M,x_0)$., Therefore it follows that $\tilde\gamma$: $\pi_0 \operatorname{Diff}^+(M)\to$ Out $\pi_1(M,x_0)$ where $\tilde\gamma$ is γ after moding out by Inn $\pi_1(M,x_0)$. It follows that $\pi_0 \operatorname{Diff}^+(M)$ is a subgroup of Out $\pi_1(M,x_0)$.

We now use the above information to find $\operatorname{Im} p_{\#_0}$ in sequence (5.1). The homomorphism $\partial_{\#_1}$ maps each $g\in\pi_1(M,x_0)$ to an equivalence class of diffeomorphisms in $\pi_0 \operatorname{Diff}^+_{x_0}(M)$. By exactness $i_{\#_0}\partial_{\#_1} = 1$; hence the action of $\partial_{\#_1}(g)$ as an outer automorphism of $\pi_1(M,x_0)$ is trivial. It follows that the action of $\partial_{\#_1}(g)$ on $\pi_1(M,x_0)$ is given by the inner automorphism $\partial_{\#_1}(g)h = g'hg'^{-1}$, where for a fixed $g\in\pi_1(M,x_0)$, g' is a fixed element of $\pi_1(M,x_0)$ and h is an element of $\pi_1(M,x_0)$. Therefore, the image of $\pi_1(M,x_0)$ in $\pi_0 \operatorname{Diff}^+_{x_0}(M)$ is contained in Inn $\pi_1(M,x_0)$. The group of inner automorphisms Inn $\pi_1(M,x_0)$ is equal to

$$\pi_1(M,x_0)/\operatorname{Center}(\pi_1(M,x_0)).$$

Since the image of $\pi_1(M,x_0)$ is contained in Inn $\pi_1(M,x_0)$, it follows from Theorem 2.1 that

$$\operatorname{Im}\partial_{\#_1} = A/\operatorname{Center}(\pi_1(M,x_0)) ,$$

where $A\leqslant\pi_1(M,x_0)$. Hence

$$\operatorname{Im}\partial_{\#_1} = \frac{\pi_1(M,x_0)}{\operatorname{Im} p_{\#_1}} = \frac{A}{\operatorname{Center}(\pi_1(M,x_0))} .$$

Suppose $A\neq\pi_1(M,x_0)$, then

$$|A| < |\pi_1(M,x_0)| ,$$
$$\operatorname{Im} p_{\#_1}\leqslant\operatorname{Center}(\pi_1(M,x_0))$$

(See Ref. 30). Hence

$$|\operatorname{Im} p_{\#_1}|\leqslant|\operatorname{Center}(\pi_1(M,x_0))| .$$

Therefore

$$|A|\,|\operatorname{Im} p_{\#_1}| < |\operatorname{Center}(\pi_1(M,x_0))|\,|\pi_1(M,x_0)| ,$$

a contradiction to

$$\frac{A}{\operatorname{Center}(\pi_1(M,x_0))} = \frac{\pi_1(M,x_0)}{\operatorname{Im} p_{\#_1}} .$$

Therefore $A = \pi_1(M,x_0)$ and $\operatorname{Im} p_{\#_1} = \operatorname{Center}(\pi_1(M,x_0))$. So sequence (5.1) becomes

$$1\to\operatorname{Inn}(\pi_1(M,x_0))\to \pi_0 \operatorname{Diff}^+_{x_0}(M)\to \pi_0 \operatorname{Diff}^+(M)\to 1 ,$$

where

$$\pi_0 \operatorname{Diff}^+_{x_0}(M) \leqslant \operatorname{Aut}(\pi_1(M,x_0))$$

and

$$\pi_0 \operatorname{Diff}^+(M) \leqslant \operatorname{Out}(\pi_1(M,x_0)).$$

Theorem 5.2:

$$\pi_0 \operatorname{Diff}^+_{x_0}(S^3/G) = \begin{cases} 0, & \text{if } G = D^*_8, T^*, \text{ or } O^*, \\ D_{4n}, & \text{if } G = D^*_{4n},\ n \geqslant 3, \\ I, & \text{if } G = I^*, \\ O \times \mathbb{Z}_2, & \text{if } G = D^*_8 \times \mathbb{Z}_p, T^* \times \mathbb{Z}_p, \\ & \quad \text{or } O^* \times \mathbb{Z}_p, \\ I \times \mathbb{Z}_2, & \text{if } G = I^* \times \mathbb{Z}_p, \\ D_{4n} \times \mathbb{Z}_2, & \text{if } G = D^*_{4n} \times \mathbb{Z}_p,\ n \geqslant 3. \end{cases}$$

Proof: If $G = O^*$ or I^*, Theorem 4.1 tells us that $\pi_0 \operatorname{Diff}(S^3/G) = 1$. Hence our sequence is $1 \to \operatorname{Inn} G \to \pi_0 \operatorname{Diff}_{x_0}(M) \to 1$, and $\pi_0 \operatorname{Diff}_{x_0}(M) = \operatorname{Inn} G$.

If $G = D^*_8$ or T^*, the sequence is $1 \to \operatorname{Inn} G \to \pi_0 \operatorname{Diff}_{x_0}(M) \to \operatorname{Out} G \to 1$ because $\operatorname{Out} D^*_8 = P_3$ and $\operatorname{Out} T^* = \mathbb{Z}_2$, and Theorem 3.1 gives $\pi_0 \operatorname{Diff}(S^3/D^*_8) = P_3$ and $\pi_0 \operatorname{Diff}(S^3/T^*) = \mathbb{Z}_2$. Hence $\pi_0 \operatorname{Diff}_{x_0}(M) = \operatorname{Aut} G$, where $\operatorname{Aut} D^*_8 = 0$ and $\operatorname{Aut} T^* = 0$. Therefore, $\pi_0 \operatorname{Diff}_{x_0}(M) = 0$ for $G = D^*_8$ or T^*.

If $G = D^*_{4n} = \langle x, y : x^{2n} = 1,\ y^4 = 1,\ yxy^{-1} = x^{-1} \rangle$, $n \geqslant 3$, then Theorem 4.1 implies $\pi_0 \operatorname{Diff}(M) = \mathbb{Z}_2$. The action of $\pi_0 \operatorname{Diff}(M)$ on $\pi_1(M,x_0)$ is given by $e^{\pi i/2n} A e^{-\pi i/2n} = A$ and $e^{\pi i/2n} B e^{-\pi i/2n} = AB$, where $A = e^{\pi i/n}$ and $B = j$. All the elements of D^*_{4n} are of form x^r and $x^r y$. To find the action of $\operatorname{Inn} D^*_{4n}$ on D^*_{4n}, just conjugate the generators by the elements of D^*_{4n}:

$$x^r x x^{-r} = x, \quad x^r y x^{-r} = x^{2r} y,$$

and

$$x^r y x y^{-1} x^{-r} = x^{-1}, \quad x^r y y y^{-1} x^{-r} = x^{2r} y.$$

Clearly all inner automorphisms are generated by

$$a = \begin{cases} f(x) = x, \\ f(y) = x^2 y, \end{cases}$$

and

$$b = \begin{cases} h(x) = x^{-1}, \\ h(y) = y, \end{cases}$$

where a and b satisfy the relations $a^n = 1 = b^2$, and $bab^{-1} = a^{-1}$. The action of $\pi_0 \operatorname{Diff}(M)$ on D^*_{4n} is given by

$$d = \begin{cases} k(x) = x, \\ k(y) = xy, \end{cases}$$

$$d^2 = \begin{cases} kk(x) = k(x) = x, \\ kk(y) = k(xy) = k(x)k(y) = x^2 y, \end{cases}$$

$$bdb^{-1}d = \begin{cases} hkhk(x) = hkh(x) = hk(x^{-1}) \\ \qquad = h(k(x)^{-1}) = h(x^{-1}) = x, \\ hkhk(y) = hkh(xy) = hk(x^{-1}y) \\ \qquad = h(k(x)^{-1} k(y)) = h(x^{-1}xy) = y. \end{cases}$$

Hence $d^2 = a$ and $bdb^{-1}d = 1$. Therefore, $\pi_0 \operatorname{Diff}_{x_0}(M) = D_{4n}$ for $n \geqslant 3$. Q.E.D.

Theorem 5.3:

$$\pi_0 \operatorname{Diff}^+_{x_0}\left(\frac{S^3}{D'_{2^k(2n+1)} \times \mathbb{Z}_p}\right) = D_{4(2n+1)} \times \mathbb{Z}_2.$$

Proof: Recall $D'_{2^k(2n+1)} = \langle x, y : x^{2^k} = 1,\ y^{2n+1} = 1,\ xyx^{-1} = y^{-1} \rangle$. Every element of $D'_{2^k(2n+1)}$ is of the form $y^s x^r$, where s and r are integers. If m is even, then $(y^r x^s)^m = y(rm/2)(1 + (-1)^s) x^{sm}$, and if m is odd, then $(y^r x^s)^m = y(r/2)(m + 1 + (m-1)(-1)^s) x^{sm}$. The center of $D'_{2^k(2n+1)}$ is the subgroup with elements $y^s x^r$ such that $(y^s x^r) y^p x^q (x^{-r} y^{-s}) = y^p x^q$, for all p and q, and

$$\begin{aligned} y^s x^r y^p x^q x^{-r} y^{-s} &= y^s x^r y^p x^q x^{-r} y^{-s} = y^s y^{p(-1)^r} x^q y^{-s} \\ &= y^s y^{p(-1)^r} y^{-s(-1)^q} x^q \\ &= y^{s(1-(-1)^q)} y^{p(-1)^r} x^q \\ &= y^p x^q, \end{aligned} \tag{5.2}$$

for all p and q. Expression (5.2) holds only when $s = 0$ and r is even. Therefore, the center is generated by x^2 and is a cyclic group of order 2^{k-1}. It follows that the center of $D'_{2^k(2n+1)} \times \mathbb{Z}_p$ is $\mathbb{Z}_{2^{k-1}} \times \mathbb{Z}_p$. The group

$$\operatorname{Inn}(D'_{2^k(2n+1)} \times \mathbb{Z}_p) = \frac{D'_{2^k(2n+1)} \times \mathbb{Z}_p}{\mathbb{Z}_{2^{k-1}} \times \mathbb{Z}_p}$$

is just $D_{2(2n+1)}$, since $(D'_{2^k(2n+1)} \times \mathbb{Z}_p)/\mathbb{Z}_{2^{k-1}} \times \mathbb{Z}_p$ has two distinct types of elements y^p and $y^p x \bmod \mathbb{Z}_{2^{k-1}} \times \mathbb{Z}_p$. The explicit form of inner automorphisms is found by conjugation on the generators x and y by elements of $D'_{2^k(2n+1)} \times \mathbb{Z}_p$. The $\mathbb{Z}_p$ part does nothing since it commutes so we only need to look at elements in $D'_{2^k(2n+1)}$. The general element in $D'_{2^k(2n+1)}$ is of the form $y^p x^q$ so $(y^p x^q) x (x^{-q} y^{-p}) = y^{2p} x$ and $(y^p x^q) y (x^{-q} y^{-p}) = y^{(-1)^q}$.

Therefore, all inner automorphisms are of the form $f(x) = y^{2p} x$ and $f(y) = y^{\pm 1}$, where p is an integer. Let

$$a = \begin{cases} f(x) = y^2 x, \\ f(y) = y, \end{cases} \quad \text{and} \quad b = \begin{cases} g(x) = x, \\ g(y) = y^{-1}. \end{cases}$$

Then a and b generate $\operatorname{Inn} D'_{2^k(2n+1)}$, $b^2 = 1$, $a^{2n+1} = 1$, and $bab^{-1} = a^{-1}$.

The group

$$\pi_0 \operatorname{Diff}(S^3/D'_{2^k(2n+1)} \times \mathbb{Z}_p) = \mathbb{Z}_2 \times \mathbb{Z}_2$$

is generated by $(e^{i\pi/2(2n+1)}, 1)$ and $(1, j)$, see Theorem 3.3. The action of $(e^{i\pi/2(2n+1)}, 1)$ is given by

$$\begin{aligned} &(e^{i\pi/2(2n+1)}, 1)\ \tilde{a}\ (e^{-i\pi/2(2n+1)}, 1) \\ &\quad = (e^{i\pi/(2n+1)}, -1) = \tilde{a}, \\ &(e^{i\pi/2(2n+1)}, 1)\ \tilde{b}\ (e^{-i\pi/2(2n+1)}, 1) \\ &\quad = (e^{i\pi/2(2n+1)} j e^{-i\pi/2(2n+1)}, e^{i\pi/2^{k-1}}) = \tilde{a}\tilde{b}^{2^{k-1}} \tilde{b}, \end{aligned}$$

where $\tilde{a} = (e^{i\pi/(2n+1)}, -1)$ and $\tilde{b} = (j, e^{i\pi/2^{k-1}})$. Here $g(\tilde{a}) = \tilde{a}$ and $g(\tilde{b}) = \tilde{a}\tilde{b}^{2^{k-1}} \tilde{b}$. The action of $(1, j)$ is given by

$$\begin{aligned} &(1,j)\tilde{a}(1,-j) = (e^{i\pi/(2n+1)}, -1) = \tilde{a}, \\ &(1,j)\tilde{b}(1,-j) = \tilde{a}^{2n+1} \tilde{b}^{2^{k-1}} \tilde{b}^{-1}. \end{aligned}$$

Let $h(\tilde{a}) = \tilde{a}$ and $h(\tilde{b}) = \tilde{a}^{2n+1} \tilde{b}^{2^{k-1}} \tilde{b}^{-1}$, then

$$c=\begin{cases}h(x)=x^{2^{k-1}}x^{-1},\\ h(y)=y,\end{cases}\quad d=\begin{cases}g(x)=yx^{2^{k-1}}x,\\ g(y)=y.\end{cases}$$

Now the relations between the generators are derived:

$$ca=\begin{cases}hf(x)=h(y^2x)=(h(y))^2h(x)=y^2x^{2^{k-1}}x^{-1},\\ hf(y)=h(y)=y,\end{cases}$$

$$ac=\begin{cases}fh(x)=f(x^{2^{k-1}}x^{-1})=(y^2x)^{2^{k-1}-1}\\ \quad=y^{(2^{k-1}-1+1+(2^{k-1}-1-1)(-1))}x^{2^{k-1}}x^{-1}\\ \quad=y^2x^{2^{k-1}}x^{-1},\\ fh(y)=f(y)=y,\end{cases}$$

$$cb=\begin{cases}hg(x)=h(x)=x^{2^{k-1}}x^{-1},\\ hg(y)=y^{-1},\end{cases}$$

$$bc=\begin{cases}gh(x)=g(x^{2^{k-1}}x^{-1})=x^{2^{k-1}}x^{-1},\\ gh(y)=g(y)=y^{-1},\end{cases}$$

$$d^2=\begin{cases}kk(x)=k(yx^{2^{k-1}+1})=k(y)(k(x))^{2^{k-1}+1}\\ \quad=yy^{(1/2)(2^{k-1}+1+1+(2^{k-1}+1-1)(-1)^{2^{k-1}+1})}\\ \quad\times x^{(2^{k-1}+1)^2},\\ kk(y)=k(y)=y,\end{cases}$$

$$d^2=\begin{cases}kk(x)=yy^{(1/2)2}x^{(2^{k-1}+1)^2}=y^2x,\\ kk(y)=y,\end{cases}$$

$$c^2=\begin{cases}hh(x)=h(x^{2^{k-1}-1})=(x^{2^{k-1}-1})^{2^{k-1}-1}=x,\\ hh(g)=h(y)=y,\end{cases}$$

$$dc=\begin{cases}hh(x)=h(yx^{2^{k-1}+1})=y(x^{2^{k-1}-1})^{2^{k-1}+1}\\ \quad=yx^{-1},\\ hk(y)=h(y)=y,\end{cases}$$

$$cd=\begin{cases}kh(x)=k(x^{2^{k-1}+1})=y(x^{2^{k-1}-1})^{2^{k-1}+1}\\ \quad=yx^{-1}\\ kh(y)k(y)=y\end{cases}$$

$$bdb^{-1}d=\begin{cases}gkgk(x)=gkg(yx^{2^{k-1}+1})\\ \quad=gk(g(y)(g(x))^{2^{k-1}+1})\\ \quad=gk(y^{-1}x^{2^{k-1}+1}),\\ gkgk(y)=gky(y)=gk(y^{-1})=g(y^{-1})\\ \quad=y,\end{cases}$$

$$bdb^{-1}d=\begin{cases}gkgk(x)=gk(y^{-1}x^{2^{k-1}+1})\\ \quad=g((k(y))^{-1}(k(x))^{2^{k-1}+1})\\ \quad=g(y^{-1}(yx^{2^{k-1}+1})^{2^{k-1}+1})\\ \quad=g(y^{-1}yx)=x,\\ ghgk(y)=y.\end{cases}$$

Hence, $ca=ac$, $bc=cb$, $d^2=a$, $c^2=1$, $dc=cd$, and $bdb^{-1}d=1$. Therefore, $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)=D_{4(2n+1)}\times Z_2$.

The above technique used in Theorem 5.2 and 5.3 will work for any space S^3/G, but it becomes quite lengthy when the group G is complicated as in Theorem 5.3. There is another technique that is simple to apply when the group Inn G has no center, as is the case for $T'_{8\cdot 3^k}$.

Lemma 5.4: Given a centerless group G and another group Γ, there is a 1–1 correspondence between inequivalent exact sequences E of the form

$$E:\ 1\to G\to P\to\Gamma\to 1$$

and elements of $\mathrm{Hom}(\Gamma,\mathrm{Out}(G))$. Two sequences E and E' are equivalent if and only if there exists a homomorphism ρ such that

$$\begin{array}{ccccccc}E:\ 1\to & G & \to & P & \to & \Gamma & \to 1\\ & {\scriptstyle \mathrm{id}_G}\Big\| & & {\scriptstyle\rho}\Big\downarrow & & \Big\|{\scriptstyle \mathrm{id}_\Gamma} & \\ E':\ 1\to & G & \to & P' & \to & \Gamma & \to 1\end{array}$$

commutes.

Proof: Let E: $1\to G\xrightarrow{f}P'\xrightarrow{h}\Gamma\to 1$ be an exact sequence of groups. Exactness implies $f(G)$ is a normal subgroup of P'. Since $f(G)\trianglelefteq P'$ and $G\approx f(G)$, it follows that $\psi(p)g\equiv f^{-1}(pf(g)p^{-1})$ is an automorphism of G for each $p\in P'$. This implies $\psi: P'\to\mathrm{Aut}(G)$ is a group homomorphism and $\psi(f(G))=\mathrm{Inn}(G)$. Let $\gamma\in\Gamma$, define $\varphi(\gamma)$ to be $\psi(pf(G))$, where $h(p)=\gamma$. Then $\psi(pf(G))=\psi(p)\psi(f(G))$, and $\psi(f(G))=\mathrm{Inn}(G)$ implies $\varphi(\gamma)$ is an outer automorphism for each $\gamma\in\Gamma$. It follows that φ: $\Gamma\to\mathrm{Out}(G)$ is a group homomorphism, i.e., $\varphi\in(\mathrm{Hom}(\Gamma,\mathrm{Out}(G))$.

Now, suppose $\varphi\in\mathrm{Hom}(\Gamma,\mathrm{Out}(G))$ and G is centerless. For a centerless group the exact sequence $1\to\mathrm{Center}(G)\to G\xrightarrow{\tau}\mathrm{Aut}(G)\to\mathrm{Out}(G)\to 1$ reduces to $1\to G\xrightarrow{\tau}\mathrm{Aut}(G)\to\mathrm{Out}(G)\to 1$, where $\tau(g)h=ghg^{-1}$. For $\varphi\in\mathrm{Hom}(\Gamma,\mathrm{Out}(G))$ construct the pullback P

$$\begin{array}{ccccccc} & & G & & P & \xrightarrow{p_2} & \Gamma\\ & & {\scriptstyle\mathrm{id}_G}\Big\| & & {\scriptstyle p_1}\Big\downarrow & & \Big\downarrow\\ 1\to & G & \xrightarrow{\tau} & \mathrm{Aut}\ (G) & \xrightarrow{\delta} & \mathrm{Out}\ (G), & \end{array}\qquad(5.3)$$

$$P\equiv\{(a,\gamma)\in\mathrm{Aut}(G)\times\Gamma\,|\,\delta(a)=\varphi(\gamma)\},$$

$$p_1(a,\gamma)=a,\ \text{and}\ p_2(a,\gamma)=\gamma$$

Now

$$\delta(aa')=\delta(a)\delta(a')=\varphi(a)\varphi(a')=\varphi(aa').$$

Hence, (a,γ), $(a',\gamma')\in P$ implies $(aa',\gamma\gamma')\in P$. Thus $\delta(a^{-1})=(\delta(a))^{-1}=(\varphi(\gamma))^{-1}=\varphi(\gamma^{-1})$. Hence $(a,\gamma)\in P$ implies $(a^{-1},\gamma^{-1})=(a,\gamma)^{-1}\in P$, and $\delta(1)=\varphi(1)$ implies $(1,1)\in P$. Therefore P is a group and it follows that p_1 and p_2 are group homomorphisms, which makes diagram (5.3) commute. Now, fill in the map from G to P in the diagram

$$\begin{array}{ccccccc} & & G & \overset{i}{\dashrightarrow} & P & \xrightarrow{p_2} & \Gamma\\ & & {\scriptstyle\mathrm{id}_G}\Big\| & & {\scriptstyle p_1}\Big\downarrow & & \Big\downarrow\\ 1\to & G & \xrightarrow{\tau} & \mathrm{Aut}(G) & \xrightarrow{\delta} & \mathrm{Out}(G) & \to 1,\end{array}$$

where $i(g)\equiv(\tau(g),1)$. That the left square commutes is im-

plied by $\tau(\mathrm{id}_G(g)) = \tau(g) = p_2 i(g)$, where $p_2 i(g) = p_2((\tau(g),1)) = 1$. Now find all $(a,\gamma) \in P$ such that $p_2(a,\gamma) = 1$; $p_2(a,\gamma) = 1$ implies $\gamma = 1$; $\delta(a) = \varphi(1) = 1$; $\delta(a) = 1$ if and only if $a \in \mathrm{Inn}(G)$. Hence, $\ker p_2 = i(G) = \mathrm{Im}\, i$. It follows that the diagram

$$\begin{array}{ccccccccc} 1 & \to & G & \xrightarrow{i} & P & \xrightarrow{p_2} & \Gamma & \to 1 \\ & & \| & & \downarrow & & \downarrow & \\ & & G & \xrightarrow{\tau} & \mathrm{Aut}(G) & \xrightarrow{\delta} & \mathrm{Out}(G) & \to 1 \end{array}$$

commute and the top row is exact. Therefore, given $\varphi \in \mathrm{Hom}(\Gamma, \mathrm{Out}(G))$, there is an exact sequence E.

Now, suppose E': $1 \to G \to P' \to \Gamma \to 1$ is an exact sequence. Let $\lambda(p') \equiv (\psi(p'), h(p'))$, where ψ: $P' \to \mathrm{Aut}(G)$ as defined above. Then $\delta(\psi(p')) = \alpha(h(p'))$ implies λ: $P' \to P$. Clearly

$$\begin{array}{ccccccccc} 1 & \to & G & \to & P' & \to & \Gamma & \to 1 \\ & & \| & & \downarrow \lambda & & \| & \\ & & G & \to & P & \to & \Gamma & \to 1 \end{array}$$

commutes. Q.E.D.

Using Lemma 5.4, the group $\pi_0 \mathrm{Diff}^+_{x_0}(S^3/T'_{8\cdot 3^k} \times \mathbf{Z}_p)$ is calculated.

Theorem 5.5: $\pi_0 \mathrm{Diff}^+_{x_0}(S^3/T'_{8\cdot 3^k} \times \mathbf{Z}_p) = 0$.

Proof: The center of $\tilde{T}$ for $p = 1$ is generated by $\tilde{a}^3 = (\exp[(\pi/3)(i+j+k)/\sqrt{3}], \exp(i\pi/3^k))^3$ and $\tilde{b}^2 = (j,1)^2$. Hence the center of $T'_{8\cdot 3^k}$ is generated by A^2 and B^2. The center of $T'_{8\cdot 3^k}$ is therefore equal to $\mathbf{Z}_{2\cdot 3^{k-1}}$:

$$\begin{aligned} \mathrm{Inn}\, T'_{8\cdot 3^k} &= \frac{T'_{8\cdot 3^k}}{\mathbf{Z}_{2\cdot 3^k}} \\ &= \langle \hat{A}, \hat{B} \colon \hat{A}^3 = 1, \hat{B}^2 = (\hat{B}\hat{C})^2 = \hat{C}^2 = 1, \\ &\qquad \hat{A}\hat{B}\hat{A}^{-1} = \hat{C},\ \hat{A}\hat{C}\hat{A} = \hat{B}\hat{C} \rangle . \end{aligned}$$

Therefore $\mathrm{Inn}\, T'_{8\cdot 3^k} = T$. Theorem 4.5 tells us $\pi_0 \mathrm{Diff}(S^3/T'_{8\cdot 3^k}) = \mathbf{Z}_2$. Hence, sequence (5.1) becomes $1 \to T \to \pi_0 \mathrm{Diff}_{x_0}(M) \to \mathbf{Z}_2 \to 1$. Here T is centerless, hence the number of choices is $|\mathrm{Hom}(\mathbf{Z}_2, \mathrm{Out}\, T)| = |\mathrm{Hom}(\mathbf{Z}_2, \mathbf{Z}_2)| = 2$. One choice is $1 \to T \to T \times \mathbf{Z}_2 \to \mathbf{Z}_2 \to 1$, another is $1 \to T \to 0 \to \mathbf{Z}_2 \to 1$. The action of $\pi_0 \mathrm{Diff}(S^3/T'_{8\cdot 3^k})$ on $\pi_1(S^3/T'_{8\cdot 3^k}, x_0)$ is given by

$$z = \begin{cases} A \to A^{-1}BC, \\ B \to C, \\ C \to B^3 \end{cases}$$

(see the proof of Theorem 3.4). Here z does not commute with every inner automorphism, for example, let

$$f = \begin{cases} A \to A, \\ B \to C, \\ C \to BC, \end{cases} \qquad f \in \mathrm{Inn}\, T'_{8\cdot 3^k},$$

conjugation by A, $zf \neq fz$. Hence the choice is $1 \to T \to 0 \to \mathbf{Z}_2 \to 1$.

If $\pi_1(M, x_0) = T'_{8\cdot 3^k} \times \mathbf{Z}_p$, then the center is center $(T'_{8\cdot 3^k}) \times \mathbf{Z}_p$ and $\pi_0 \mathrm{Diff}(S^3/T'_{8\cdot 3^k} \times \mathbf{Z}_p) = \mathbf{Z}_2$, so again we have $1 \to T \to \pi_0 \mathrm{Diff}_{x_0}(M) \to \mathbf{Z}_2 \to 1$ with the $\mathbf{Z}_2$ action being nontrivial. So we have $\pi_0(S^3/T'_{8\cdot 3^k} \times \mathbf{Z}_p) = 0$. Q.E.D.

Theorem 5.6: $\pi_0 \mathrm{Diff}^+(L(p,q)) = \pi_0 \mathrm{Diff}^+_{x_0}(L(p,q))$.

Proof: Since the center of $\pi_1 L(p,q)$ equals the whole group, we have

$$\mathrm{Inn}\, \pi_1(L(p,q)) = \pi_1(L(p,q))/\pi_1(L(p,q)) = 1 .$$

Hence

$$1 \to 1 \to \pi_0 \mathrm{Diff}^+_{x_0}(L(p,q)) \to \pi_0 \mathrm{Diff}^+(L(p,q)) \to 1 .$$

Q.E.D.

Table III gives the zeroth homotopy groups of the group of orientation preserving diffeomorphisms fixing a point. The only results depending on the weak conjecture from Sec. IV are $\pi_1 M = T^*$, I^*, $T'_{8\cdot 3^k}$, and those with a direct product of $\mathbf{Z}_p$. It should be also noted that $T'_{8\cdot 3} = T^*$, so the conjecture is really only open in two cases.

VI. $\pi_0 \mathrm{Diff}_F\,(S^3/G)$

We now look at the symmetry groups of the asymptotically flat hypersurfaces of the form $S = S^3/G$-point, these are the groups $\pi_0 \mathrm{Diff}_F(S^3/G)$. They will be calculated using the results of Secs. III and V combined with an exact sequence. The reason for using an exact sequence to calculate $\pi_0 \mathrm{Diff}_F(S^3/G)$ is that none of the elements of $\pi_0 \mathrm{Diff}_F(S^3/G)$ can be realized by isometries for G noncyclic.

Since every element of $\mathrm{Diff}_F(M)$ fixes the tangent space at x_0, it follows that $\mathrm{Diff}_F(M) \leqslant \mathrm{Diff}^+(M)$. Let $\hat{p}$: $\mathrm{Diff}^+(M) \to F^+(M)$ be the projection defined by $\hat{p}(f) = (f(x_0), fF)$, where ρ: $F^+(M) \to M$ is the bundle of oriented frames over M with structure group $\mathrm{SO}(3)$, and fF is the action of the diffeomorphism f on the frame F at x_0. The fiber of the projection $\hat{p}$ is $\mathrm{Diff}_F(M)$. We can show this is a fiber bundle. We have the following exact sequence of groups because $\hat{p}$: $\mathrm{Diff}^+(M) \to F^+(M)$ is a principal bundle:

TABLE III. The groups of path components of point fixing diffeomorphisms of spherical spaces, $M = S^3/G$.

$\pi_1 M$	$\pi_0 \mathrm{Diff}^+_{x_0}(M)$
D^*_{4n}, $n>3$	D_{4n}
D^*_8	O
T^*	O
O^*	O
I^*	I
$D^*_{4n} \times \mathbf{Z}_p$, $n>3$	$D_{4n} \times \mathbf{Z}_2$
$D^*_8 \times \mathbf{Z}_p$	$\mathrm{O} \times \mathbf{Z}_2$
$T^* \times \mathbf{Z}_p$	$\mathrm{O} \times \mathbf{Z}_2$
$O^* \times \mathbf{Z}_p$	$\mathrm{O} \times \mathbf{Z}_2$
$I^* \times \mathbf{Z}_p$	$I \times \mathbf{Z}_2$
$D'_{2^k(2n+1)}$	$D_{4(2n+1)} \times \mathbf{Z}_2$
$T'_{8\cdot 3^k}$	O
$D'_{2^k(2n+1)} \times \mathbf{Z}_p$	$D_{4(2n+1)} \times \mathbf{Z}_2$
$T'_{8\cdot 3^k} \times \mathbf{Z}_p$	O
$\mathbf{Z}_p$	$\pi_0 \mathrm{Diff}^+(M)$

$$\cdots\to\pi_i\,\mathrm{Diff}_F(M)\xrightarrow{j_{\#_i}}\pi_i\,\mathrm{Diff}^+(M)\xrightarrow{\hat{p}_{\#_i}}\pi_i F^+(M)$$

$$\xrightarrow{\partial_{\#_i}}\pi_{i-1}\,\mathrm{Diff}_F(M)\to\cdots\to\pi_1 F^+(M)\xrightarrow{\partial_{\#_1}}\pi_0\,\mathrm{Diff}_F(M)$$

$$\xrightarrow{j_{\#_0}}\pi_0\,\mathrm{Diff}^+(M)\to 1\,.$$

All three manifolds have a global frame which implies $F^+(M)\approx M\times\mathrm{SO}(3)$. Hence

$$\pi_i F^+(M)=\pi_i M\times\pi_i\,\mathrm{SO}(3).$$

The only part of the sequence we need is

$$\pi_2 M\times\pi_2\,\mathrm{SO}(3)\to\pi_1\,\mathrm{Diff}_F(M)\to\pi_1\,\mathrm{Diff}^+(M)$$
$$\to\pi_1 M\times\pi_1\,\mathrm{SO}(3)\to\pi_0\,\mathrm{Diff}_F(M)$$
$$\to\pi_0\,\mathrm{Diff}^+(M)\to 1\,.$$

Using the fact that $\pi_1(M)=\pi_1(S^3/G)=G$ and $\pi_2(S^3/G)=\pi_2(S^3)=1$, then

$$1\to\pi_1\,\mathrm{Diff}_F(M)\to\pi_1\,\mathrm{Diff}^+(M)\xrightarrow{\hat{p}_{\#_1}}\pi_1 M\times\mathbf{Z}_2$$

$$\xrightarrow{\partial_{\#_1}}\pi_0\,\mathrm{Diff}_F(M)$$

$$\xrightarrow{j_{\#_0}}\pi_0\,\mathrm{Diff}^+(M)\to 1$$

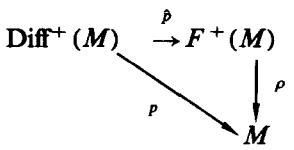
is an exact sequence of groups where $M=S^3/G$.

The exactness of the above sequence implies $\ker j_{\#_0}=\mathrm{Im}\,\partial_{\#_1}$ and $\ker\partial_{\#_1}=\mathrm{Im}\,\hat{p}_{\#_1}$. So $\mathrm{Im}\,\partial_{\#_1}=\pi_1 M\times\mathbf{Z}_2/\ker\partial_{\#_1}$. Hence $\mathrm{Im}\,\partial_{\#_1}=\pi_1 M\times\mathbf{Z}_2/\mathrm{Im}\,\hat{p}_{\#_1}$. Since $\ker j_{\#_0}=\mathrm{Im}\,\partial_{\#_0}$, it follows that the sequence

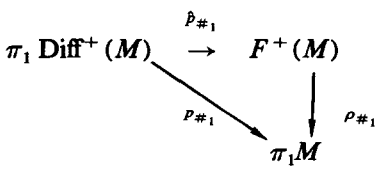
$$1\to\frac{\pi_1 M\times\mathbf{Z}_2}{\mathrm{Im}\,\hat{p}_{\#_1}}\to\pi_0\,\mathrm{Diff}_F(M)\to\pi_0\,\mathrm{Diff}^+(M)\to 1 \tag{6.1}$$

is an exact sequence of groups. If we determine $\mathrm{Im}\,\hat{p}_{\#_1}$, then $\pi_0\,\mathrm{Diff}_F(M)$ can be calculated using sequence (6.1).

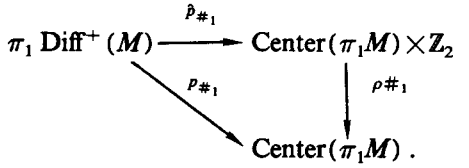
The projection ρ: $F^+(M)\to M$ is just $\rho(m,s)=m$, where $m\in M$ and $s\in\mathrm{SO}(3)$. The projection p: $\mathrm{Diff}^+(M)\to M$ is the composition of ρ and $\hat{p}$, $\rho\hat{p}$: $\mathrm{Diff}^+(M)\to M$. This means the diagram



commutes. Since $p=\rho\hat{p}$, it follows that

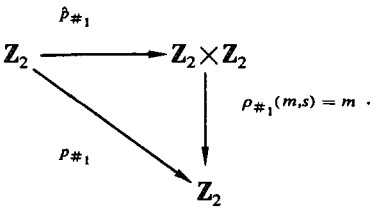


commutes. From Sec. V, we know that, $\mathrm{Im}\,p_{\#_1}=\mathrm{Center}(\pi_1 M)$, $\pi_1 F^+(M)=\pi_1 M\times\mathbf{Z}_2$, and $\rho_{\#_1}$ is just the projection $\rho_{\#_1}(m,s)=m$. Hence, the above diagram becomes



(6.2)

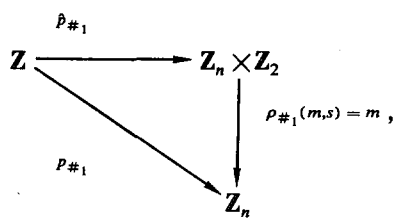
Table II implies $\pi_1\,\mathrm{Diff}^+(M)$ has one generator for $\pi_1 M=\pi_1(S^3/G)=G$, a noncyclic group. In fact the groups $\pi_1\,\mathrm{Diff}^+(M)$ are either $\mathbf{Z}_2$ or $\mathbf{Z}$. If $G\leqslant\mathrm{SU}(2)$ and noncyclic, then diagram (6.2) becomes



From this diagram, it follows that there are only two choices for $\hat{p}_{\#_1}$, namely $\hat{p}_{\#_1}(a)=(a,1)$ and $\hat{p}_{\#_1}(a)=(a,a)$, where $a^2=1$ and $a\neq 1$. For either choice, $\mathrm{Im}\,\hat{p}_{\#_1}=\mathbf{Z}_2$. Hence sequence (6.1) takes the form

$$1\to\pi_1 M\times\mathbf{Z}_2/\mathbf{Z}_2\to\pi_0\,\mathrm{Diff}_F(M)\to\pi_0\,\mathrm{Diff}^+(M)\to 1\,, \tag{6.3}$$

where $\pi_1 M$ is a finite noncyclic subgroup of SU(2). This implies $\pi_0\,\mathrm{Diff}_F(S^3/G)$ has twice the order of $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)$ for G noncyclic. If G is a noncyclic subgroup of SO(4) that is not contained in SU(2), then $\pi_1\,\mathrm{Diff}^+(S^3/G)=\mathbf{Z}$. The center of G is a cyclic group $\mathbf{Z}_n$, $n>2$. Diagram (6.2) becomes



where $\mathbf{Z}_n=\mathrm{Center}(\pi_1 M)$. Again there are only two choices for $\hat{p}_{\#_1}$. One is $\hat{p}_{\#_1}(a)=([a]_n,1)$, and the other is $\hat{p}_{\#_1}(a)=([a]_n,[a]_2)$, where $[\ ]_n$ and $[\ ]_2$ denote the equivalence classes in $\mathbf{Z}_n$ and $\mathbf{Z}_2$, respectively; and a generates $\mathbf{Z}$. One can check that for either choice $\mathrm{Im}\,\hat{p}_{\#_1}=\mathrm{Center}(\pi_1 M)$. Hence sequence (6.1) takes the form

$$1\to\pi_1 M\times\mathbf{Z}_2/\mathrm{Center}(\pi_1 M)$$
$$\to\pi_0\,\mathrm{Diff}_F(M)\to\pi_0\,\mathrm{Diff}^+(M)\to 1\,, \tag{6.4}$$

where $\pi_1 M$ is a finite noncyclic subgroup of SO(4) not contained in SU(2). This implies $\pi_0\,\mathrm{Diff}_F(S^3/G)$ has twice the order of $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)$ for G noncyclic. Using sequences (6.3) and (6.4) the groups $\pi_0\,\mathrm{Diff}_F(S^3/G)$ are now calculated for G noncyclic.

Theorem 6.1:

$$\pi_0 \operatorname{Diff}_F(S^3/G)$$

$$= \begin{cases} O^*, & \text{if } G = D^*_8,\ T^*,\ O^*, \text{ or } T'_{8\cdot 3^k}\times\mathbf{Z}_p,\ p\geqslant 1, \\ D^*_{8n}, & \text{if } G = D^*_{4n}, n\geqslant 3, \\ I^*, & \text{if } G = I^*, \\ O^*\times\mathbf{Z}_2, & \text{if } G = D^*_8\times\mathbf{Z}_p,\ T^*\times\mathbf{Z}_p, \text{ or } O^*\times\mathbf{Z}_p,\ p>1, \\ D^*_{8n}\times\mathbf{Z}_2, & \text{if } G = D^*_{4n}\times\mathbf{Z}_p, p>1 \text{ and } n\geqslant 3, \\ I^*\times\mathbf{Z}_2, & \text{if } G = I^*\times\mathbf{Z}_p, p>1, \\ D^*_{8(2n+1)}\times\mathbf{Z}_2, & \text{if } G = D'_{2^k(2n+1)}\times\mathbf{Z}_p, p\geqslant 1. \end{cases}$$

Proof: If $G\leqslant \mathrm{SU}(2)$, then sequence (6.3) implies

$$1\to G\times\mathbf{Z}_2/\mathbf{Z}_2\to\pi_0 \operatorname{Diff}_F(M)\to\pi_0 \operatorname{Diff}^+(M)\to 1,$$

where $M = S^3/G$. We must decide which choice of $\hat{p}_{\#_1}$ is the correct one. If we think about the definition of $\hat{p}$, it maps f to $(f(x_0), fF)$. The choice $\hat{p}_{\#_1}(a) = (a,1)$ says that $\hat{p}$ does nothing to the elements in $\pi_1 \mathrm{SO}(3)$, and the other, $\hat{p}_{\#_1}(a) = (a,a)$, tells us that $\hat{p}$ acts simultaneously on $\pi_1 M$ and $\pi_1 \mathrm{SO}(3)$. This agrees with $(f(x_0), fF)$; the details will appear in Ref. 31. Hence the correct choice is $\hat{p}_{\#_1}(a) = (a,a)\cdot G\times\mathbf{Z}_2/\mathbf{Z}_2 \approx G$. The exact sequence becomes

$$1\to G\to\pi_0 \operatorname{Diff}_F(S^3/G)\to\pi_0 \operatorname{Diff}^+(S^3/G)\to 1.$$

In the case $G = O^*$ or I^*, $\pi_0 \operatorname{Diff}^+(M) = 1$. Hence the sequence gives $\pi_0 \operatorname{Diff}_F(S^3/G) = G$, for $G = O^*$ or I^*.

If $G = D^*_{4n}$, $n\geqslant 3$, then

$$1\to D^*_{4n}\to\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n})\to\pi_0 \operatorname{Diff}^+(S^3/D^*_{4n})\to 1$$

is exact, and $\pi_0 \operatorname{Diff}^+(S^3/D^*_{4n}) = \mathbf{Z}_2$, for $n\geqslant 3$. Hence $1\to D^*_{4n}\to\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n})\to\mathbf{Z}_2\to 1$ is exact. This means D^*_{4n} is a normal subgroup of $\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n})$. Thus conjugation by elements of $\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n})$ gives automorphisms of D^*_{4n}. Since $\pi_0 \operatorname{Diff}^+(S^3/D^*_{4n}) = \mathbf{Z}_2$, there exist a $\gamma\in \operatorname{Diff}_F(S^3/D^*_{4n})$ for which γ is an outer automorphism of D^*_{4n}. Further, γ and D^*_{4n} will generate $\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n})$. The action of γ on D^*_{4n} is known from the proof of Theorem 3.2. It is $\gamma^2 = x$, $\gamma x\gamma^{-1} = x$, and $\gamma x\gamma^{-1} = xy$. Therefore,

$$\pi_0 \operatorname{Diff}_F(S^3/D^*_{4n}) = \langle \gamma,\ y:\ \gamma^{4n} = 1,\ y^4 = 1,\ y\gamma y^{-1} = \gamma^{-1}\rangle = D^*_{8n},$$

for $n\geqslant 3$.

If $G = T^*$, then $1\to T^*\to\pi_0 \operatorname{Diff}_F(S^3/T^*)\to\pi_0 \operatorname{Diff}^+(S^3/T^*)\to 1$ is exact, and $\pi_0 \operatorname{Diff}^+(S^3/T^*) = \mathbf{Z}_2$. Hence the following sequence is exact: $1\to T^*\to\pi_0 \operatorname{Diff}_F(S^3/T^*)\to\mathbf{Z}_2\to 1$, and T^* is normal subgroup of $\pi_0 \operatorname{Diff}_F(S^3/T^*)$. Again there is a $\gamma\in\pi_0 \operatorname{Diff}_F(S^3/T^*)$ such that γ is an outer automorphism of T^*; γ and T^* generate $\pi_0 \operatorname{Diff}_F(S^3/T^*)$. The action of γ on T^* follows from Theorem 3.2. It follows that $\pi_0 \operatorname{Diff}_F(S^3/T^*) = O^*$.

If $G = D^*_8$, then $1\to D^*_8\to\pi_0 \operatorname{Diff}_F(S^3/D^*_8)\to P_3\to 1$ is exact. Again $D^*_8\trianglelefteq \operatorname{Diff}_F(S^3/D^*_8)$. The action of P_3 on D^*_8 is just conjugation by elements O^*/D^*_8. It follows that $\pi_0 \operatorname{Diff}_F(S^3/D^*_8) = O^*$, because $\pi_0 \operatorname{Diff}_F(S^3/D^*_8)$ is generated by D^*_8 and $\pi_0 \operatorname{Diff}^+(S^3/D^*_8)$.

If G is a noncyclic subgroup of SO(4) not contained in SU(2), then sequence (6.4) is exact:

$$1\to G\times\mathbf{Z}_2/\mathrm{Center}(G)\to\pi_0 \operatorname{Diff}_F(S^3/G)\to\pi_0 \operatorname{Diff}^+(S^3/G)\to 1.$$

Since $p(f) = (f(x_0), fF)$, the choice is $p_{\#_1}(a) = ([a]_n, [a]_2)$. If $G = H\times\mathbf{Z}_p$, where $H\leqslant \mathrm{SU}(2)$ is noncyclic, then the sequence becomes

$$1\to (H\times\mathbf{Z}_p)\times\mathbf{Z}_2/(\mathbf{Z}_p\times\mathbf{Z}_2)\to\pi_0 \operatorname{Diff}_F(S^3/H\times\mathbf{Z}_p)\to\pi_0 \operatorname{Diff}^+(S^3/H)\times\mathbf{Z}_2\to 1,$$

because $\pi_0 \operatorname{Diff}^+(S^3/H\times\mathbf{Z}_p) = \pi_0 \operatorname{Diff}^+(S^3/H)\times\mathbf{Z}_2$.

Therefore $1\to H\to\pi_0 \operatorname{Diff}_F(S^3/H\times\mathbf{Z}_p)\to\pi_0 \operatorname{Diff}(S^3/H)\times\mathbf{Z}_2\to 1$. Theorem 3.2 implies that the $\mathbf{Z}_2$ action is trivial on H. It follows that $\pi_0 \operatorname{Diff}_F(S^3/H\times\mathbf{Z}_p) = \pi_0 \operatorname{Diff}_F(S^3/H)\times\mathbf{Z}_2$ for $H\leqslant \mathrm{SU}(2)$ and noncyclic.

If $G = D'_{2^k(2n+1)}$ or $T'_{8\cdot 3^k}$, then sequence (6.4) becomes

$$1\to D'_{2^k(2n+1)}\times\mathbf{Z}_2/\mathrm{Center}(D'_{2^k(2n+1)})\to\pi_0 \operatorname{Diff}_F(S^3/D'_{2^k(2n+1)})\to\mathbf{Z}_2\times\mathbf{Z}_2\to 1$$

and

$$1\to T'_{8\cdot 3^k}\times\mathbf{Z}_2/\mathrm{Center}(T'_{8\cdot 3^k})\to\pi_0 \operatorname{Diff}_F(S^3/T'_{8\cdot 3^k})\to\mathbf{Z}_2\to 1.$$

It follows that $1\to D^*_{4n}\to\pi_0 \operatorname{Diff}_F(S^3/D'_{2^k(2n+1)})\to\mathbf{Z}_2\times\mathbf{Z}_2\to 1$ and that $1\to T^*\to\pi_0 \operatorname{Diff}_F(S^3/T'_{8\cdot 3^k})\to\mathbf{Z}_2\to 1$ are exact. Checking the action of the $\mathbf{Z}_2$ on T^*, it follows that $\pi_0 \operatorname{Diff}_F(S^3/T'_{8\cdot 3^k}) = O^*$. The action of one of the $\mathbf{Z}_2$ factors in the first sequence is just a trivial action because it comes from $N_{\mathrm{SU}(2)}(S^1)$ and the other comes from the D^*_{8n} factor in the normalizer. It follows that $\pi_0 \operatorname{Diff}_F(S^3/D'_{2^k(2n+1)}) = D^*_{8n(2n+1)}\times\mathbf{Z}_2$.

If $G = D'_{2^k(2n+1)}\times\mathbf{Z}_p$ or $T'_{8\cdot 3^k}\times\mathbf{Z}_p$, then $\pi_0 \operatorname{Diff}_F(S^3/G)$ is the same as the groups for D' and T' alone, because $\pi_0 \operatorname{Diff}(S^3/G)$ is the same with $p = 1$ and $p>1$ (see Table II). Q.E.D.

Let δ: $\pi_0 \operatorname{Diff}_F(M)\to\pi_0 \operatorname{Diff}^+_{x_0}(M)$ be the map $\delta([f]_F) = [f]_{x_0}$, where $[\ ]_F$ and $[\ ]_{x_0}$ denote equivalence classes in $\pi_0 \operatorname{Diff}_F(M)$ and $\pi_0 \operatorname{Diff}^+_{x_0}(\mathrm{M})$, respectively. Let $f'\in[f]_F$. Then $f' = fg$, where g is in the identity component, and g must also be in the identity of $\operatorname{Diff}^+_{x_0}(M)$. Hence $f'\in[f]_{x_0}$. Therefore, $\delta([f']_F) = [f']_{x_0} = [f]_{x_0} = \delta([f]_F)$, i.e., is well defined:

$$\delta([f_1]_F)\delta([f_2]_F) = [f_1]_{x_0}[f_2]_{x_0},$$
$$\delta([f_1f_2]_F) = [f_1f_2]_{x_0} = [f_1]_{x_0}[f_2]_{x_0}.$$

Hence

$$\delta([f_1f_2]_F) = \delta([f_1]_F)\delta([f_2]_F).$$

Therefore δ is a group homomorphism. Using Theorem 6.1 and the results of Table III, observe that

$$1\to\mathbf{Z}_2\to\pi_0 \operatorname{Diff}_F(S^3/G)\xrightarrow{\delta}\pi_0 \operatorname{Diff}^+_{x_0}(M)\to 1,$$

for G a noncyclic group. The $\mathbf{Z}_2$ corresponds to the 2π rotation $R_{2\pi}$ parallel to a two-sphere at x_0. For spaces S^3/G with

G noncyclic, this implies the 2π rotation is nontrivial in $\pi_0\,\mathrm{Diff}_F(S^3/G)$.

Now suppose our manifold is a lens space. Then $\mathrm{Center}(\pi_1 L(p,q)) = \pi_1 L(p,q)$. Sequence (6.2) becomes

$$\begin{array}{ccc} \pi_1\,\mathrm{Diff}^+(L(p,q)) & \xrightarrow{\rho_{\#_1}} & \pi_1 L(p,q)\times\mathbf{Z}_2 \\ & \searrow^{\rho_{\#_1}} \quad \downarrow{\scriptstyle \rho_{\#_1}} & \\ & \pi_1 L(p,q)\,, & \end{array}$$

with $\mathrm{Im}\,\rho_{\#_1} = \pi_1 L(p,q)$. It follows that $\mathrm{Im}\,\rho_{\#_1}$ is either equal to $\pi_1 L(p,q)$ or to $\pi_1 L(p,q)\times\mathbf{Z}_2$. Hence, there are two choices for δ, namely

$$1\to \pi_1 L(p,q)\times\mathbf{Z}_2/\pi_1 L(p,q) \to \mathrm{Diff}_F(L(p,q))$$

$$\xrightarrow{\delta} \pi_0\,\mathrm{Diff}^+_{x_0}(L(p,q))\to 1,$$

and

$$1\to \pi_1 L(p,q)\times\mathbf{Z}_2/\pi_1 L(p,q)\times\mathbf{Z}_2 \to \pi_0\,\mathrm{Diff}_F(L(p,q))$$

$$\xrightarrow{\delta} \pi_0\,\mathrm{Diff}^+_{x_0}(L(p,q))\to 1.$$

The first choice is

$$1\to\mathbf{Z}_2\to\pi_0\,\mathrm{Diff}_F(L(p,q))\xrightarrow{\delta}\mathrm{Diff}^+_{x_0}(L(p,q))\to 1,$$

which corresponds to the rotation $R_{2\pi}$ being nontrivial in $\pi_0\,\mathrm{Diff}_F(L(p,q))$. The 2π rotation is trivial in $\pi_0\,\mathrm{Diff}_F(L(p,q))$ (see Ref. 31 and 32). Therefore, the second choice is the correct one. So $1\to 1\to\pi_0\,\mathrm{Diff}_F(L(p,q))$ $\xrightarrow{\delta}\pi_0\mathrm{Diff}^+_{x_0}(L(p,q))\to 1$. The following theorem is a result of this argument.

Theorem 6.2: $\pi_0\,\mathrm{Diff}_F(L(p,q)) = \pi_0\,\mathrm{Diff}^+(L(p,q))$.

Proof: Because $R_{2\pi}$ is nontrivial in $\pi_0\,\mathrm{Diff}_F(L(p,q))$, $\pi_0\,\mathrm{Diff}_F(L(p,q)) = \pi_0\,\mathrm{Diff}^+_{x_0}(L(p,q))$. Theorem 5.6 implies $\pi_0\,\mathrm{Diff}^+_{x_0}(L(p,q)) = \pi_0\,\mathrm{Diff}(L(p,q))$.

Q.E.D.

Table IV summarizes the results of Secs. IV, V, and VI. The results for $\pi_1 M = T^*, I^*, T'_{8\cdot 3^k}$, and the direct product of these with $\mathbf{Z}_p$ depend on the weak conjecture. This table also gives corrections to Table II in Ref. 10 for the spaces with $\pi_1 M = D^*_{4(2n+1)}$, and $\pi_1 M = \mathbf{Z}_p$ with $q^2\equiv -1 \bmod p$, and fills in the cases that were unknown at the time that paper was published.

VII. CONCLUSION

Comparing the groups $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)$ and $\pi_0\,\mathrm{Diff}_F(S^3/G)$ for a noncyclic group G (Table IV), we see that $\pi_0\,\mathrm{Diff}_F(S^3/G)$ double covers $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)$. Moreover, it is an SU(2) double covering of the factor of $\pi_0\,\mathrm{Diff}^+_{x_0}(S^3/G)$ coming from the noncyclic SO(3) crystal group. The SU(2) covering is due to the 2π rotation being nontrivial in $\pi_0\,\mathrm{Diff}_F(S^3/G)$ for G noncyclic. Physically, $R_{2\pi}$ nontrivial allows half-integral angular momentum states to arise as Friedman and Sorkin have pointed out.[4] The SU(2) double covering is what one would expect for half-integral angular momentum.

The remaining mathematical problem is to calculate $\pi_0\,\mathrm{Diff}_F(M)$ for $M = S^3/G_1\#\cdots\#S^3/G_k\#S^2\times S^1\#\cdots\#S^2\times S^1$. These calculations would involve $\pi_0\,\mathrm{Diff}(S^3/G_i)$ and $\pi_0\,\mathrm{Diff}(S^2\times S^1)$, maps coming from the interchange of factors, and slides of the factors along curves on M. The groups $\pi_0\,\mathrm{Diff}_F(M)$ correspond to the symmetry group of quantum states on $S = (S^3/G_1\#\cdots\#S^3/G_k\#S^2\times S^1\#\cdots\#S^2\times S^1)$-point. These groups are of physical interest because all classical asymptotically flat vacuum space-times seem to have spacelike hypersurfaces with this topology.[33]

TABLE IV. The groups of path components of the diffeomorphisms fixing a frame, point, and nothing for spherical spaces $M = S^3/G$.

$\pi_1 M$	$\pi_0\,\mathrm{Diff}_F(M)$	$\pi_0\,\mathrm{Diff}^+_{x_0}(M)$	$\pi_0\,\mathrm{Diff}^+(M)$
D^*_{4n}, $n>3$	D^*_{8n}	D_{4n}	$\mathbf{Z}_2$
D^*_8	O^*	O	O
T^*	O^*	O	$\mathbf{Z}_2$
O^*	O^*	O	1
I^*	I^*	I	1
$D^*_{4n}\times\mathbf{Z}_p$, $n>3$	$D^*_{8n}\times\mathbf{Z}_2$	$D_{4n}\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
$D^*_8\times\mathbf{Z}_p$	$O^*\times\mathbf{Z}_2$	$O\times\mathbf{Z}_2$	$P_3\times\mathbf{Z}_2$
$T^*\times\mathbf{Z}_p$	$O^*\times\mathbf{Z}_2$	$O\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
$O^*\times\mathbf{Z}_p$	$O^*\times\mathbf{Z}_2$	$O\times\mathbf{Z}_2$	$\mathbf{Z}_2$
$I^*\times\mathbf{Z}_p$	$I^*\times\mathbf{Z}_2$	$I\times\mathbf{Z}_2$	$\mathbf{Z}_2$
$D'_{2^k(2n+1)}$	$D^*_{8(2n+1)}\times\mathbf{Z}_2$	$D_{4(2n+1)}\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
$T'_{8\cdot 3^k}$	O^*	O	$\mathbf{Z}_2$
$D'_{2^k(2n+1)}\times\mathbf{Z}_p$	$D^*_{8(2n+1)}\times\mathbf{Z}_2$	$D_{4(2n+1)}\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
$T'_{8\cdot 3^k}\times\mathbf{Z}_p$	O^*	O	$\mathbf{Z}_2$
$\mathbf{Z}_p$, $q^2\equiv 1 \bmod p$ with $q\not\equiv\pm 1 \bmod p$	$\mathbf{Z}_2\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$	$\mathbf{Z}_2\times\mathbf{Z}_2$
$\mathbf{Z}_p$, $p=1$ or 2	1	1	1
$\mathbf{Z}_p$, remaining cases	$\mathbf{Z}_2$	$\mathbf{Z}_2$	$\mathbf{Z}_2$

ACKNOWLEDGMENTS

I am indebted to J. Arnold, J. Friedman, J. Rubinstein, R. Sorkin, and J. Wagner for helpful discussions and especially J. Friedman for his patience and encouragement.

This research was supported by the University of Wisconsin at Milwaukee Graduate School, and by National Science Foundation Grant No. PHY-8303597-03.

[1]R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).
[2]C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. B **63**, 334 (1976).
[3]C. J. Isham and G. Kunstatter, J. Math. Phys. **23**, 1668 (1982).
[4]J. L. Friedman and R. D. Sorkin, Gen. Relativ. Gravit. **14**, 615 (1982).
[5]C. J. Isham, Phys. Lett. **1068**, 188 (1981).
[6]C. J. Isham, in *Quantum Structure of Space and Time*, edited by M. J. Duff and C. J. Isham (Cambridge U. P., New York, 1982).
[7]A manifold is *open* if it has no boundary and no compact components.
[8]In Yang–Mills theory the Gauss constraint $D_b E^b_\alpha = 0$ with D_b the gauge-covariant derivative, replaces $D_b \pi^{ab} = 0$. Then, in the Schrödinger picture, where $E^b_\alpha = (\hbar/i)(\delta/\delta A^\alpha_b)$ integrating the constraint with test isovector fields ξ^α yields $\Psi \circ G^* = \Psi$, where G is a gauge transformation connected to the identity.
[9]A *closed* manifold is a compact manifold without boundary.
[10]J. L. Friedman and D. M. Witt, Phys. Lett. B **120**, 324 (1983).
[11]A. Hatcher, Ann. Math. **117**, 553 (1983).
[12]Two spaces X and Y have some *homotopy type* if and only if there exists f: $X \to Y$ and g: $Y \to X$ such that $fg \simeq 1$ and $gf \simeq 1$.
[13]A. Hatcher, Proc. Int. Congr. Math. Helsinki, **1978**, 463.
[14]N. V. Ivanov, Sov. Math. Dokl. **20**, 129 (1979).
[15]J. H. Rubinstein, Trans. Am. Math. Soc. **251**, 129 (1979).
[16]J. H. Rubinstein and J. S. Birman, "Homeotopy groups of some non-Haken manifolds," preprint, 1982.
[17]C. J. Hodgson and J. H. Rubinstein, "Involutions and diffeomorphisms on lens spaces," to be published.
[18]Let R: $M^3 - \mathring{B}^3_{x_o} \to M^3 - \mathring{B}^3_{x_o}$ be a diffeomorphism with $R|_{\partial(M^3 - \mathring{B}^3_{x_o}) \approx S^2} = \mathrm{id}$, and support in a neighborhood N of $\partial(M^3 - \mathring{B}^3_{x_o})$, where $\mathring{B}^3_{x_o}$ is the interior of a ball enclosing x_0. Then $R|_N$ is a diffeomorphism of $N \approx S^2 \times I$, with $R|_{\partial N} = \mathrm{id}$ and R is a 2π rotation if and only if $R|_N$ is a 2π rotation of $S^2 \times \{0\}$ with respect to $S^2 \times \{1\}$ on $S^2 \times I$.
[19]T. W. Hungerford, *Algebra* (Springer, New York, 1974).
[20]$\mathbf{H}$ is the division of algebra of quaternions. If $q \in \mathbf{H}$, then $q = al + bi + cj + dk$, where $a, b, c, d \in \mathbf{R}$, $q^* = al - bi - cj - dk$, $ij = k = -ji$, $1q = q1 = q$, and $i^2 = j^2 = k^2 = -1$. The correspondence between $SU(2) = \{\left(\begin{smallmatrix} a & b \\ -b^* & a^* \end{smallmatrix}\right) | a, b \in \mathbf{C},\ aa^* + bb^* = 1\}$ and $S^3 = \{q \in \mathbf{H} | qq^* = 1\}$ is

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ and } k = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

[21]Given any two groups G_1 and G_2 their *direct product*, denoted $G_1 \times G_2$, is the set $G_1 \times G_2 = \{(g_1 g_2) | g_1 \in G_1, g_2 \in G_2\}$ with $(g_1, g_2)(g_1', g_2') = (g_1 g_1', g_2 g_2')$.
[22]C. B. Thomas, in Proc. Sympos. Pure Math. Vol. **32**, part I, (Am. Math. Soc., Providence, RI, 1978) pp. 125–130.
[23]A *spherical space form* is just an orbit space S^3/G, where $G \lhd SO(4)$.
[24]N. Steenrod, *The Topology of Fiber Bundles* (Princeton U.P., Princeton, 1974).
[25]S. T. Hu, *Homotopy Theory* (Academic, New York 1957).
[26]P. Orlik, *Seifert Manifolds, Lecture Notes in Mathematics*, Vol. 291 (Springer, Berlin, 1972).
[27]If $H_1 \trianglelefteq G_1$ and $H_2 \trianglelefteq G_2$, then $G_1 \times G_2 / H_1 \times H_2 \approx (G_1/H_1) \times (G_2/H_2)$, see, for example, Ref. 19.
[28]The *degree* of a map f between two closed manifolds f: $M^n \to N^n$ is the unique integer $\deg f$ that satisfies $f_*(Z_M) = (\deg f) Z_N$, where f_*: $H_n(M^n) \to H_n(N^n)$, Z_M is the fundamental class of $H_n(M^n)$, and Z_N is the fundamental class of $H_n(N^n)$. This is equivalent to $\int_M f^* \omega = (\deg f) \int_N \omega$ for deRham cohomology.
[29]A. Dold, *Lectures on Algebraic Topology* (Springer, New York, 1980).
[30]G. S. McCarty, Jr., Trans. Am. Math. Soc. **106**, 293 (1963).
[31]J. L. Friedman and D. M. Witt, Topology **25** (1), (1986).
[32]H. Hendricks, Bull. Soc. Math. France **53**, 81 (1977).
[33]R. Schoen and S. T. Yau, Phys. Rev. Lett. **43**, 1457 (1979).

Necessary versus sufficient conditions for exact solubility of statistical models on lattices

P. Lochak
Centre de Mathématiques de l'Ecole Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 05, France

J. M. Maillard
Laboratóire de Physique Théorique et Hautes Energies de l'Université Paris VI, Tour 16, ler étage, 4, place Jussieu 75230 Paris Cedex 05, France[a)]

(Received 10 June 1985; accepted for publication 30 September 1985)

It is shown that under rather mild conditions the triangle relation represents a necessary condition for the existence of commuting transfer matrices of arbitrary size. The cases of spin models and vertex models are treated separately.

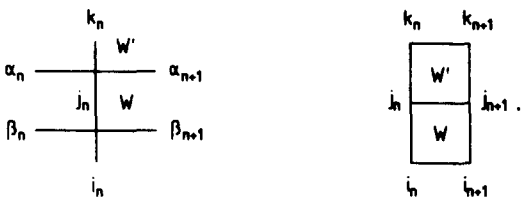
I. INTRODUCTION

The problem of the parametrization of the models is one of the most important in exactly solvable models. All the solutions known in the literature are parametrized by elliptic, trigonometric, or even rational functions, but solutions involving curves of genus bigger than one or even surfaces are still unknown. There is not proof (nor even good arguments) that only genus one curves should occur in the solutions of the Yang–Baxter equations; it is possible to argue (see Appendix A) that one has to deal with algebraic varieties, but it seems very difficult to prove that it is necessary to deal with Abelian varieties. For that reason our approach is a very general one: there are no assumptions like the existence of a unique spectral parameter or the reduction of the Boltzmann weight to a simple transposition for a special value of the parameters. Therefore the proof is completely algebraic. The reader should be told that it is certainly possible to find simpler but less general proofs of the previous equivalence.

II. THE MAIN RESULT

A. Statement of the theorem

Following many authors (see, e.g., Refs. 1–11), it is quite simple to show that the star-triangle relation (for the Boltzman weights W, W', W'') implies the commutation of the transfer matrices with periodic boundary conditions $T_N(W)$ and $T_N(W')$, whatever their size N. The proof leads to a distinction between the case of the vertex models (see Fig. 1) and the case of the spin models (see Fig. 2). The configurations of the spin $i_1 \cdots i_N, k_1 \cdots k_N$ in Figs. 1 and 2 are fixed and we sum all the configurations of the remaining spins $(j_1 \cdots j_N, \alpha_i, \beta_i)$. These two figures represent the product of the two transfer matrices $T_N(W)$ and $T_N(W')$ for vertex and spin models, respectively. In the case of the Potts model (with spins belonging to Z_q), the transfer matrices are thus $q^N \times q^N$ matrices with coefficients $T_N(W)_{i_1,\ldots,i_N;j_1,\ldots,j_N}$ and $T_N(W')_{j_1,\ldots,j_N;k_1,\ldots,k_N}$. The commutation of the transfer matrices means that for any configuration of the spins $i_1 \cdots i_N, k_1 \cdots k_N$, the partition function of the two graphs on both sides of the equality are equal. Let us introduce the two matrices $M_{\alpha_n,\beta_n;\alpha_{n+1},\beta_{n+1}}(i_n,k_n)$ and $M_{j_n,j_{n+1}}(i_n,i_{n+1};k_n,k_{n+1})$ associated with the two following graphs



These two matrices (associated with the vertex and spin models, respectively) are $q^2 \times q^2$ (resp. $q \times q$) matrices and there are q^2 (resp. q^4) of them [as many as the number of configurations for (i_n,k_n) and $(i_n,i_{n+1};k_n,k_{n+1})$]. From now on these matrices will be denoted by M_{I_n} and $M_{I_n,I_{n+1}}$ $[I_n = (i_n,k_n)]$. We add a prime to denote the same matrices with the two Boltzmann weights W and W' permuted. With these notations the commutation of $T_N(W)$ and $T_N(W')$ is equivalent to

$$\mathrm{Tr}\,(M_{I_1}M_{I_2}\cdots M_{I_n}) = \mathrm{Tr}\,(M'_{I_1}M'_{I_2}\cdots M'_{I_N}) \tag{1}$$

and

$$\mathrm{Tr}\,(M_{I_1I_2}M_{I_2I_3}\cdots M_{I_NI_1}) = \mathrm{Tr}\,(M'_{I_1I_2}M'_{I_2I_3}\cdots M'_{I_NI_1}) \tag{2}$$

for any configuration of the I_n's, that is to say for any configurations of the i_n's and k_n's that index the coefficients of the matrices $T_N(W)T_N(W')$ and $T_N(W')T_N(W)$.

We want to establish an equivalence between the existence of a star-triangle relation and the commutation of the transfer matrices $T_N(W)$ and $T_N(W')$ for arbitrary size N; in other words, we want to show that when relation (1) [resp. (2)] is satisfied—for all I_n's and N—there necessarily exists a star-triangle relation. With the above notations it is equivalent to saying that there exists a $q^2 \times q^2$ matrix R (resp. q^2 $q \times q$ matrices R_I) such that

$$RM_{I_n} = M'_{I_n}R \tag{3}$$

$$(\text{resp. } R_IM_{IJ} = M'_{IJ}R_J). \tag{4}$$

B. Proof in the case of the vertex models

The case of the vertex models is the simpler case to deal with. Switching to a slightly more convenient notation $(I_n \to n)$, the question is easily seen to be reduced to the following theorem.

Theorem 1: Let $\mathcal{M}$ and $\mathcal{M}'$ be two subalgebras of $M_n(\mathbb{C})$

[a)] Laboratoire associé au Centre National de la Recherche Scientifique LA 280.

 0022-2488/86/020593-06$02.50 © 1986 American Institute of Physics

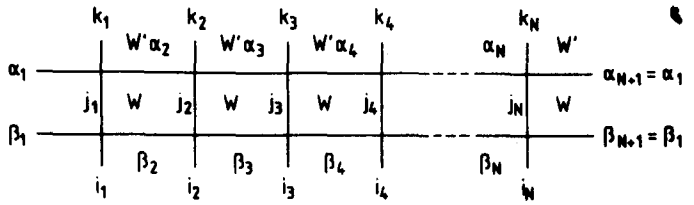


FIG. 1. Pictorial representation of the commutation of the two transfer matrices of size N, $T_N(W)$, and $T_N(W')$ in the case of vertex models.

(the $n\times n$ complex matrices) and φ: $\mathcal{M}\to\mathcal{M}'$, a surjective algebra homomorphism satisfying the following property:

$$\forall\, M_1,\ldots,M_k\in\mathcal{M},$$

$$\mathrm{Tr}\,(M_1\cdots M_k)=\mathrm{Tr}\,[\varphi(M_1\cdots M_k)].$$

Suppose further that there is no nontrivial invariant subspace of $E\simeq\mathbb{C}^n$ under the action of $\mathcal{M}$, then, there exists $R\in\mathrm{GL}_n(\mathbb{C})$ so that

$$\forall M\in\mathcal{M},\quad M'\,(\equiv\varphi(M))=RMR^{-1}.$$

In Appendix A we discuss the problem of the existence of nontrivial invariant subspaces.

Proof: For any $M\in\mathcal{M}$, the corresponding spectral projection operators are elements of $\mathcal{M}$ (being polynomials in M).

We shall need the following lemma.

Lemma: There exists in $\mathcal{M}$ a matrix with (n) distinct eigenvalues.

Proof of the lemma: For any M in $\mathcal{M}$, we set

$$\nu(M)\equiv\Sigma\,(\text{dim. spectral subspace}-1)$$

and $\nu\equiv\inf\nu(M)$, $M\in\mathcal{M}$.

The lemma is then equivalent to $\nu=0$. Suppose $\nu\neq0$ and take $M\in\mathcal{M}$ such that $\nu(M)=\nu$; select further V a spectral subspace of M of dimension $\geqslant2$ and let π be the associated spectral projection operator. One has the following simple proposition.

Proposition: $\forall M_\lambda\in\mathcal{M}$, $\pi M_\lambda\pi$ has only one eigenvalue when considered as an operator on V.

For if not, consider operators of the form

$$\tilde{M}=(1-\pi)M(1-\pi)+\pi(N-k)\pi,$$

where $k\in\mathbb{C}$ and $\pi N\pi$ has more than one eigenvalue on V. Then, for suitable k, $\nu(\tilde{M})<\nu(M)=\nu$, a contradiction.

Any M_λ in $\mathcal{M}$ can thus be written as

$$\pi M_\lambda\pi=k_\lambda\mathbb{1}_\nu+N_\lambda,$$

$k_\lambda\in\mathbb{C}$, N_λ nilpotent on V ($\mathbb{1}_\nu$ is the identity operator on V). Adding $(1-k_\lambda)\mathbb{1}_\nu$, we obtain a family of operators on V of the form $\mathbb{1}_\nu+N_\lambda$, stable under multiplication. The Engel theorem[12] provides us with a vector $v\in V$ such that

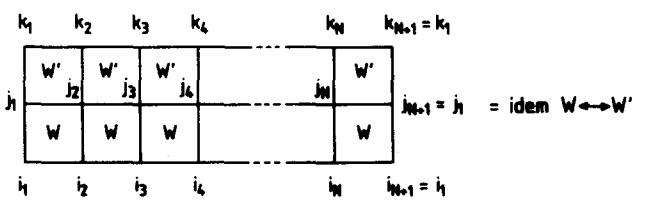


FIG. 2. Pictorial representation of the commutation of two transfer matrices of size N in the case of spin models.

$N_\lambda v=0$ $(\forall\lambda)$. The subspace $(Mv)_{M\in\mathcal{M}}$ is then a nontrivial invariant subspace for $\mathcal{M}$, contradicting the assumption. This finishes the proof of the lemma.

Returning to the main proof, we let $M\in\mathcal{M}$ be a matrix such that $\nu(M)=0$ and let $(\pi_i)_{i=1}^{i=n}$ be the corresponding one-dimensional projection operators

$$\pi_i\cdot\pi_j=\pi_j\cdot\pi_i=\delta_{ij};\quad\sum_{i=1}^{i=n}\pi_i=\mathbb{1}_E.$$

Let $\pi_i'=\varphi(\pi_i)$ and choose $e_i\in\mathrm{Ran}\,\pi_I$, $e_i'\in\mathrm{Ran}\,\pi_i'$. Setting $e_i'\equiv\tilde{R}e_i$ we shall prove that $\tilde{R}$ is the intertwining operator up to scaling, that is,

$$R=D\tilde{R},\quad D=\mathrm{diag}\,(\alpha_1,\ldots,\alpha_n),$$

for some nonzero α_i's.

To prove this, for M_λ in $\mathcal{M}$, we denote by $(m_{ij}^{(\lambda)})$ and $(m_{ij}'^{(\lambda)})$ the matrices of M_λ and $M'_\lambda\equiv\varphi(M_\lambda)$, with respect to the bases $(e_i)_{i=1}^{i=n}$ and $(e_i')_{i=1}^{i=n}$, respectively. The existence of D (which implies the theorem) is then equivalent to the existence of nonzero numbers $(\alpha_i)_{i=1}^{i=n}$ such that

$$m_{ij}'^{(\lambda)}=m_{ij}^{(\lambda)}\alpha_i/\alpha_j.$$

The existence of the α_i's is now proved in a sequence of simple assertions.

Assertion 1:

$$\beta_{ij}\equiv m_{ij}'^{\lambda}/m_{ij}^{(\lambda)}$$

is independent of λ.

For all $M_\lambda,M_\mu\in\mathcal{M}$, $\forall\, i,j$,

$$m_{ij}^{(\lambda)}\cdot m_{ji}^{(\mu)}=\mathrm{Tr}\,(\pi_iM_\lambda\pi_jM_\mu\pi_i),$$

hence

$$m_{ij}^{(\lambda)}\cdot m_{ji}^{(\mu)}=m_{ij}'^{(\lambda)}\cdot m_{ji}'^{(\mu)}$$

or

$$m_{ij}'^{(\lambda)}/m_{ij}^{(\lambda)}=m_{ji}^{(\mu)}/m_{ji}'^{(\mu)},$$

which demonstrates the validity of the assertion.

Assertion 2: $\forall\, i,j$, $\beta_{ij}\neq0,\infty$, i.e., $\forall i,j$, $\exists M_\lambda\in\mathcal{M}$, $m_{ij}^{(\lambda)}\neq0$.

In fact, if there existed a pair (i,j) such that $m_{ij}^{(\lambda)}=0$ for any M_λ in $\mathcal{M}$, then the subspace $(Me_j)_{M\in\mathcal{M}}$ would be a nontrivial (it would not contain e_i) invariant subspace for $\mathcal{M}$.

Assertion 3: There exist n nonzero α_i's such that $\beta_{ij}=\alpha_i/\alpha_j$.

Setting $\alpha_i\equiv\beta_{i1}$, it only remains to show that

$$\forall\, i,j,k,\quad\beta_{ij}\cdot\beta_{jk}=\beta_{ik}\ .$$

This can be written (dropping the superscript λ) as

$$\frac{m_{ij}'}{m_{ij}}\cdot\frac{m_{jk}'}{m_{jk}}=\frac{m_{ik}'}{m_{ik}}=\frac{m_{ki}}{m_{ki}'}$$

(the last equality comes from the proof of Assertion 1) or

$$m'_{ij} \cdot m'_{jk} \cdot m'_{ki} = m_{ij} \cdot m_{jk} \cdot m_{ki}.$$

This finishes the proof of the theorem.

The result has thus been established in the case of vertex models.

C. Proof in the case of the spin models

The algebra is more intricate in this case; this is why, in order to be able to obtain a neat mathematical statement we shall restrict ourselves here to the case $q = 2$. For general q, however, the result must still be valid except for very particular values of the matrices (Boltzmann weight). So let us consider the $q = 2$ case. Here we can replace the cumbersome indexation $I_n = (i_n, k_n)$ ($i_n = \pm 1$, $k_n = \pm 1$) by an index i running through the values 1,2,3,4.

We are thus given a set of sixteen 2×2 matrices $(i,j = 1,2,3,4)$ with positive coefficients, but we can only form "chain products" of the form $M_{i_1 i_2} M_{i_2 i_3} \cdots M_{i_{n-1} i_n}$, returning to the same index i_1. We shall apply Theorem 1 to the algebra generated by multiplying chains starting, and finishing, with the same *fixed* index, but we first need to find a condition that ensures that the hypothesis on the nonexistence of invariant subspaces is satisfied. Since the matrices are 2×2, this is equivalent to the nonexistence of a common eigenspace; we shall also see below that the condition is independent of the length of the chains we consider.

The only possibility we need to explore is the following: Whatever i_1 ($i_1 = 1,2,3,4$), there exists a common eigenvector V_{i_1} for the matrices $M_{i_1 i_2} \cdot M_{i_2 i_3} \cdots M_{i_{n-1} i_n} \cdot M_{i_n i_1}$ (with variable $i_2,\ldots,i_n$).

The M_{ij}'s induce homographic transformations on $\mathbb{P}^1(\mathbb{C})$, which we still call M_{ij} when there is no risk of confusion. The existence of the four vectors V_i is then equivalent to the existence of four points F_i ($i = 1,2,3,4$) for $\mathbb{P}^1(\mathbb{C})$ such that

$$M_{i_1 i_2} \cdots M_{i_n i_1}(F_{i_1}) = F_{i_1}$$

multiplying by $M_{i_n i_1}$ we get

$$M_{i_n i_1} M_{i_1 i_2} \cdots M_{i_{n-1} i_n} \cdot (M_{i_n i_1}(F_{i_1})) = M_{i_n i_1}(F_{i_1}).$$

This shows that we can assume that the F_i's are permuted under the action of M_{ij}'s:

$$\forall\, i,j \in \{1,2,3,4\}, \quad M_{ij}(F_j) = F_i.$$

Also, recalling that the M_{ij}'s have real positive coefficients, we find that each M_{ij} has two real fixed points, one negative and one positive (possibly ∞), and that the real positive axis (including ∞) is stable under their action. This shows that the F_i's are all positive or all negative real numbers. In the latter case, we can replace all the M_{ij}'s by $S^t M_{ij} S$ $[S \equiv \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)]$ and this allows us to assume that the F_i's are all real positive. In Appendix B we describe a pair of families (M_{ij}) and (M'_{ij}) arising in this fashion, which *do not* satisfy the intertwining property to be shown below; they are seen to be essentially the only possible ones.

Let us now state our result in the case of spin models.

Theorem 2: Let (M_{ij}) and (M'_{ij}) be two families of sixteen 2×2 matrices such that the following hold.

(i) All M_{ij}'s and M'_{ij}'s have positive elements and are invertible matrices.

(ii) $\forall n$, $\forall i_1,\ldots,i_n \in \{1,2,3,4\}$,

$$\mathrm{Tr}\,(M_{i_1 i_2} \cdots M_{i_{n-1} i_n} M_{i_n i_1}) = \mathrm{Tr}\,(M'_{i_1 i_2} \cdots M'_{i_{n-1} i_n} M'_{i_n i_1}).$$

(iii) There do not exist four—all positive or all negative—real numbers F_i ($i = 1,2,3,4$), some of the F_i's possibly ∞ so that $M_{ij}(F_j) = M_i$, M_{ij} being viewed as a projective transformation on $\mathbb{P}^1(\mathbb{C})$. Then there exist four matrices R_i ($i = 1,2,3,4$), $R_i \in \mathrm{GL}_2(\mathbb{C})$ with the property

$$\forall\, i,j \in \{1,2,3,4\}, \quad R_i M_{ij} = M'_{ij} R_j.$$

Remarks:

(1) Assumption (iii) can be made on any one of the two families; if it holds for one, it will also be satisfied by the other.

(2) The F_i's can be replaced by vectors $V_i = (1, F_i)$ (or $V_i = 0,1$) if $F_i = \infty$ so that $M_{ij} V_j = \lambda_{ij} V_i$.

(3) The validity of (iii), intricate as it looks, is nonetheless very easy to check. In fact, each F_i is simply one of the two fixed points of M_{ii}; compute these, and check (iii) for the two disjoint sets of the positive fixed points and negative ones.

Proof of the theorem: Since (iii) is satisfied, we can apply Theorem 1 to the algebra generated by the $M_{i_1 i_2} \cdots M_{i_{n-1} i_n} M_{i_n i_1}$ (i_1 fixed), and we choose $n = 2$ (any fixed n is allowed); we also set $i_1 = 1$, without loss of generality. Assumption (iii) means that the $(M_{ij} M_{jk} M_{k1})_{j,k=1,2,3,4}$ have no common eigenspace. Theorem 1 then asserts the existence of R_1 such that

$$\forall\, j,k, \quad R_1 M_{1j} M_{jk} M_{k1} = M'_{ij} M'_{jk} M'_{k1} R_1.$$

Now, *define* R_j by $R_1 M_{1j} = M'_{1j} R_j$, i.e., $R_j \equiv M'^{-1}_{ij} \cdot R_1 \cdot M_{1j}$. We need to check that $R_i M_{ij} = M'_{ij} R_j$, $\forall\, i,j \in \{1,2,3,4\}$. But we can write

$$\begin{aligned} R_i M_{ij} &= M'^{-1}_{1i} R_1 M_{1i} M_{ij} = M'^{1}_{1i} \cdot R_1 M_{1i} M_{ij} M_{j1} \cdot M^{-1}_{j1} \\ &= M'^{-1}_{1i} \cdot M'_{1i} M'_{ij} M'_{j1} R_1 M^{-1}_{j1} \\ &= M'_{ij} M'_{j1} R_1 M^{-1}_{j1}, \end{aligned}$$

and thus we only need to prove that

$$R_j M_{j1} = M'_{j1} R_1, \quad \forall\, j \in \{1,2,3,4\}.$$

By the very definition of R_j, the left-hand side is equal to $M'^{-1}_{ij} R_1 M_{1j} M_{j1}$ and the equality to be shown is therefore equivalent to

$$M'_{1j} R_1 M_{1j} M_{j1} = M'_{j1} R_1 \quad \text{or} \quad R_1 M_{1j} M_{j1} = M'_{1j} M'_{j1} R_1,$$

which in turn can be reduced to

$$R_1 M_{1j} M_{j1} M_{11} = M'_{ij} M'_{j1} R_1 M_{11}.$$

Using the definition of R_1, the left-hand side is equal to $M'_{1j} M'_{j1} M'_{11} R_1$ and we only have to prove that $M'_{11} R_1 = R_1 M_{11}$. But, we already know that $M'^{3}_{11} R_1 = R_1 M^{3}_{11}$. Since both matrices M_{11} and M'_{11} have real positive coefficients, it is easy to show that the desired equality follows, finishing the proof of the theorem.

We thus arrive at (4), which is equivalent to the existence of a star-triangle relation for spin models.

We should note that in both cases (spin and vertex models) the star-triangle relation is implied by the commutation of the transfer matrices for only a *finite* number of sizes N.

This is similar to the result of Parke[13] according to which the existence of only three conserved quantities in involution implies the existence of an infinity of conserved quantities.

III. PROSPECTS

In the previous sections we have tried to specify the equivalence between the commutation of transfer matrices and the star-triangle relations. This amounts to reducing the complete integrability property to a simple local relation. On the other hand, the commutation of transfer matrices of specific sizes leads to the determination of algebraic invariants [cf. Appendix A(a)] that constitute constraining conditions. This explains the results of the search for models satisfying a star-triangle relation, namely, that there exist very few such models. For instance in the case of vertex models with two valued spins the general case is essentially given by the Baxter model and the free fermions models of Fan and Wu.[14] Such an analysis underlines the exceptional occurence of solvable models.

This study also calls for a generalization in dimension 3. In this respect we would like to establish a similar equivalence between the commutation of transfer matrices of finite sizes and the so-called tetrahedron relation[6,15]; this looks like a nontrivial extension. However these commutations of the transfer matrices of finite sizes $([T_{N,M}(W), T_{N,M}(W')] = 0$ are still necessary conditions for the validity of the tetrahedron relation; in particular this includes the conditions that pertain to the two-dimensional models $([T_N(W), T_N(W')] = 0; M = 1)$, and these have been shown to imply the star-triangle relation. This imposes severe restrictions on the possible solutions of the tetrahedron relation that, in a way, appear as extensions of the—sparse—solutions of the star-triangle relation.

The above discussion may give the impression that the domain of validity of the star-triangle and tetrahedron relation is indeed very restricted.

However, if the commutation of transfer matrices allows their simultaneous diagonalization (Bethe ansatz), thereby leading to the calculation of the partition function, we can imagine weaker condition that still make this calculation possible. In fact there already exist simple examples that illustrate this idea; these are the so-called disorder (or crystal-growth) solutions.[16–18] These solutions lead unfortunately to simple analytical expressions for the partition function; however, we should notice that one condition for the existence of such disorder solutions is very similar to a constraintful relation occurring in the framework of exactly solvable models [compare Eq. (2.10) of Ref. 19 and the so-called Frobenius relation[20,21]].

More precisely, if we look carefully at the construction of the Bethe ansatz for the Baxter model,[2] we can see that only relations similar to the so-called Frobenius relations are used [Eqs. (C.34a) and (C.34b) of Ref. 2] and not the full Yang–Baxter structure. We could therefore imagine that a model involving a higher-dimensional theta function would not satisfy the Yang–Baxter equations,[22,23] but that it would actually be possible to build a Bethe ansatz for that model (because of the Frobenius relations) leading to a commutation of transfer matrices only in a *subspace* of the space on which the matrices act; the case of disorder solutions corresponds to a one-dimensional subspace.

IV. CONCLUSION

We have thus shown the equivalence between the existence of a star-triangle relation and that of a family of commuting transfer matrices of arbitrary size; this has been established under conditions mild enough to be almost always satisfied in physical cases. Moreover we have proved that it suffices to check the commutation of the transfer matrices for a finite number of sizes. It may be interesting to look for the three-dimensional generalization of the above results.

In two dimensions, the above equivalence fully legitimizes the tentatively exhaustive studies that are currently done on the star-triangle relation.[24,25] In this framework we have also touched upon the problem of finding simple, algebraic, necessary conditions for the existence of the star-triangle relation (see Appendix A). Such relations, which appear very stringent, are directly related to one of the major problems concerning exactly solvable models: that of the parametrization of these models (rational or elliptic uniformization, Abelian varieties).

Finally these studies on the star-triangle relation seem to show that this is really a rarity; it is thus desirable to extend the notion of integrability beyond it, and to introduce new local criteria.

ACKNOWLEDGMENT

It is a pleasure to thank A. Douady for very helpful conversations, in particular for providing a complete answer to question (a) of Appendix A.

APPENDIX A: ALGEBRAIC VARIETIES AND COMMUTATION OF MATRICES

In this appendix, we briefly describe a solution to three elementary but important questions. The approach is both theoretical and practical, in that it readily provides effective algorithms. However, being as the size of the different matrices involved is a very important feature of the problem for practical purposes, there may be more powerful methods of solution in a given situation.

Let A and B in $M_n(\mathbb{C})$ be two complex $n \times n$ matrices, which we also view as linear operators on $E \simeq \mathbb{C}^n$ with basis $(e_i)_{i=1}^{i=n}$. The three questions are the following.

(a) Can we find a list of invariants that ensure commutation of A and B? By this we mean expressions $(\varphi_k)_{k=1}^{k=m}$, algebraic in the coefficients of A and B, such that

$$\{\varphi_k(A) = \varphi_k(B);\ k = 1,...,m\} \Leftrightarrow \{AB = BA\}.$$

(b) Can we find an easy way to detect a nontrivial invariant subspace under the action of A and B?

(c) This is the same question as (b) in the one-dimensional case, namely, when do A and B have a common (one-dimensional) eigenspace?

(a) We restrict ourselves to the case when A and B are both diagonalizable with distinct eigenvalues [we denote this subset of $M_n(\mathbb{C})$ by $\tilde{M}_n\mathbb{C}$], that is, we discard the codimension one algebraic variety in $M_n(\mathbb{C})$ given by the vanishing of the discriminant of the characteristic polynomial; the invariants will have poles on this surface.

Now, let $M \in \tilde{M}_n(\mathbb{C})$ have eigenvectors $(w_i)_{i=1}^{i=n}$. There is a natural map φ, given by the composition

$$\tilde{M}_n(\mathbb{C}) \to \mathbb{C}^{n\times n} \to (\mathbb{P}^{n-1}(\mathbb{C}))^n/\sigma_n,$$

$$M \to (W_1,\ldots,W_n) \to (\overline{W}_1,\ldots,\overline{W}_n) \bmod \sigma_n,$$

where the bar denotes the natural fibration map $\mathbb{C}^n \to \mathbb{P}^{n-1}(\mathbb{C})$.

Clearly we have the following proposition.

Proposition: $\forall A, B \in \tilde{M}_n(\mathbb{C})$, $AB = BA \Leftrightarrow \varphi(A) = \varphi(B)$.

It remains therefore to give an explicit description of the map φ. To this end we use the embedding

$$(\mathbb{P}^{n-1}(\mathbb{C}))^n/\sigma_n \to \mathbf{P}(S^nE)$$

$$(\overline{W}_1,\ldots,\overline{W}_n) \bmod \sigma_n \to \overline{W_1 \otimes \cdots \otimes W_n},$$

and consider the map $i\circ\varphi$: $M_n(\mathbb{C}) \to \mathbf{P}(S^nE)$. This is easily seen to be described by the following proposition.

Proposition: $(i\circ\varphi)(M)$, $\tilde{M} \in M_n(\mathbb{C})$, represents the one-dimensional eigenspace of $M^{\otimes n}$ (the nth symmetric power of M) for the eigenvalue $\det(M)$.

Proof: If $MW_i = \mu_i W_i$, we have

$$\begin{aligned} M^{\otimes n} W_1 \otimes \cdots \otimes W_n &= MW_1 \otimes \cdots \otimes MW_n \\ &= \mu_1 \cdots \mu_n \cdot W_1 \otimes \cdots \otimes W_n \\ &= \det(M)\, W_1 \otimes \cdots \otimes W_n. \end{aligned}$$

The recipe is thus the following: Compute $M^{\otimes n}$ acting on S^nE [of dimension $\binom{2n-1}{n}$] and find the eigenvector of this matrix for the eigenvalue $\det(M)$, which appears as a polynomial in the variables $(e_i)_{i=1}^{i=n}$, homogeneous of degree n. The quotients of the coefficients of this polynomial by any one of them represent the sought after invariants.

Example: $n = 2$, $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $\det(M) = ad - bc$,

$$M^{\otimes 2} = \begin{bmatrix} a^2 & 2ab & b^2 \\ ac & ad+bc & bd \\ c^2 & 2cd & d^2 \end{bmatrix};$$

the eigenvector of $M^{\otimes 2}$ with eigenvalue $\det(M)$ is given by

$$\Phi = \lambda\,(be_1 \otimes e_1 + (d-a)e_1 \otimes e_2 - ce_2 \otimes e_2).$$

This gives the (projective) invariants $(b, d-a, -c)$ and we may take

$$\varphi_1(M) = b/c; \quad \varphi_2(M) = (d-a)/c.$$

The validity of this result of course can be readily checked by direct computation.

Important remark: The φ's we have found are enormously redundant for $n > 2$. In fact, there should be $n(n-1)$ [$= \dim(\mathbb{P}^{n-1}(\mathbb{C}))^n/\sigma_n$] of them, whereas our result gives $\binom{2n-1}{n} - 1$. It would be interesting to know what is the minimum possible number, a question equivalent to finding "better" embeddings of $(\mathbb{P}^{n-1}(\mathbb{C}))^n/\sigma_n$ in projective varieties. Can the optimum $(n(n-1))$ be achieved?

(b) This question is reduced to the next by the following obvious proposition.

Proposition: There is an equivalence between the following statements: (i) M has an invariant subspace of dimension j generated by $(U_1,\ldots,U_j)$; and (ii) $\Lambda^j M$ [the jth exterior power of M, dimension $\binom{n}{j}$] has $U_1 \wedge U_2 \wedge \cdots \wedge U_j$ as an eigenvector.

(c) We make again the hypothesis that A and B are in $\tilde{M}_n(\mathbb{C})$ and let $U_1,\ldots,U_j$ (resp. $V_1,\ldots,V_j$) be the eigenvectors—unique up to scalar multiplication—of A (resp. B). Then $U_1 \otimes \cdots \otimes U_n$ and $V_1 \otimes \cdots \otimes V_n$ are viewed as two polynomials in $\mathbb{C}[e_1,\ldots,e_n]$. We have the following equivalence: (i) A and B possess a common eigenvector; and (ii) $U_1 \otimes \cdots \otimes U_n$ and $V_1 \otimes \cdots \otimes V_n$ have a linear factor in common.

To check (ii), simply use $\mathbb{C}[e_1,\ldots,e_n] \approx \mathbb{C}[e_1,\ldots,\hat{e}_j,\ldots,e_n][e_j]$ for some j (any j will do) and perform the Euclidean algorithm. (This, of course, relies heavily on the fact that we know *a priori* that the polynomials we are working on can be decomposed into a product of linear factors.)

APPENDIX B: DEGENERATE CASE FOR ISING SPIN MODEL

Here we describe two families of 2×2 matrices

$$M_{ij} = \begin{pmatrix} a_{ij} & b_{ij} \\ c_{ij} & d_{ij} \end{pmatrix}$$

(resp. M'_{ij}) with the following properties:

(i) The M_{ij}'s and M'_{ij}'s have real positive elements.

(ii) The M_{ij}'s (resp. M'_{ij}'s) generate $M_2(\mathbb{C})$ as a vector space.

(iii) $\forall n, \forall i_1,\ldots,i_n$ $(1 \leqslant i_k \leqslant 4)$,

$$\mathrm{Tr}(M_{i_1i_2} \cdots M_{i_{n-1}i_n} M_{i_ni_1}) = \mathrm{Tr}(M'_{i_1i_2} \cdots M'_{i_{n-1}i_n} M'_{i_ni_1}).$$

(iv) There *do not* exist matrices $(R_i)_{i=1}^{i=1}$ such that $R_i M_{ij} = M'_{ij} R_j$. The constructed families will be seen to be essentially the only ones possessing these properties.

We first choose four points $0 < f_1 < F_2 < F_3 < F_4 \in \mathbf{P}^1(\mathbb{C})$ on the positive real axis (possibly with $F_4 = \infty$), corresponding to four vectors V_1, V_2, V_3, V_4 [for example, take $V_i = (1, F_i)$ and $V_4 = (0,1)$ if $F_4 = \infty$] and we also select 16 strictly positive numbers λ_{ij}. The M_{ij}'s and M'_{ij}'s will be constructed in order to satisfy

(a) $M_{ij} \cdot V_j = \lambda_{ij} V_i$; $M'_{ij} \cdot V_j = \lambda_{ij} \cdot V_i$,

(b) $\det M_{ij} = \det M'_{ij}$.

Proposition: (a) and (b) imply condition (iii).

In fact, $M_{i_1i_2} \cdots M_{i_{n-1}i_n} \cdot M_{i_ni_1}$ and $M'_{i_1i_2} \cdots M'_{i_{n-1}i_n} \cdot M_{i_ni_1}$ will have the same determinant and one eigenvalue in common, namely $\lambda_{i_1i_2} \cdots \lambda_{i_{n-1}i_n} \lambda_{i_ni_j}$.

Next, we prove the simple following lemma.

Lemma: $\forall \lambda, \mu; \lambda > 0, \mu > 0$, there exists a one parameter family of 2×2 matrices with real positive elements such that $\det M = \Delta$, Δ some fixed strictly positive number; and $M\binom{\lambda}{1} = \gamma\binom{\mu}{1}$, γ fixed, positive, with $\gamma^2 > \Delta\lambda/\mu$.

Proof: The corresponding homographic transformation looks like

$$M(Z) = \mu + \alpha(Z-\lambda)/(cZ+d), \quad \alpha > 0, \quad c > 0, \quad d > 0.$$

Now

$$M = \begin{pmatrix} \mu c + \alpha & d\mu - \alpha\lambda \\ c & d \end{pmatrix} \det M = \alpha(\lambda c + d),$$

$$\gamma = \lambda c + d.$$

We have therefore $\alpha \equiv \Delta\alpha$ and d *arbitrary* inside $(\Delta\lambda/\mu\gamma, \gamma)$ so that $d\mu - \alpha\lambda > 0$ and $C \equiv (1/\lambda)(\gamma - d) > 0$.

Repeat the above construction for all pairs F_i, F_j, keeping the d_{ij}'s as a set of variables. For the M_{ij}'s and M'_{ij}'s, we

shall take matrices of this form, with different values of the d_{ij}'s. Conditions (i) and (iii) are automatically satisfied. Condition (ii) also is, except for very special values of the d_{ij}'s [and elementary calculations show that these can be chosen so that (iv) also holds]. In fact R_i (resp. R_j) intertwines M_{ii} and M'_{ii} (resp. M_{jj} and M'_{jj}) and we can choose d_{ij} such that $R_i M_{ij}$ and $M'_{ij} R_j$ are different for any R_i and R_j satisfying the intertwining property.

[1]R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).
[2]R. J. Baxter, Ann. Phys. NY **70**, 193 (1972); **76**, 25 (1973).
[3]A. B. Zamolodchikov and A. B. Zamolodchikov, Ann. Phys. NY **120**, 243 (1979).
[4]C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[5]P. Kulish, N. Y. Reshetikhin, and E. K. Sklyanin, Lett. Math. Phys. **5**, 400 (1981).
[6]A. B. Zamolodchikov, Commun. Math. Phys. **79**, 489 (1981).
[7]V. V. Bazhanov and Y. G. Stroganov, Nucl. Phys. B **205**, 505 (1982).
[8]T. Miwa and M. Jimbo, "Classification of solutions of the star-triangle relation for a class of 3 and 4 state IRF models," preprint, 1984.
[9]R. J. Baxter, Ann. Phys. NY **70**, 193 (1972).
[10]P. W. Kasteleyn, *Fundamental Problems in Statistical Mechanics III: Proceedings of the Wageningen Summer School*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1975).
[11]P. P. Kulish and E. K. Sklyanin, "Quantum spectral transform method. Recent developments," in *Lecture Notes in Physics*, Vol. 151 (Springer, Berlin, 1981).
[12]M. Naimark and A. Stern, *Théorie des représentations des groupes* (Mir, Moscow, 1979).
[13]S. Parke, Phys. Lett B **88**, 287 (1979).
[14]C. Fan. and F. Y. Wu, Phys. Rev. B **2**, 723 (1970).
[15]R. J. Baxter, Commun. Math. Phys. **88**, 185 (1982).
[16]T. R. Welberry and R. Galbraith, J. Appl. Crystallogr. **6**, 87 (1973).
[17]I. G. Enting, J. Phys. A **10**, 1023 (1977).
[18]A. M. W. Verhagen, J. Stat. Phys. **15**, 219 (1976).
[19]R. J. Baxter, J. Phys. A **18**, L 911 (1984).
[20]M. Gaudin "La fonction d'onde de Bethe," in *Collection CEA, série scientifique* (Masson, Paris, 1983), pp. 229 and 308.
[21]F. G. Frobenius, J. Reine Angew. Math. **89**, 185 (1885) and *Gesammelte Abhandlungen Band II*, edited by J. P. Serre (Springer, Berlin, 1968).
[22]H. C. Øttinger and J. Honerkamp, Phys. Lett. A **88**, 339 (1982).
[23]R. E. Schrock, Phys. Lett. A **91**, 157 (1982).
[24]I. M. Krichever, Func. Anal. Appl. Math. **15**, 92 (1981).
[25]K. Sogo, M. Uchinami, Y. Akutsu, and M. Wadati, Prog. Theor. Phys. **68**, 508 (1982).

The density of occupied nearest neighbor pairs on a rectangular $2\times N$ lattice space

R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201

J. L. Hock
Department of Electrical Engineering and Computer Science, Marquette University, Milwaukee, Wisconsin 53233

(Received 7 May 1985; accepted for publication 20 September 1985)

Utilizing a 15-term recursion that describes exactly the composite nearest neighbor degeneracy for simple, indistinguishable particles distributed on a $2\times N$ lattice space, $\langle\theta_{11}\rangle$, the expectation of the normalized number of occupied nearest neighbor pairs, is calculated as a function of coverage and the reduced interaction potential.

I. INTRODUCTION

In a recent paper,[1] we derived a 15-term, four-variable recursion relation that yields exactly the composite nearest neighbor degeneracy for simple, indistinguishable particles distributed on a $2\times N$ rectangular lattice space. In the present paper, we utilize the aformentioned recursion to determine the grand canonical partition function and $\langle\theta_{11}\rangle$, the expectation of the normalized number of occupied nearest neighbor pairs as a function of the lattice coverage and of the reduced particle–particle interaction potential.

II. DETERMINATION OF THE GRAND CANONICAL PARTITION FUNCTION

We have shown previously[1] that $A\,[N, q, n_{11}, n_{00}]$, the number of arrangements of q simple, indistinguishable particles on a $2\times N$ rectangular lattice that exhibit n_{11} occupied nearest neighbor pairs *and* n_{00} vacant nearest neighbor pairs (as well as n_{01} mixed nearest neighbor pairs), satisfies the recursion

$$\begin{aligned}
&A\,[N+3,q+3,n_{11}+4,n_{00}+4]\\
&\quad=A\,[N+2,q+3,n_{11}+4,n_{00}+1]\\
&\qquad+A\,[N+2,q+2,n_{11}+4,n_{00}+4]\\
&\qquad+A\,[N+2,q+2,n_{11}+3,n_{00}+3]\\
&\qquad+A\,[N+2,q+1,n_{11}+1,n_{00}+4]\\
&\qquad+A\,[N+1,q+2,n_{11}+4,n_{00}+1]\\
&\qquad-A\,[N+1,q+2,n_{11}+3,n_{00}]\\
&\qquad+A\,[N+1,q+1,n_{11}+3,n_{00}+3]\\
&\qquad-A\,[N+1,q+1,n_{11}+1,n_{00}+1]\\
&\qquad+A\,[N+1,q,n_{11}+1,n_{00}+4]\\
&\qquad-A\,[N+1,q,n_{11},n_{00}+3]\\
&\qquad-A\,[N,q,n_{11}+3,n_{00}+3]\\
&\qquad+3A\,[N,q,n_{11}+2,n_{00}+2]\\
&\qquad-3A\,[N,q,n_{11}+1,n_{00}+1]\\
&\qquad+A\,[N,q,n_{11},n_{00}]\,. \qquad (1)
\end{aligned}$$

The initial conditions for the recursion contained in Eq. (1) are contained in Table I.

Utilizing Eq. (1), we form the polynomials

$$f_{N,q}(x,y)\equiv\sum_{n_{11},n_{00}} A\,[N,q,n_{11},n_{00}]\,x^{n_{11}}y^{n_{00}} \qquad (2)$$

[where

$$x\equiv\exp[-V_{11}/kT]\ \text{and}\ y\equiv\exp[-V_{00}/kT]\,,$$

in which V_{11} and V_{00} are, respectively, the interaction potentials for the occupied nearest neighbor pairs and for the vacant nearest neighbor pairs] and obtain a recursion for the generating (canonical partition) function:

TABLE I. The initial conditions for the recursion given in Eq. (1).

N	q	n_{11}	n_{00}	A
1	0	0	1	1
1	1	0	0	2
1	2	1	0	1
2	0	0	4	1
2	1	0	2	4
2	2	0	0	2
2	2	1	1	4
2	3	2	0	4
2	4	4	0	1
3	0	0	7	1
3	1	0	4	2
3	1	0	5	4
3	2	0	2	4
3	2	0	3	4
3	2	1	2	1
3	2	1	3	4
3	2	1	4	2
3	3	0	0	2
3	3	1	1	4
3	3	1	2	4
3	3	2	1	4
3	3	2	2	6
3	4	2	0	4
3	4	2	1	1
3	4	3	0	4
3	4	3	1	4
3	4	4	1	2
3	5	4	0	2
3	5	5	0	4
3	6	7	0	1

 0022-2488/86/020599-04$02.50 © 1986 American Institute of Physics

$$
\begin{aligned}
f_{N+3,q+3}(x,y) &= y^3 f_{N+2,q+3}(x,y) + [1+xy] f_{N+2,q+2}(x,y) \\
&\quad + x^3 f_{N+2,q+1}(x,y) + y^3[1-xy] f_{N+1,q+2}(x,y) \\
&\quad + xy[1-x^2y^2] f_{N+1,q+1}(x,y) \\
&\quad + x^3[1-xy] f_{N+1,q}(x,y) - xy[1-xy]^3 f_{N,q}(x,y),
\end{aligned} \tag{3}
$$

with the initial conditions

$$
\begin{aligned}
&f_{N,0} = y^3 f_{N-1,0}, \quad N>2, \\
&f_{N,1} = y^3 f_{N-1,1} + 2 f_{N-1,0}, \quad N>3, \\
&f_{N,2} = y^3 f_{N-1,2} + [1+xy] f_{N-1,1}, \\
&\qquad + y^3[1-xy] f_{N-2,1} + xy f_{N-2,0}, \quad N>3, \\
&f_{1,0} = y, \quad f_{2,1} = 4y^2, \\
&f_{2,2} = 2[1+2xy], \quad f_{3,1} = 2x^4[1+2y], \\
&f_{3,3} = 2[1+2xy^2+2xy+3x^2y^2+2x^2y], \\
&f_{3,4} = 4x^3+4x^2+2yx^4+4yx^3+yx^2, \\
&f_{3,5} = 2x^4[1+2x], \quad f_{3,6} = x^7.
\end{aligned} \tag{4}
$$

In Eqs. (3) and (4), we adopt the convention that $f_{N,q}(x,y) = 0$, if $q<0$ or if $q>2N$.

The grand canonical (bivariate generating) function is written as

$$
g_N(x,y,z) = \sum_{q=0}^{2N} f_{N,q}(x,y) z^q, \tag{5}
$$

where

$$
z \equiv m \exp[\mu/kT],
$$

in which μ is the chemical potential, m is the adsorbed particle partition function

$$
m = m(x)m(y)m(z) \exp[-V_0/kT],
$$

and V_0 is the interaction potential between the particle and the surface. It should be mentioned that in the Langmuir model for adsorption,[2] $m(x)$, $m(y)$, and $m(z)$ are single-particle harmonic oscillator partition functions.

We can now obtain $g_N(x,y,z)$ by substituting Eq. (3) into Eq. (5):

$$
\begin{aligned}
g_N(x,y,z) &= [y^3 + z(1+xy) + x^3z^2] g_{N-1}(x,y,z) \\
&\quad + [y^3(1-xy)z + xy(1-x^2y^2)z^2 \\
&\quad + x^3z^3(1-xy)] g_{N-2}(x,y,z) \\
&\quad - [xy(1-xy)^3 z^3] g_{N-3}(x,y,z),
\end{aligned} \tag{6}
$$

where the initial conditions are

$$
g_1(x,y,z) = y + 2z + xz^2, \tag{7a}
$$

$$
g_2(x,y,z) = y^4 + 4y^2z + 2z^2(1+2xy) + 4x^2z^3 + x^4z^4, \tag{7b}
$$

$$
\begin{aligned}
g_3(x,y,z) &= y^7 + 2y^4[1+2y]z \\
&\quad + [4y^3 + 4y^2 + 2xy^4 + 4xy^3 + xy^2]z^2 \\
&\quad + 2[1 + 2xy^2 + 2xy + 3x^2y^2 + 2x^2y]z^3 \\
&\quad + [4x^3 + 4x^2 + 2x^4y + 4x^3y + x^2y]z^4 \\
&\quad + 2x^4[1+2x]z^5 + x^7z^6.
\end{aligned} \tag{7c}
$$

To obtain an explicit relation for $g_N(x,y,z)$, the grand canonical partition function, we first form the polynomials

$$
\begin{aligned}
h(x,y,z,\eta) &\equiv \sum_{N=1}^{\infty} g_N(x,y,z)\eta^N \\
&= \eta\left\{\frac{a_0 + a_1\eta + a_2\eta^2}{b_0 + b_1\eta + b_2\eta^2 + b_3\eta^3}\right\} = \eta\frac{r(\eta)}{s(\eta)},
\end{aligned} \tag{8}
$$

in which

$$
a_0 = y + 2z + xz^2, \tag{9a}
$$

$$
a_1 = z[y^2(4-x) - y(1+2y^2)] + z^2[y(2x-x^3) - xy^3] + z^3[x^2(4-y) - x(1+2x^2)], \tag{9b}
$$

$$
a_3 = z^3[4xy^2 + 4x^2y + 2x^2y^2 - 4x^2y^3 - 4x^3y^2 - 4xy - x^3y - xy^3 + x^4y^2 + x^2y^4 + 2x^3y^3], \tag{9c}
$$

and

$$
b_0 = 1, \tag{10a}
$$

$$
b_1 = -[y^3 + z(1+xy) + z^2x^3], \tag{10b}
$$

$$
b_2 = -[z(1-xy)][y^3 + xyz(1+xy) + z^2x^3], \tag{10c}
$$

$$
b_3 = xyz^3(1-xy)^3, \tag{10d}
$$

and $r(\eta)$ is a quadratic function of η, while $s(\eta)$ is a cubic function of η. From

$$
g_N(x,y,z) = \frac{1}{N!}\left|\frac{\partial^N h}{\partial \eta^N}\right|_{\eta=0}, \tag{11}
$$

and using a partial fraction expansion of $h(x,y,z,\eta)$, we obtain

$$
g_N(x,y,z) = \sum_{j=1}^{3} k_j R_j^N, \tag{12}
$$

and the k_j's are given by

$$
k_j = -r(R_j)/s'(R_j), \tag{13}
$$

and where the R_j's are the reciprocals of the roots of the cubic [see Eq. (8)]

$$
s(\eta) = b_0 + b_1\eta + b_2\eta^2 + b_3\eta^3 = 0. \tag{14}
$$

If η_1 is the smallest root of $s(\eta)$, then, as $N\to\infty$, Eq. (12) becomes

$$
g_N(x,y,z) \simeq k_1 R_1^N, \tag{15}
$$

where $R_1 \equiv \eta_1^{-1}$.

This explicit expression for the grand canonical partition function can now be used to determine the expectation of the lattice coverage and the normalized fraction of the number of occupied nearest neighbor pairs.

III. THE DETERMINATION OF $\langle\theta\rangle$

To determine the expectation of the lattice coverage we define $\langle\theta\rangle$ to be

$$
\langle\theta\rangle_N \equiv \langle q\rangle_N/2N, \tag{16}
$$

where

$$
\langle q\rangle = \left\{\sum_q \sum_{n_{11},n_{00}} q A[N,q,n_{11},n_{00}] x^{n_{11}} y^{n_{00}} z^q\right\} \div \left\{\sum_q \sum_{n_{11},n_{00}} A[N,q,n_{11},n_{00}] x^{n_{11}} y^{n_{00}} z^q\right\}. \tag{17}
$$

Thus

$$\langle\theta\rangle_N = \frac{z}{2N}\frac{\partial}{\partial z}\ln[g_N(x,y,z)]$$

$$= \frac{z}{2N}\frac{\partial}{\partial z}\left[\ln(k_1 R_1^N\right] . \tag{18}$$

For $N\to\infty$, we may write

$$\langle\theta\rangle = \frac{z}{2R_1}\frac{\partial R_1}{\partial z} = -\frac{z}{2\eta_1}\frac{\partial\eta_1}{\partial z} . \tag{19}$$

From Eq. (14), we see that

$$\frac{\partial\eta_1}{\partial z} = \left\{\eta_1\frac{\eta b_1}{\partial z} + \eta_1^2\frac{\partial b_2}{\partial z} + \eta_1^3\frac{\partial b_3}{\partial z}\right\}$$

$$\div\{b_1 + 2b_2\eta_1 + 3b_3\eta_1^2\} . \tag{20}$$

Utilizing Eq. (10), and assuming that $y = 1$ (no vacancy–vacancy interaction), Eq. (19) becomes

$$\langle\theta\rangle = \frac{z}{2}\{[1 + x + 2x^3z] + (1-x)[1 + 2xz + 2x^2z$$

$$+ 3x^3z^2]\eta_1 - 3z^2x(1-x)^3\eta_1^2\}$$

$$\div\{[1 + z(1+x) + z^2x^3] + 2\eta_1[z(1-x)]$$

$$\times[1 + xz(1+x) + z^2x^3] - 3\eta_1^2xz^3(1-x)^3\} , \tag{21}$$

which, for $x = 1$, i.e., no particle–particle interaction, reduces to

$$\langle\theta\rangle = z/(1+z) , \tag{22}$$

the Langmuir isotherm. Figure 1 shows $\langle\theta\rangle$ as a function of $\log_e(u)$, where $u = x^{3/2}z$, for several values of x.

IV. THE DETERMINATION OF $\langle\theta_{11}\rangle$

We wish to determine $\langle\theta_{11}\rangle$, the expectation of the normalized number of occupied nearest neighbor pairs. In this calculation we will assume that $y\equiv 1$, i.e., that there is no interaction potential between vacant pairs on the lattice. Here we define $\langle\theta_{11}\rangle$ to be

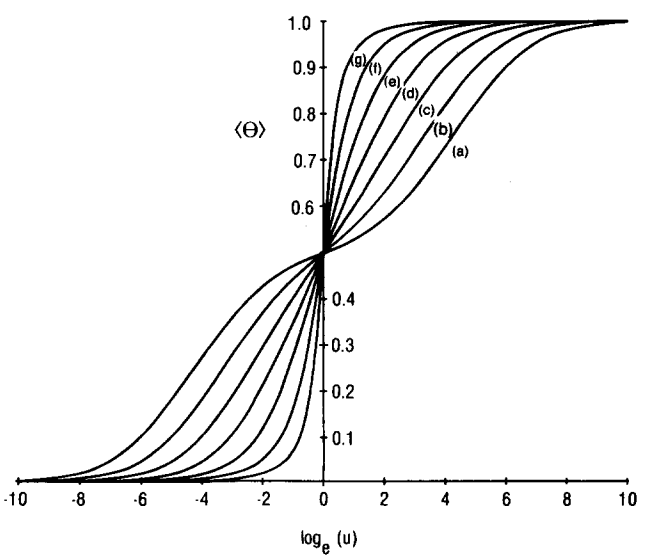


FIG. 1. Shows $\langle\theta\rangle$ as a function of $\log_e(u)$ for various values of $x\equiv V_{11}/kT$. (a) $x = 0.0625$; (b) $x = 0.125$; (c) $x = 0.25$; (d) $x = 0.5$; (e) $x = 1.0$; (f) $x = 2.0$; (g) $x = 4.0$.

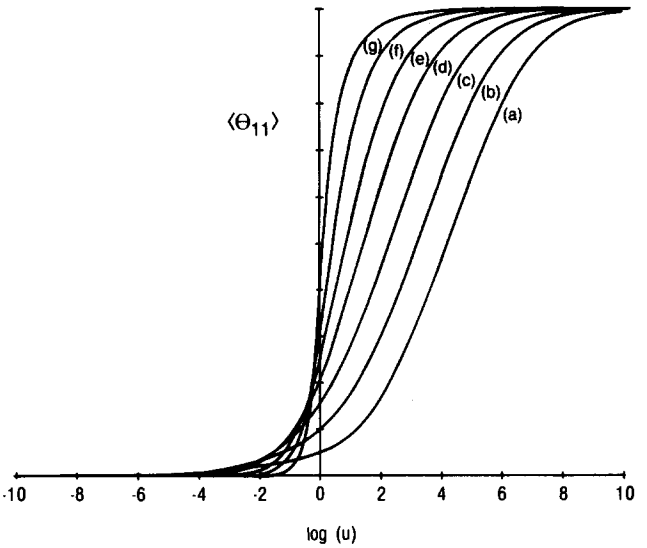


FIG. 2. Shows $\langle\theta_{11}\rangle$ as a function of $\log_e(u)$ for various values of $x\equiv V_{11}/kT$. (a) $x = 0.0625$; (b) $x = 0.125$; (c) $x = 0.25$; (d) $x = 0.50$; (e) $x = 1.0$; (f) $x = 2.0$; (g) $x = 4.0$.

$$\langle\theta_{11}\rangle_N \equiv \langle n_{11}\rangle_N/(3N-2) , \tag{23}$$

where

$$\langle n_{11}\rangle \equiv \left\{\sum_q\sum_{n_{11}} n_{11}A\,[N,q,n_{11},n_{00}]x^{n_{11}}z^q\right\}$$

$$\div\left\{\sum_q\sum_{n_{11}} A\,[N,q,n_{11},n_{00}]x^{n_{11}}z^q\right\} \tag{24}$$

$$= x\frac{\partial}{\partial x}\ln[g_N(x,z)] .$$

Thus, from Eqs. (23) and (24),

$$\langle\theta_{11}\rangle_N = \frac{x}{3N-2}\frac{\partial}{\partial x}\ln[k_1 R_1^N] . \tag{25}$$

As $N\to\infty$, Eq. (25) becomes

$$\langle\theta_{11}\rangle = \frac{x}{3R_1}\frac{\partial R_1}{\partial x} = -\frac{x}{3\eta_1}\frac{\partial\eta_1}{\partial x} . \tag{26}$$

From Eq. (14)

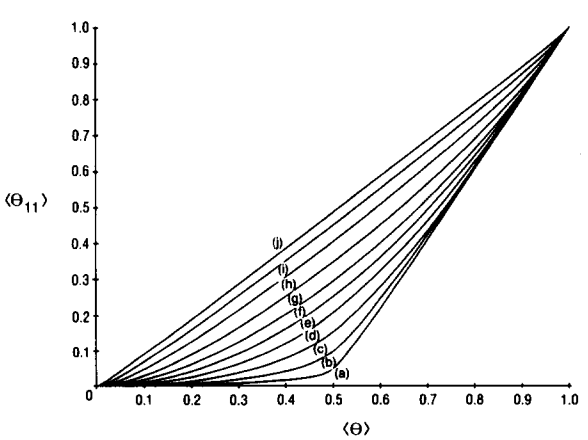


FIG. 3. Shows $\langle\theta_{11}\rangle$ as a function of $\langle\theta\rangle$ for various values of $x\equiv V_{11}/kT$. (a) $x = 0.0625$; (b) $x = 0.125$; (c) $x = 0.25$; (d) $x = 0.50$; (e) $x = 1.0$; (f) $x = 2.0$; (g) $x = 4.0$; (h) $x = 8.0$; (i) $x = 16.0$; (j) $x = 32.0$.

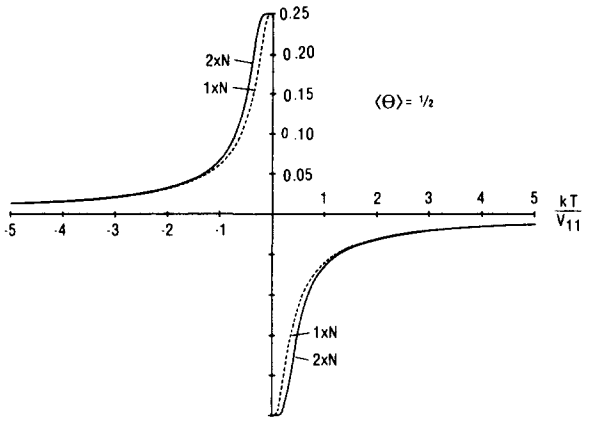


FIG. 4. Shows $\gamma \equiv \langle \theta_{11} \rangle - \langle \theta \rangle^2$ as a function of kT/V_{11}, the reduced temperature for a $2 \times N$ lattice (solid line) and for a $1 \times N$ lattice (dashed line), for $\langle \theta \rangle = \frac{1}{2}$.

$$\frac{\partial \eta_1}{\partial x} = -\left\{\eta_1\left(\frac{\partial b_1}{\partial x}\right) + \eta_1^2\left(\frac{\partial b_2}{\partial x}\right) + \eta_1^3\left(\frac{\partial b_3}{\partial x}\right)\right\} \div \{b_1 + 2b_2\eta_1 + 3b_3\eta_1^2\}, \tag{27}$$

so that Eq. (26) becomes [see Eqs. (10)]

$$\begin{aligned}\langle \theta_{11} \rangle &= \frac{x}{3}\left\{\left(\frac{\partial b_1}{\partial x}\right) + \eta_1\left(\frac{\partial b_2}{\partial x}\right) + \eta_1^2\left(\frac{\partial b_3}{\partial x}\right)\right\} \\ &\quad \div \{b_1 + 2b_2\eta_1 + 3b_3\eta_1^2\} \\ &= \left(\frac{x}{3}\right)\{+z(1+3zx^2) - z[1 + z(3x^2-1) \\ &\quad + z^2x^2(4x-3)\eta_1 + z^3[(4x-1)(x-1)^2]\eta_1^2\} \\ &\quad \div \{[1 + z(1+x) + z^2x^3] + 2z(1-x)[1 + xz(1+x) \\ &\quad + z^2x^3]\eta_1 - 3xz^3(1-x)^3\eta_1^2\}. \end{aligned} \tag{28}$$

For $x = 1$, i.e., for no particle–particle interaction, Eq. (28) reduces to the expected result:

$$\langle \theta_{11} \rangle_{x=1} = [z/(1+z)]^2 = \langle \theta \rangle_{x=1}^2 . \tag{29}$$

Figure 2 shows $\langle \theta_{11} \rangle$ plotted as a function of $\log_e(u)$ for several values of x.

In Fig. 3, we show $\langle \theta_{11} \rangle$ as a function of $\langle \theta \rangle$ for several values of x. Curve (e) ($x = 1.0$) is the parabola $\langle \theta_{11} \rangle = \langle \theta \rangle^2$. We see that for large repulsive potentials $\langle \theta_{11} \rangle$ is small until the coverage is greater than $\frac{1}{2}$, beyond which occupied nearest neighbor pairs are formed in spite of the repulsive interaction. For large attractive potentials, occupied nearest neighbor pairs are easily formed even when the coverage is low.

V. COMPARISON WITH A $1 \times N$ LATTICE SPACE

We define γ to be the deviation of $\langle \theta_{11} \rangle$ from its random value, i.e.,

$$\gamma \equiv \langle \theta_{11} \rangle - \langle \theta \rangle^2 . \tag{30}$$

Figure 4 shows γ as a function of kT/V_{11}, the reduced temperature, when $\langle \theta \rangle = \frac{1}{2}$. For comparison we show γ for a $2 \times N$ lattice and for a $1 \times N$ lattice for which

$$\gamma = -\tfrac{1}{4}\tanh[V_{11}/4kT] . \tag{31}$$

VI. CONCLUSION

We have developed exact statistics for occupied nearest neighbor pairs when simple, indistinguishable particles are distributed on a rectangular $2 \times N$ lattice. Expressions have also been derived that yield exactly the expectation of the coverage and of the density of occupied nearest neighbor pairs.

Results are compared with analogous results for a one-dimensional lattice.

[1]R. B. McQuistan and J. L. Hock, J. Math. Phys. **25**, 261 (1984).
[2]T. L. Hill, *Statistical Thermodynamics* (Addison–Wesley, Reading, MA, 1960).

On the generalized Gibbs formula and the thermodynamic solution of the Boltzmann equation

M. Chen
Vanier College, 821 Ste. Croix Boulevard, St-Laurent, Quebec, Canada H4L 3X9

(Received 28 November 1984; accepted for publication 19 September 1985)

The purpose of this paper is to examine the relationship between the entropy balance equation, the Gibbs formula, and the Boltzmann equation. Consider a system of a mixture of gases contained in an arbitrary region Ω with volume V, where no chemical reactions take place. Let f_i be the one-particle distribution function of species i. First, suppose there exist some f_i, such that the entropy density ρS, the entropy flux $\mathbf{J}_s$, the entropy production σ, and the Boltzmann H-function, $H(t)$, satisfy, respectively, the entropy balance equation and the Boltzmann H-equation under appropriate boundary conditions on the surface $\partial\Omega$ of Ω. Then f_i can be shown to satisfy the Boltzmann equation. Under the functional hypothesis, where f_i depends on time t and the spatial coordinates $\mathbf{r}$ only in terms of the thermodynamic variables—particle density of species i, ρ_i, hydrodynamic velocity $\mathbf{v}$, energy density E, stress tensor $\overset{*}{\pi}_i$, heat flux $\mathbf{Q}_i'$, and mass flux $\mathbf{J}_i$, and possibly the spatial derivatives of $\{\rho_i, \mathbf{v}, E, \overset{*}{\pi}_i, \mathbf{Q}_i', \mathbf{J}_i\}$, the entropy balance equation together with the semipositive definiteness of the entropy production, $\sigma \geqslant 0$, then provides an alternative method of solving the Boltzmann equation. The thermodynamic variables, in turn, are governed by their corresponding evolution equations with appropriate boundary conditions. Second, to the linear order of the spatial gradients of the temperature T, the hydrodynamic velocity $\mathbf{v}$, and the ratio of the particle number n_i/n, the entropy balance equation then yields a generalized Gibbs formula and a nonlinear solution of f_i in terms of the thermodynamic variables, such that $\sigma \geqslant 0$. The generalized Gibbs formula is an exact one-form of the thermodynamic variables that contains the equilibrium Gibbs formula. Furthermore, if f_i is linearized, it is identical to the expression given by Grad's 13-moment method. Finally, we consider the stability problem of the evolution equations for $\overset{*}{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$.

I. INTRODUCTION

It is well known that the Gibbs formula plays an important role in equilibrium thermodynamics.[1] Whether the same assertion holds true or not for nonequilibrium thermodynamics has been a controversial question for some time. In 1949, Prigogine[2] showed that, to the first order, the Chapman–Enskog solution[3] of the Boltzmann equation[4] was consistent with the Gibbs formula of equilibrium form. Recently, De Groot and Mazur[5] analyzed the entropy balance equation in terms of the Chapman–Enskog solution of the Boltzmann equation. They concluded that, beyond the first order, the statistical expression of the entropy density contained the spatial gradients of the thermodynamic variables, which inevitably led to contradiction with the Gibbs formula. This inconsistency can be attributed to either an inappropriate approach in the series expansion and the constraints imposed on the Chapman–Enskog method or incompatibility of the Gibbs formula with the Boltzmann equation in general. It would be ideal if a satisfactory theory of irreversible thermodynamics can be constructed from kinetic theory in terms of the Boltzmann equation. With this aim in mind, recently Eu[6] proposed a modified moment method similar to Grad's 13-moment method.[7] In order to conform with the second law of thermodynamics, Eu emphasized the importance of the Gibbs formula and the entropy balance equation, thereby closing the gap between kinetic theory and irreversible thermodynamics. Based on these considerations, he was able to obtain some interesting results in nonlinear transport processes and a new formulation in nonlinear irreversible thermodynamics.[8]

The main purpose of this paper is to further examine the relationship between the entropy balance equation, the Gibbs formula, and the Boltzmann equation.

In the following discussions we shall adopt the same notations and definitions as given in Eu's paper[6] except for some minor changes. For convenience, these notations are given in Appendix A.

Consider a system of gases with r components contained in an arbitrary region Ω with volume V, where no chemical reactions take place. Let $f_i(t, \mathbf{u}_i, \mathbf{r})$ be the one-particle distribution function of species i. The Boltzmann equation for the system can be written as

$$\frac{\partial f_i}{\partial t} + \mathbf{u}_i \cdot \nabla f_i = \sum_j C(f_i, f_j), \tag{1}$$

where $\mathbf{u}_i$ is the velocity of molecular species i and $C(f_i, f_j)$ is the Boltzmann collision integral.

By Eq. (1) we can derive the following set of evolution equations for the infinite hierarchy of moments, such as ρ, $\mathbf{v}$, E, $\overset{*}{\pi}_i$, $\mathbf{Q}_i'$, $\mathbf{J}_i$, etc.:

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v}, \quad \rho \frac{dc_i}{dt} = -\nabla \cdot \mathbf{J}_i \quad \left(c_i = \frac{\rho_i}{\rho}\right), \tag{2}$$

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla \cdot \overleftrightarrow{\mathbf{P}}, \tag{3}$$

$$\rho \frac{dE}{dt} = -\nabla \cdot \mathbf{Q} - \overleftrightarrow{\mathbf{P}} : \nabla \mathbf{v}, \tag{4}$$

 0022-2488/86/020603-09$02.50 © 1986 American Institute of Physics

$$\frac{d\overleftrightarrow{\pi}_i}{dt} = \overleftrightarrow{\mathbf{Z}}_i^{(p)} + \sum_j \overleftrightarrow{\Lambda}_{ij}^{(p)}, \tag{5}$$

$$\frac{d\mathbf{Q}_i'}{dt} = \mathbf{Z}_i^{(h)} + \sum_j \Lambda_{ij}^{(h)}, \tag{6}$$

$$\frac{d\mathbf{J}_i}{dt} = \mathbf{Z}_i^{(d)} + \sum_j \Lambda_{ij}^{(d)}, \tag{7}$$

$$\frac{dP_i}{dt} = Z_i^{(b)} + \sum_j \Lambda_{ij}^{(b)}, \tag{8}$$

$\vdots$,

where $\overleftrightarrow{\mathbf{Z}}_i^{(p)}, \mathbf{Z}_i^{(h)}, \mathbf{Z}_i^{(d)}, Z_i^{(b)}, \overleftrightarrow{\Lambda}_{ij}^{(p)}, \Lambda_{ij}^{(h)}, \Lambda_{ij}^{(d)}$, and $\Lambda_{ij}^{(b)}$ are given in Appendix A and $d/dt = \partial/\partial t + \mathbf{v}\cdot\nabla$ is the substantial differentiation.

Equations (2)–(8) are the evolution equations considered by Grad and by Eu. We note that each evolution equation depends on the rest of the set. Henceforth, without a closure relation, the infinite set of evolution equations must be considered. However, it is interesting to note that by the definition of the moments, we can rederive the Boltzmann equation (1) from each evolution equation except for the conserved equations (2)–(4), where the collision contribution vanishes. In fact it will be proved in Appendix B that the infinite hierarchy of evolution equations for the moments of f_i is formally equivalent to the Boltzmann equation. Hence, if f_i can be expressed as a series of moments of f_i, then solving the set of evolution equations of the moments will be equivalent to solving the Boltzmann equation. However, from a practical point of view, it is impossible to study the infinite hierarchy of the moment equations. In order to study the hydrodynamic state of a fluid, Grad truncated the hierarchy by including only the 13 moments $\{\rho, \mathbf{v}, E, \overleftrightarrow{\pi}, \mathbf{Q}\}$. Inspired by Grad's method, recently Eu developed a generalized moment method for nonequilibrium thermodynamics, where the state of the thermodynamic system (of mixtures) is described by $W = \{\rho_i, \mathbf{v}, E, \overleftrightarrow{\pi}_i, Q_i', \mathbf{J}_i\}$. Based on Eqs. (2)–(8), a general closure relation can be written as $f_i = f_i(\mathbf{u}_i, \rho_i, \mathbf{v}, E, \overleftrightarrow{\pi}_i, Q_i', \mathbf{J}_i, \nabla\alpha)$, where $\nabla\alpha$ denotes the spatial gradients of W. If f_i can be determined uniquely in terms of W and $\nabla\alpha$, then Eqs. (2)–(8) form a complete set equivalent to the Boltzmann equation.

In order to determine f_i in terms of W and $\nabla\alpha$ (approximately), we consider the following nonlinear functionals of f_i: ρS, $\mathbf{J}_s$, and σ defined, respectively, by

$$\rho S = -K\sum_i \int_{\Omega_i} d\mathbf{u}_i f_i (\log f_i - 1), \tag{9}$$

$$\mathbf{J}_s = -K\sum_i \int_{\Omega_i} d\mathbf{u}_i (\mathbf{u}_i - \mathbf{v}) f_i (\log f_i - 1), \tag{10}$$

$$\sigma = -K\sum_{ij} \int_{\Omega_i} d\mathbf{u}_i\, C(f_i, f_j)\log f_i, \tag{11}$$

where K is the Boltzmann constant and Ω_i is an arbitrary region of $\mathbf{u}_i$.

By Eq. (1) and the definitions of ρS, $\mathbf{J}_s$, and σ we can easily derive the following entropy balance equation:

$$\rho\frac{dS}{dt} = -\nabla\cdot\mathbf{J}_s + \sigma. \tag{12}$$

Conversely, by Eqs. (9)–(12) we can obtain

$$\sum_i \int_{\Omega_i} d\mathbf{u}_i \left\{\frac{\partial f_i}{\partial t} + \mathbf{u}_i\cdot\nabla f_i - \sum_j C(f_i, f_j)\right\}\log f_i = 0.$$

Since no chemical reactions take place, the f_i are independent for all i. Thus

$$\int_{\Omega_i} d\mathbf{u}_i \left\{\frac{\partial f_i}{\partial t} + \mathbf{u}_i\cdot\nabla f_i - \sum_j C(f_i, f_j)\right\}\log f_i = 0. \tag{12'}$$

We now show that Eq. (12') leads to Eq. (1). Let the Boltzmann H-function, $H(t)$, be given by

$$H(t) = \sum_i \int_\Omega d\mathbf{r} \int_{\Omega_i} d\mathbf{u}_i f_i \log f_i,$$

where Ω is the region of the vessel containing the system with volume V.

Consider an arbitrary Ω. The boundary $\partial\Omega$ of Ω may be moving or stationary. If $\partial\Omega$ is moving, the boundary condition of f_i on $\partial\Omega$ can be set up as described by Cercignani,[9] whereas, if $\partial\Omega$ is stationary, the boundary condition can be set up as described by Darrozès and Guiraud.[10]

Suppose there exist some f_i, such that ρS, $\mathbf{J}_s$, and σ, defined by (9)–(11), satisfy (12) and $H(t)$ satisfies the Boltzmann H-equation

$$\frac{dH}{dt} = -\sum_i \int_{\Omega_i} d\mathbf{u} \int_{\partial\Omega} (\mathbf{u}_i f_i \log f_i)\cdot d\mathbf{A} + \sum_{ij} \int_\Omega d\mathbf{r} \int_{\Omega_i} d\mathbf{u}_i\, C(f_i, f_j)\log f_i,$$

but

$$\frac{\partial f_i}{\partial t} + \mathbf{u}_i\cdot\nabla f_i - \sum_j C(f_i, f_j) = h_i \in L^1, \tag{1'}$$

where $d\mathbf{A}$ is a surface element of $\partial\Omega$. Then

$$\int_\Omega d\mathbf{r} \int_{\Omega_i} d\mathbf{u}_i\, h_i(t, \mathbf{u}_i, \mathbf{r}) = \int_{\Omega_i} d\mathbf{u}_i \int_\Omega d\mathbf{r}\, h_i(t, \mathbf{u}_i, \mathbf{r}) = \int\int h_i = 0.$$

Since Ω and Ω_i are arbitrary with finite measure (volume) and $h_i \in L^1$, thus $h_i = 0$ almost everywhere.

Proposition 1: Consider a dilute system of a mixture of gases contained in an arbitrary Ω with volume V, where no chemical reactions take place. Suppose there exist some f_i, such that, ρS, $\mathbf{J}_s$, and σ satisfy (12) and $H(t)$ satisfies the Boltzmann H-equation with appropriate boundary conditions as described by Cercignani, or by Darrozès and Guiraud. Then f_i satisfies the Boltzmann equation (almost everywhere).

Notice first that in the proof of Proposition 1, no boundary condition of Ω_i is required. Thus, Ω_i is completely arbitrary. However, from a physical point of view, Ω_i should be sufficiently large so that f_i vanishes outside the region Ω_i. In that case, ρS is the entropy density, $\mathbf{J}_s$ the entropy flux, σ the entropy production, and Eq. (12) is usually referred to as the entropy balance equation. Furthermore, the condition $\int\int h_i = 0$ is equivalent to the conservation of mass (of the system) as can be easily verified by (1'). Second, the Boltzmann equation is valid only for dilute systems, whereas the

entropy balance equation is practically valid for all systems. Third, if there are no oblique stresses of the gases exerting on the surface $\partial\Omega$, nor is there net energy flow from the gases into the solid body that constitutes the boundary of the vessel, Cercignani and Darrozès and Guiraud have, respectively, proved the Boltzmann H-theorem based on the Boltzmann H-equation.

Hence, instead of solving the Boltzmann equation directly,[11] we look for some f_i indirectly through Eq. (12), which satisfies the Boltzmann H-equation with appropriate boundary condition. In order to achieve this goal, we assume the functional hypothesis (closure relation) that f_i depends on t and $\mathbf{r}$ only in terms of the thermodynamic variables W and possibly their spatial gradients, where the thermodynamic variables W, in turn, are governed by the evolution equations (2)–(8). The boundary condition of f_i must be reformulated in terms of the boundary conditions of the thermodynamic variables W for the evolution equations (2)–(8). Since this is a very delicate and difficult problem, we shall not discuss it in this paper. We assume that the boundary conditions can be set up so that Eqs. (2)–(8) can be solved. On the other hand, the second law of thermodynamics requires that $\sigma \geqslant 0$. Henceforth, by the functional hypothesis, we look for some f_i, which satisfies Eq. (12) and the condition $\sigma \geqslant 0$, where the thermodynamic variables W are determined by the evolution equations under appropriate boundary conditions. The functions f_i obtained in this manner certainly satisfy the Boltzmann equation. We therefore call them the thermodynamic solutions of the Boltzmann equation. In the next section we show how f_i can be obtained (approximately) in this alternative approach, where the Gibbs formula plays an important role.

II. GIBBS FORMULA AND THE THERMODYNAMIC SOLUTION OF THE BOLTZMANN EQUATION

Let $f_i = f_i(\rho_i, \mathbf{v}, E, \nabla\beta) = f_i^{(0)}(1+\phi_i)$, where $\nabla\beta$ denotes the spatial gradients of ρ_i, $\mathbf{v}$, and E. Equation (12) then yields

$$\rho\left\{\frac{dS}{dt} - \frac{1}{T}\left[\frac{dE}{dt} + p\frac{dv}{dt} - \sum_i \mu_i \frac{dc_i}{dt}\right]\right\}$$
$$= -\left[\mathbf{Q} - \sum_i \mu_i \mathbf{J}_i\right]\cdot\nabla\left(\frac{1}{T}\right) + \frac{1}{T}\ddot{\pi}:\nabla\mathbf{v} + \sum_i \frac{1}{T}\mathbf{J}_i\cdot\nabla(\mu_i)$$
$$- K\sum_i \nabla\cdot\left\{\int d\mathbf{u}_i(\mathbf{u}_i - \mathbf{v}) f_i \log(1+\phi_i)\right\} + \sigma, \quad (13)$$

where the divergence is with respect to the spatial coordinates, $v = \rho^{-1}$ is the specific volume, μ_i is the chemical potential of species i per unit mass, and ρ_i, $\rho\mathbf{v}$, and ρE are defined by

$$\begin{Bmatrix}\rho\\ \rho\mathbf{v}\\ \rho E\end{Bmatrix} = \sum_i \int d\mathbf{u}_i\, m_i f_i \begin{Bmatrix}1\\ \mathbf{u}_i\\ \frac{1}{2}(\mathbf{u}_i - \mathbf{v})^2\end{Bmatrix} = \sum_i \int d\mathbf{u}_i\, m_i f_i^{(0)} \begin{Bmatrix}1\\ \mathbf{u}_i\\ \frac{1}{2}(\mathbf{u}_i - \mathbf{v})^2\end{Bmatrix}. \quad (14)$$

Equation (14) then implies the following constraints on ϕ_i:

$$\int d\mathbf{u}_i f_i^{(0)}\phi_i \begin{Bmatrix}1\\ (\mathbf{u}_i - \mathbf{v})\\ \frac{1}{2}(\mathbf{u}_i - \mathbf{v})^2\end{Bmatrix} = 0. \quad (15)$$

By the definition of the entropy density, we have

$$S = \frac{1}{T}\left\{E + pv - \sum_i \mu_i c_i\right\} - \rho^{-1}K\sum_i \int d\mathbf{u}_i f_i \log(1+\phi_i). \quad (16)$$

Hence, S becomes a first-degree homogeneous function of the extensive variables (ρ_i, v, E), which gives rise to the local equilibrium Gibbs formula

$$T\frac{dS}{dt} = \frac{dE}{dt} + p\frac{dv}{dt} - \sum_i \mu_i \frac{dc_i}{dt}, \quad (17)$$

if and only if

$$\int d\mathbf{u}_i f_i \log(1+\phi_i) = 0. \quad (18)$$

We now define the thermodynamic solution of the Boltzmann equation as the solution of the entropy balance equation (13), which satisfies the Gibbs formula (or the generalized Gibbs formula to be defined later) and the condition $\sigma \geqslant 0$.

With the Gibbs formula (17), it is evident that f_i is a thermodynamic solution of the Boltzmann equation if

$$\sigma = -\left\{\frac{1}{T}\ddot{\pi}:\nabla\mathbf{v} + \frac{1}{T}\mathbf{Q}\cdot\nabla\log T + \sum_i \mathbf{J}_i\cdot\nabla\left(\frac{\mu_i}{T}\right)\right\} + K\sum_i \nabla\cdot\left\{\int d\mathbf{u}_i(\mathbf{u}_i - \mathbf{v}) f_i \log(1+\phi_i)\right\}, \quad (19)$$

and $\sigma \geqslant 0$. Conversely, we can solve Eq. (19) for ϕ_i subject to the constraint (15). If ϕ_i satisfies the condition (18), then f_i certainly satisfies the Gibbs formula (17). Therefore, f_i is a solution of Eq. (13). By examining the conditions $\int d\mathbf{u}_i f_i \times\log(1+\phi_i) = \int d\mathbf{u}_i f_i^{(0)}\phi_i$, we notice that either $\phi_i = 0$, the trivial solution, or $(1+\phi_i)\log(1+\phi_i) \sim \phi_i$. Equation (19) then yields

$$\sigma = -\left\{\frac{1}{T}\ddot{\pi}:\nabla\mathbf{v} + \frac{1}{T}\mathbf{Q}\cdot\nabla\log T + \sum_i \mathbf{J}_i\cdot\nabla\left(\frac{\mu_i}{T}\right)\right\} \geqslant 0, \quad (20)$$

which can be shown to be the first-order solution of the Chapman–Enskog method. Hence, by the entropy balance equation (13), we have recovered the following well-known result.

Proposition 2: Suppose $f_i = f_i(\rho_i, \mathbf{v}, E, \nabla\beta) = f_i^{(0)}(1+\phi_i)$. Then f_i is a thermodynamic solution of the Boltzmann equation with respect to the Gibbs formula (17) if, and only if, either (i) $f_i = f_i^{(0)}$ or (ii) f_i is the first-order solution of the Chapman–Enskog method.

By Proposition 2, f_i is called the linear thermodynamic solution of the Boltzmann equation, because if can only describe linear irreversible thermodynamics.[5] This drawback is attributed to the following possibilities: (i) the constraints on ϕ_i are too restrictive; or (ii) the Gibbs space spanned by the thermodynamic variables (ρ_i, v, E) is inadequate. Re-

cently, in order to study nonlinear irreversible thermodynamics, Eu proposed a modified moment method to overcome the above difficulties. The ensuing discussion is a reformulation of Eu's method.

Let us now enlarge the hydrodynamic variables to $(\rho_i, \mathbf{v}, E, \overleftrightarrow{\pi}_i, \mathbf{Q}_i, \mathbf{J}_i, \nabla\alpha)$. Assume $f_i = f_i(\rho_i, \mathbf{v}_i, E, \overleftrightarrow{\pi}_i, \mathbf{Q}_i', \mathbf{J}_i, \nabla\alpha) = f_i^{(0)}(1+\phi_i)$, where $\rho_i, \rho\mathbf{v}$, and ρE are still defined by Eq. (14). In Eq. (13), ρ_i, $\mathbf{v}$, E, $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, $\mathbf{J}_i$, and $\nabla\alpha$ are now considered as independent variables. According to the definitions of $\overleftrightarrow{\Lambda}_{ij}^{(p)}$, $\Lambda_{ij}^{(h)}$, $\Lambda_{ij}^{(d)}$, $\overleftrightarrow{\psi}_i^{(p)}$, and $\psi_i^{(h)}$ given in the Appendix, we can eliminate the $\log(1+\phi_i)$ term in Eq. (13) by setting

$$\begin{aligned}\log(1+\phi_i) &= -\frac{1}{K}[(\mathbf{u}_i-\mathbf{v})(\mathbf{u}_i-\mathbf{v})]^{(2)}:\overleftrightarrow{\mathbf{X}}_i^{(p)} \\ &\quad -\frac{1}{K}\left[\frac{1}{2}m_i(\mathbf{u}_i-\mathbf{v})^2-\frac{5}{2}KT\right](\mathbf{u}_i-\mathbf{v})\cdot\mathbf{X}_i^{(h)} \\ &\quad -\frac{m_i}{K}(\mathbf{u}_i-\mathbf{v})\cdot\mathbf{X}_i^{(d)}-\epsilon(\mathbf{u}_i-\mathbf{v})^4, \qquad (21)\end{aligned}$$

where $\overleftrightarrow{\mathbf{X}}_i^{(p)}$, $\mathbf{X}_i^{(h)}$, and $\mathbf{X}_i^{(d)}$ are independent of $\mathbf{u}_i$, and the superscripts p, h, and d represent pressure tensor, heat, and diffusion. The last term $-\epsilon(\mathbf{u}_i-\mathbf{v})^4$, with ϵ as an infinitesimal real number, is needed to ensure the normalizability of f_i. That is, if we write

$$\log f_i^{(0)} = -\frac{1}{KT}(H_i-\mu_i) = -\frac{1}{KT}\left\{\frac{1}{2}m_i(\mathbf{u}_i-\mathbf{v})^2-\mu_i\right\}$$

and $(1+\phi_i)=\exp(-(1/KT)H_i^{(1)})$, we can include $\exp[-\epsilon(\mathbf{u}_i-\mathbf{v})^4]$ into $f_i^{(0)}$ and define μ_i by

$$\exp(-\beta\mu_i) = \lim_{\epsilon\to 0} n_i^{-1}\int d\mathbf{u}_i \exp[-\epsilon(\mathbf{u}_i-\mathbf{v})^4] \times\exp(-\beta H_i-\beta H_i^{(1)}),$$

with $\beta = 1/KT$. Henceforth, in the following discussions we shall drop the term $-\epsilon(\mathbf{u}_i-\mathbf{v})^4$.

With Eq. (21), we have

$$\begin{aligned}&-K\int d\mathbf{u}_i(\mathbf{u}_i-\mathbf{v})f_i\log(1+\phi_i) \\ &\quad=\overleftrightarrow{\mathbf{X}}_i^{(p)}:\overleftrightarrow{\psi}_i^{(p)}+\mathbf{X}_i^{(h)}\cdot\psi_i^{(h)}+\mathbf{X}_i^{(d)}\cdot\psi_i^{(d)},\end{aligned}$$

and

$$\begin{aligned}\sigma &= \sum_{ij}\{\overleftrightarrow{\mathbf{X}}_i^{(p)}:\overleftrightarrow{\Lambda}_{ij}^{(p)}+\mathbf{X}_i^{(h)}\cdot\Lambda_{ij}^{(h)}+\mathbf{X}_i^{(d)}\cdot\Lambda_{ij}^{(d)}\} \\ &= \sum_i\left\{\overleftrightarrow{\mathbf{X}}_i^{(p)}:\frac{d\overleftrightarrow{\pi}_i}{dt}+\mathbf{X}_i^{(h)}\cdot\frac{d\mathbf{Q}_i'}{dt}+\mathbf{X}_i^{(d)}\cdot\frac{d\mathbf{J}_i}{dt}-\overleftrightarrow{\mathbf{X}}_i^{(p)}:\mathbf{Z}_i^{(p)}\right. \\ &\quad\left.-\mathbf{X}_i^{(h)}\cdot\mathbf{Z}_i^{(h)}-\mathbf{X}_i^{(d)}\cdot\mathbf{Z}_i^{(d)}\right\}. \qquad (22)\end{aligned}$$

Equation (13) then reduces to

$$\begin{aligned}&\frac{\rho}{T}\left\{T\frac{dS}{dt}-\frac{dE}{dt}-p\frac{dv}{dt}+\sum_i\mu_i\frac{dc_i}{dt}-\frac{T}{\rho}\sum_i\left[\overleftrightarrow{\mathbf{X}}_i^{(p)}:\frac{d\overleftrightarrow{\pi}_i}{dt}\right.\right. \\ &\quad\left.\left.+\mathbf{X}_i^{(h)}\cdot\frac{d\mathbf{Q}_i'}{dt}+\mathbf{X}_i^{(d)}\cdot\frac{d\mathbf{J}_i}{dt}\right]\right\} \\ &=\frac{1}{T}\overleftrightarrow{\pi}:\nabla\mathbf{v}+\frac{1}{T}\mathbf{Q}\cdot\nabla\log T+\sum_i\mathbf{J}_i\cdot\nabla\left(\frac{\mu_i}{T}\right) \\ &\quad-\sum_i\{\mathbf{X}_i^{(p)}:\mathbf{Z}_i^{(p)}+\mathbf{X}_i^{(h)}\cdot\mathbf{Z}_i^{(h)}+\mathbf{X}_i^{(d)}\cdot\mathbf{Z}_i^{(d)}\}-\nabla\cdot\mathbf{D},\end{aligned}$$

where

$$\mathbf{D}=\sum_i\{\overleftrightarrow{\mathbf{X}}_i^{(p)}:\overleftrightarrow{\psi}_i^{(p)}+\mathbf{X}_i^{(h)}:\psi_i^{(h)}+\mathbf{X}_i^{(d)}\cdot\overleftrightarrow{\mathbf{P}}_i\}. \qquad (23)$$

If $\overleftrightarrow{\mathbf{X}}_i^{(p)}$, $\mathbf{X}_i^{(h)}$, and $\mathbf{X}_i^{(d)}$ can be determined in terms of ρ_i, $\mathbf{v}$, E, $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$, then f_i is independent of $\nabla\alpha$. By Eq. (9), S becomes a first degree homogeneous function of the extensive variables ρ_i, ν, E, $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$. We therefore look for such an approximate solution f_i of Eq. (23) that satisfies the condition $\sigma\geqslant 0$. Let

$$\begin{aligned}\overleftrightarrow{\mathbf{X}}_i^{(p)} &= \overleftrightarrow{\mathbf{X}}_i^{(p_0)}(\rho_i,\mathbf{v},E,\overleftrightarrow{\pi}_i,\mathbf{Q}_i',\mathbf{J}_i)+\text{terms involving }\nabla\alpha,\\ \mathbf{X}_i^{(h)} &= \mathbf{X}_i^{(h_0)}(\rho_i,\mathbf{v},E,\overleftrightarrow{\pi}_i,\mathbf{Q}_i',\mathbf{J}_i)+\text{terms involving }\nabla\alpha,\\ \mathbf{X}_i^{(d)} &= \mathbf{X}_i^{(d_0)}(\rho_i,\mathbf{v},E,\overleftrightarrow{\pi}_i,\mathbf{Q}_i',\mathbf{J}_i)+\text{terms involving }\nabla\alpha,\\ \overleftrightarrow{\mathbf{Z}}_i^{(p)} &= \overleftrightarrow{\mathbf{Z}}_i^{(p_0)}+\text{higher-order terms of }\nabla\alpha\\ &= -2p_i[\nabla\mathbf{v}]^{(2)}+\cdots,\\ \mathbf{Z}_i^{(h)} &= \mathbf{Z}_i^{(h_0)}+\cdots=-\frac{5}{2}\frac{Kp_i}{m_i}\nabla T+\cdots,\\ Z_i^{(d)} &= Z_i^{(d_0)}+\cdots=p\left[\nabla\left(\frac{n_i}{n}\right)+\left(\frac{n_i}{n}-\frac{\rho_i}{\rho}\right)\nabla\log p\right]+\cdots\\ &= \nabla(\mu_i/T)+\cdots.\end{aligned}$$

Equation (23) then becomes

$$\begin{aligned}&\frac{\rho}{T}\left\{T\frac{dS}{dt}-\frac{dE}{dt}-p\frac{dv}{dt}+\sum_i\left\{\mu_i\frac{dc_i}{dt}-\hat{X}_i^{(p_0)}:\frac{d\overleftrightarrow{\pi}_i}{dt}\right.\right. \\ &\quad\left.\left.-\mathbf{X}_i^{(h_0)}\cdot\frac{d\mathbf{Q}_i'}{dt}-\hat{X}_i^{(d_0)}\cdot\frac{d\mathbf{J}_i}{dt}\right\}\right\}+\cdots \\ &=\sum_i\left\{\left[2p_i\mathbf{X}_i^{(p_0)}+\frac{1}{T}\overleftrightarrow{\pi}_i\right]:[\nabla\mathbf{v}]^{(2)}+\left[-\frac{5Kp_i}{2m_i}\mathbf{X}_i^{(h_0)}\right.\right. \\ &\quad\left.-\frac{1}{T^2}\mathbf{Q}_i'\right]\cdot\nabla T+\left[p\mathbf{X}_i^{(d_0)}+\frac{nK}{m_in_i}\mathbf{J}_i\right] \\ &\quad\left.\cdot\left[\nabla\left(\frac{n_i}{n}\right)+\left(\frac{n_i}{n}-\frac{\rho_i}{\rho}\right)\nabla\log p\right]\right\} \\ &\quad+\text{higher-order terms of }\nabla\alpha, \qquad (24)\end{aligned}$$

where $\hat{X}_i^{(p_0)}=(T/\rho)\overleftrightarrow{\mathbf{X}}_i^{(p_0)}$, $\hat{X}_i^{(h_0)}=(T/\rho)\mathbf{X}_i^{(h_0)}$, $\hat{X}_i^{(d_0)}=(T/\rho)\mathbf{X}_i^{(d_0)}$. It is evident that the left-hand side (lhs) of (24) is a generalization of the lhs of (13) such that S depends on t and r only in terms of the thermodynamic variables $W=(\rho_i,\mathbf{v},E,\overleftrightarrow{\pi}_i,\mathbf{Q}_i',\mathbf{J}_i)$. Since the spatial gradients of W are independent quantities, to the linear order of $[\nabla\mathbf{v}]^{(2)}$, ∇T, and $\nabla(n_i/n)$, (24) in conjunction with (2)–(8) yields lhs = rhs = 0. Consequently,

$$\begin{aligned}T\frac{dS}{dt} &= \frac{dE}{dt}+p\frac{dv}{dt}-\sum_i\mu_i\frac{dc_i}{dt}+\sum_i\left\{\hat{X}_i^{(p_0)}:\frac{d\overleftrightarrow{\pi}_i}{dt}\right. \\ &\quad\left.+\hat{X}_i^{(h_0)}\cdot\frac{d\mathbf{Q}_i'}{dt}+\hat{X}_i^{(d_0)}\cdot\frac{d\mathbf{J}_i}{dt}\right\}, \qquad (25)\end{aligned}$$

where

$$\begin{aligned}\hat{X}_i^{(p_0)} &= -\frac{1}{2p_i\rho}\overleftrightarrow{\pi}_i=-\frac{1}{2p_i}\hat{\pi}_i,\\ \hat{X}_i^{(h_0)} &= -\frac{2m_i}{5Kp_iT\rho}\mathbf{Q}_i'=-\frac{2m_i}{5Kp_iT}\hat{Q}_i', \qquad (26)\end{aligned}$$

$$\hat{X}_i^{(d_0)} = -\frac{1}{\rho_i \rho}\mathbf{J}_i = -\frac{1}{\rho_i}\hat{\mathbf{J}}_i,$$

and the distribution function f_i can be written as

$$f_i = f_i^{(0)} \exp\left\{1 + \frac{m_i}{2p_iKT}[(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} : \overleftrightarrow{\pi}_i + \frac{2m_i}{5K^2 p_i T^2}\left[\frac{1}{2}m_i(\mathbf{u}_i - \mathbf{v})^2 - \frac{5}{2}KT\right](\mathbf{u}_i - \mathbf{v}) \cdot \mathbf{Q}_i' + \frac{m_i}{\rho_i KT}(\mathbf{u}_i - \mathbf{v}) \cdot \mathbf{J}_i\right\}. \tag{27}$$

Since $\hat{X}_i^{(p_0)}$, $\hat{X}_i^{(h_0)}$, and $\hat{X}_i^{(d_0)}$ are independent of the density ρ, they are intensive variables, and thus can be defined as the conjugate variables of $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$, respectively. By Eq. (27), S becomes a first degree homogeneous function of the extensive variables $(\rho_i, \nu, E, \overleftrightarrow{\pi}_i, \mathbf{Q}_i' \cdot \mathbf{J}_i)$ and gives rise to Eq. (25), which can therefore be defined as the generalized Gibbs formula. Note that by (25), $T\,dS$ can be shown to be an exact one-form.

Second, the entropy production is given by

$$\sigma = \sum_{i,j}\{\overleftrightarrow{\mathbf{X}}_i^{(p_0)} : \overleftrightarrow{\Lambda}_{ij}^{(p_0)} + \mathbf{X}_i^{(h_0)} \cdot \Lambda_{ij}^{(h_0)} + \mathbf{X}_i^{(d_0)} \cdot \Lambda_{ij}^{(d_0)}\}$$

$$= -K\sum_{i,j}\int d\mathbf{u}_i\, C(f_i, f_j)\log(1 + \phi_i),$$

where $\overleftrightarrow{\Lambda}_{ij}^{(p_0)}$, $\Lambda_{ij}^{(h_0)}$, $\Lambda_{ij}^{(d_0)}$, are obtained from $\overleftrightarrow{\Lambda}_{ij}^{(p)}$, $\Lambda_{ij}^{(h)}$, and $\Lambda_{ij}^{(d)}$, respectively, with f_i given by Eq. (27). By the properties of the Boltzmann collision integral, one can easily show that

$$\sigma = -\frac{1}{4KT}\sum_{i,j}\int d\Gamma_{ij} f_i^{(0)} f_j^{(0)}(H_{ij}^* - H_{ij}) \times \left[e^{-H_{ij}^*} - e^{-H_{ij}}\right] \geq 0, \tag{28}$$

where $H_i = \log(1 + \phi_i)$, $H_{ij} = H_i + H_j$, H_{ij}^* is the post-collision value of H_{ij}, and $d\Gamma_{ij}$ is the measure of the collision integral.

Finally, by linearizing Eq. (27) we have

$$f_i = f_i^{(0)}\left\{1 + \frac{m_i}{2p_iKT}[(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} : \overleftrightarrow{\pi}_i + \frac{2m_i}{5K^2 p_i T^2}\left[\frac{1}{2}m_i(\mathbf{u}_i - \mathbf{v})^2 - \frac{5KT}{2}\right](\mathbf{u}_i - \mathbf{v}) \cdot \mathbf{Q}_i' + \frac{m_i}{\rho_i KT}(\mathbf{u}_i - \mathbf{v}) \cdot \mathbf{J}_i\right\}, \tag{29}$$

which can be shown to be identical to the 13-moment method of Grad. We can summarize our results in the following proposition.

Proposition 3: Suppose f_i depends on t and $\mathbf{r}$ in terms of the thermodynamic variables $W = (\rho_i, \mathbf{v}, E, \overleftrightarrow{\pi}_i, \mathbf{Q}_i', \mathbf{J}_i)$, which in turn are governed by the evolution equations (2)–(8) with appropriate boundary conditions. Then, to the linear order of $[\nabla\mathbf{v}]^{(2)}$, ∇T, and $\nabla(n_i/n)$, f_i is a nonlinear thermodynamic solution of the Boltzmann equation with respect to the generalized Gibbs formula (25), if, and only if, f_i is given by Eq. (27).

In view of the definitions given by Eqs. (9)–(11), S, $\mathbf{J}_s$, and σ can be considered as nonlinear integral transforms of f_i. Consequently, the entropy balance equation (22) is also a nonlinear integral transform of the Boltzmann equation (1). In Eq. (24), the entropy balance equation is solved to the linear order of $[\nabla\mathbf{v}]^{(2)}$, ∇T, and $\nabla(n_i/n)$, which in turn gives rise to the nonlinear solution of Eq. (1). This is in contrast to the first-order Chapman–Enskog solution of the Boltzmann equation.

By the closure relation (27), we have a closed set of evolution equations (2)–(8). If Eq. (2)–(8) can be solved, we can then obtain a complete description of the thermodynamic solution of the Boltzmann equation. Recently, in a sequence of papers, Eu[7] has extensively studied the nonlinear transport coefficients based on Eqs. (6)–(8). We shall not repeat these topics in this paper. However, it is essential to know whether or not the distribution function f_i given by Eq. (27) indeed approaches equilibrium as $t \to \infty$. In the next section we discuss the stability and asymptotic behavior of $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$.

III. THE STABILITY AND THE ASYMPTOTIC BEHAVIOR OF THE EVOLUTION EQUATIONS FOR $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, AND $\mathbf{J}_i$

In order to simplify the notation, we introduce the following column vectors H, X, ϕ, Z, Λ with $3r - 1$ components given by

$$H = [H_1^{(1)},\ldots,H_r^{(1)},H_1^{(2)},\ldots,H_r^{(2)},H_1^{(3)},\ldots,H_{r-1}^{(3)}],$$

where

$$H_i^{(1)} = [m_i(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)},$$
$$H_i^{(2)} = [\tfrac{1}{2}m_i(\mathbf{u}_i - \mathbf{v})^2 - 5KT/2](\mathbf{u}_i - \mathbf{v}),$$
$$H_i^{(3)} = m_i(\mathbf{u}_i - \mathbf{v});$$
$$X = [X_1^{(1)},\ldots,X_r^{(1)},X_1^{(2)},\ldots,X_r^{(2)},X_1^{(3)},\ldots,X_{r-1}^{(3)}],$$

where

$$X_i^{(1)} = \overleftrightarrow{\mathbf{X}}_i^{(p_0)}, \quad \mathbf{X}_i^{(2)} \equiv X_i^{(h_0)}, \quad \mathbf{X}_i^{(3)} = \mathbf{X}_i^{(d_0)},$$
$$\phi = [\overleftrightarrow{\pi}_1,\ldots,\overleftrightarrow{\pi}_r,\mathbf{Q}_1',\ldots,\mathbf{Q}_r',\mathbf{J}_1,\ldots,\mathbf{J}_{r-1}];$$
$$Z = [Z_1^{(1)},\ldots,Z_r^{(1)},Z_1^{(2)},\ldots,Z_r^{(2)},Z_1^{(3)},\ldots,Z_{r-1}^{(3)}],$$

where

$$Z_i^{(1)} = \overleftrightarrow{\mathbf{Z}}_i^{(p_0)}, \quad \mathbf{Z}_i^{(2)} = \mathbf{Z}_i^{(h_0)}, \quad \mathbf{Z}_i^{(3)} = \mathbf{Z}_i^{(d_0)};$$
$$\Lambda = [\Lambda_1^{(1)},\ldots,\Lambda_r^{(1)},\Lambda_1^{(2)},\ldots,\Lambda_r^{(2)},\Lambda_1^{(3)},\ldots,\Lambda_{r-1}^{(3)}],$$

where

$$\Lambda_i^{(\alpha)} = \sum_j \Lambda_{ij}^{(\alpha)} = \sum_j \int d\Gamma_{ij} f_i^{(0)} f_j^{(0)} H_i^{(\alpha)} \times \{e^{-y_{ij}} - e^{-x_{ij}}\}, \quad \alpha = 1,2,3,$$
$$X_{ij} = \frac{1}{KT}\sum_{\beta=1}^{3}[X_i^{(\beta)}H_i^{(\beta)} + X_j^{(\beta)}H_j^{(\beta)}];$$

and

$$y_{ij} = X_{ij}^* = \text{post collision value of } X_{ij}.$$

To the linear order of $\nabla\mathbf{v}$, ∇T, and $\nabla(\mu_i/T)$ the evolution equations for $\overleftrightarrow{\pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$ can be rewritten as

$$\frac{d}{dt}\phi_i^{(\alpha)} = \Lambda_i^{(\alpha)}(\rho_j, \mathbf{v}, T, \phi) + Z_i^{(\alpha)}(\rho_j, \mathbf{v}, T, \nabla\mathbf{v}, \nabla T, \nabla(\mu_j/T)), \tag{30}$$

which is a system of first-order quasilinear partial differential equations with the same principal part[12]

$$\frac{d\phi_i^{(\alpha)}}{dt}=\frac{\partial\phi_i^{(\alpha)}}{\partial t}+(\mathbf{v}\cdot\nabla)\phi_i^{(\alpha)}.$$

We now define the stationary solution $\phi_i^{(\alpha)}$ (s.t.) of Eq. (30) as the solution of the homogeneous equation

$$\frac{d\phi_i^{(\alpha)}}{dt}=\frac{\partial\phi_i^{(\alpha)}}{\partial t}+v_x\frac{\partial\phi_i^{(\alpha)}}{\partial X}+v_y\frac{\partial\phi_i^{(\alpha)}}{\partial y}+v_z\frac{\partial\phi_i^{(\alpha)}}{\partial z}=0. \tag{31a}$$

Thus $\phi_i^{(\alpha)}$ (s.t.) is obtained from the following system of algebraic equations:

$$Z_i^{(\alpha)}=-\Lambda_i^{(\alpha)}(\rho_j,\mathbf{v},T,\phi(\text{s.t.})),\quad i=1,2,\ldots,r,\quad \alpha=1,2,3. \tag{31b}$$

By inverting $\phi_i^{(\alpha)}$(s.t.) in terms of $Z_i^{(\alpha)}$, we then obtain the nonlinear consititutive relations. On the other hand, the characteristic equations of (31a) are given by

$$\frac{dx}{V_x}=\frac{dy}{v_y}=\frac{dz}{v_z}=dt,$$

and $\phi_i^{(\alpha)}$(s.t.) = const on the characteristics. From a physical point of view, due to the chaotic molecular collisions, $\phi_i^{(\alpha)}$ changes rapidly in space and time as compared to ρ_i, $\mathbf{v}$, T. Thus $\phi_i^{(\alpha)}$ reaches the stationary state defined by $\phi_i^{(\alpha)}$ (s.t.) within a short period of time. During this period of time, ρ_i, $\mathbf{v}$, T can be considered almost constant. Once the system is in the stationary state, $\phi_i^{(\alpha)}$ (s.t.) then depends on t and $\mathbf{r}$ in terms of ρ_i, $\mathbf{v}$, T, which in turn are governed by the conserved equations (2)–(5).

By Eq. (22) the entropy production for the stationary state becomes the following familiar expression:

$$\sigma=-\sum_{i,\alpha}X_i^{(\alpha)}Z_i^{(\alpha)}=-\frac{1}{T}\ddot{\pi}(\text{s.t.}):\nabla\mathbf{v}-\frac{1}{T}\mathbf{Q}(\text{s.t.})\cdot\nabla\log T-\sum_i\mathbf{J}_i(\text{s.t.})\cdot\nabla\left(\frac{\mu_i}{T}\right). \tag{32}$$

However, $\ddot{\pi}$(s.t.), $\mathbf{Q}$(s.t.), and $\mathbf{J}_i$(s.t.) are related to $\nabla\mathbf{v}$, ∇T, and $\nabla(\mu_i/T)$ nonlinearly, respectively, via the nonlinear constitutive relations (31b). If $\Lambda_i^{(\alpha)}$ is linearized, we then have the following linear constitutive relations:

$$Z_i^{(1)}=\frac{2}{T}\sum_{\gamma=2,3}\int d\Gamma_{ii}f_i^{(0)}f_i^{(0)}H_i^{(1)}[H_i^{(1)}-H_i^{(1)*}]X_i^{(1)}(\text{s.t.})+\sum_{j\neq i}\frac{1}{T}\int d\Gamma_{ij}f_i^{(0)}f_j^{(0)}H_i^{(1)}[H_i^{(1)}-H_j^{(1)*}]X_j^{(1)}(\text{s.t.})$$

$$=g_{ij}^{(1,1)}X_i^{(1)}(\text{s.t.})+\sum_{j\neq i}g_{ij}^{(1,1)}X_j^{(1)}(\text{s.t.}), \tag{33a}$$

$$Z_i^{(\beta)}=\frac{2}{T}\sum_{\gamma=2,3}\int d\Gamma_{ii}f_i^{(0)}f_i^{0)}H_i^{(\beta)}[H_i^{(\gamma)}-H_i^{(\gamma)*}]X_i^{(\gamma)}(\text{s.t})+\frac{1}{T}\sum_{j\neq i}\sum_{\gamma=2,3}\int d\Gamma_{ij}f_i^{(0)}f_j^{(0)}H_i^{(\beta)}[H_j^{(\gamma)}-H_j^{(\gamma)*}]X_j^{(\gamma)}(\text{s.t.})$$

$$=\sum_\gamma g_{ii}^{(\beta,\gamma)}X_i^{(\gamma)}(\text{s.t})+\sum_{j\neq i}\sum_\gamma g_{ij}^{(\beta,\gamma)}X_j^{(\gamma)}(\text{s.t.}),\quad \beta=2,3. \tag{33b}$$

Let the matrix g be defined by the block form

$$g=\begin{bmatrix}g^{(1,1)} & 0 & 0\\ 0 & g^{(2,2)} & g^{(2,3)}\\ 0 & g^{(3,2)} & g^{(3,3)}\end{bmatrix},$$

where the elements of $g^{(1,1)}, g^{(2,2)},\ldots,g^{(3,3)}$ are given by (33a) and (33b). It can be shown easily that $g_{ij}^{(1,1)}=g_{ji}^{(1,1)}$, $g_{ii}^{(\beta,\gamma)}=g_{ii}^{(\gamma,\beta)}$, and $g_{ij}^{(\beta,\gamma)}=g_{ji}^{(\gamma,\beta)}$. These are Onsager's relations.[13] By Eqs. (33a) and (33b), Eq. (32) then reduces to the linear irreversible thermodynamics given by Eq. (20), and the generalized Gibbs formula (25) becomes the equilibrium Gibbs formula (17).

The system of evolution equations (30) is equivalent to the following system of ordinary differential equations[12]:

$$\frac{dx}{v_x}=\frac{dy}{v_y}=\frac{dz}{v_z}=dt, \tag{34a}$$

$$\frac{d\phi_i^{(\alpha)}}{dt}=\Lambda_i^{(\alpha)}(\rho_j,\mathbf{v},T,\phi)+Z_i^{(\alpha)}\left(\rho_j,\mathbf{v},T,\nabla\mathbf{v},\nabla T,\nabla\left(\frac{\mu_j}{T}\right)\right). \tag{34b}$$

Strictly speaking, the whole set of evolution equations (2)–(8) must be considered. However, $\phi_i^{(\alpha)}$ changes rapidly in time toward the stationary state $\phi_i^{(\alpha)}$ (s.t.) as compared to the conserved variables ρ_i $\mathbf{v}$, T. For this reason, we can separate the $\phi_i^{(\alpha)}$ from the conserved variables, and consider only Eq. (34b) in terms of ρ_i $\mathbf{v}$, and T and their spatial gradients. For simplicity, we start with the linearized equation of (34b):

$$\frac{d\phi_i^{(\alpha)}}{dt}=\sum_{j,\beta}r_{ij}^{(\alpha,\beta)}\phi_j^{(\beta)}+Z_i^{(\alpha)}, \tag{35}$$

where

$$r_{ij}^{(\alpha,\beta)}=g_{ij}^{(\alpha,\beta)}a_j^{(\beta)},$$

with

$$a_j^{(1)}=-\frac{1}{2p_j\rho},\quad a_j^{(2)}=-\frac{1}{5Kp_jT\rho},\quad a_j^{(3)}=-\frac{1}{\rho_j\rho}.$$

Suppose ρ_i, $\mathbf{v}$, and T are uniformly bounded and continuously differentiable such that $|\rho_i|$, $|\mathbf{v}|$, and $|T'|$ are integrable over the interval $[t_0,\infty)$, and ρ_i, $\mathbf{v}$, and T approach their thermodynamic equilibrium values as $t\to\infty$. Then $r_{ij}^{(\alpha,\beta)}$ is also uniformly bounded and continuously differentiable such that

$$\int_{t_0}^{\infty}\left|\frac{dr_{ij}}{dt}\right|dt<\infty,$$

and $r_{ij}^{(\alpha,\beta)}\to a_{ij}^{(\alpha,\beta)}=\text{const}$ as $t\to\infty$.

Let $R(t)=A+R_1(t)$, where $(R_1)_{i,j}^{(\alpha,\beta)}=r_{ij}^{(\alpha,\beta)}$, $(A)_{ij}^{(\alpha,\beta)}=a_{ij}^{(\alpha,\beta)}$.

Then the matrix $R_1(t)$ has the property

$\int_t^\infty ||R_1(s)||ds < \infty$, and $\int_{t_0}^\infty ||dR_1/ds||ds < \infty$, where $||R_1(s)|| = \Sigma_i \Sigma_j |(R_1)_{ij}|$ is the norm of $R_1(t)$. Like the matrix g, both A and R_1 have the same block form.

Now the solution of Eq. (35) can be written as

$$\phi_i^{(\alpha)} = \psi_i^{(\alpha)} + \int_{t_0}^t Y(t)Y^{-1}(s)Z_i^{(\alpha)}(s)ds, \tag{36}$$

where $\psi_i^{(\alpha)}$ satisfies the homogeneous equation

$$\frac{d\psi_i^{(\alpha)}}{dt} = \sum_{j,\beta} r_{ij}^{(\alpha,\beta)}\psi_j^{(\beta)} = \sum_{j,\beta} [A + R_1(t)]_{ij}^{(\alpha,\beta)}\psi_j^{(\beta)}, \tag{35'}$$

or, in vector notation $(d/dt)\psi = [A + R_1(t)]\psi$, and $Y(t)$ is the fundamental matrix of (35') with $Y(0) = I$.

Suppose A has distinct characteristic roots λ_i. By the properties of $R_1(t)$, Eq. (35') has a fundamental system of solution[14] ψ_i, such that

$$c_2 \exp\left[\mathrm{Re}(\lambda_i t) + d_2 \int_{t_0}^t ||R_1(s)||ds\right]$$
$$\leqslant |\psi_i| = c_1 \exp\left[\mathrm{Re}(\lambda_i t) + d_1 \int_{t_0}^t ||R_1(s)||ds\right],$$

for all $t \geqslant t_0$, where c_1, c_2, d_1, d_2 are positive constants; in particular, $\log(|\psi_i|)/t \to \mathrm{Re}(\lambda_i)$ as $t \to \infty$. As a matter of fact, it can be shown that $\psi_i \to e^{\lambda_i t}\xi_i$ as $t \to \infty$, where ξ_i is a characteristic vector of A corresponding to the characteristic root λ_i (see Ref. 15). On the other hand, the system (35') is uniformly asymptotically stable[16] if and only if there exist positive constants k and c such that $||Y(t)Y^{-1}(t_1)|| \leqslant k\, e^{-c(t-t_1)}$, $t_0 < t_1 < t$. But this is true if and only if all λ_i have negative real parts. For binary mixtures, the negativity of $\mathrm{Re}(\lambda_i)$ can be confirmed by the Routh–Hurwitz method.[17] In general, it is almost impossible to prove that $\mathrm{Re}(\lambda_i) < 0$. However, if all $\mathrm{Re}(\lambda_i)$ are indeed negative, then $\phi_i^{(\alpha)}$ approaches $\phi_i^{(\alpha)}$ (s.t.) exponentially. This can be seen by setting $\phi_i^{(\alpha)} = \psi_i^{(\alpha)} + \phi_i^{(\alpha)}$(s.t.). Then

$$\frac{d}{dt}\phi_i^{(\alpha)} = \frac{d}{dt}\psi_i^{(\alpha)} = \sum_{j,\beta} r_{ij}^{(\alpha,\beta)}\psi_j^{(\beta)}$$
$$+ \left[\sum_{j,\beta} r_{ij}^{(\alpha,\beta)}\phi_j^{(\beta)}(\text{s.t.}) + Z_i^{(\alpha)}\right]$$
$$= \sum_{j,\beta} r_{ij}^{(\alpha,\beta)}\psi_j^{(\beta)}.$$

Once the system is in the stationary state defined by $\phi_i^{(\alpha)}$(s.t.), it is then identical to the first-order solution of the Chapman–Enskog method.

All characteristic roots of A may not necessarily be distinct. In that case, A can be reduced to the Jordan canonical form. Still, the system $\psi' = A\psi$ has solutions all approaching zero as $t \to \infty$, if, and only if, all $\mathrm{Re}(\lambda_i) < 0$. Alternatively, given any matrix A, there exists another matrix A' with distinct characteristic roots[18] such that $||A - A'|| \leqslant \epsilon$, where ϵ is any positive real number that can be made as small as possible (but not zero). This justifies the assumption that A has distinct characteristic roots.

Let the absolute thermodynamic equilibrium state be defined by the absolute Maxwellian distribution function ($f_i^{(0)}$ is independent of t and $\mathbf{r}$) and the thermodynamic equilibrium Gibbs formula $T\,ds = dE + p\,dv - \Sigma_i \mu_i\, d\,c_i$. Since the stationary state $\phi_i^{(\alpha)}$(s.t.) is identical to the first-order Chapman–Enskog method, where the generalized Gibbs formula becomes the local equilibrium Gibbs formula (17), we can readily employ the thermodynamic stability and hydrodynamic stability theory of Glansdorff and Prigogine[19] to f_i given by Eq. (26). If $\phi_i^{(\alpha)}$ (s.t.) is substituted into (26), then f_i can be shown to be identical to the first-order solution of the Chapman–Enskog method. To summarize, we have the following proposition.

Proposition 4: Suppose all characteristic roots of A have negative real parts. Then the solution of Eq. (35) is asymptotically stable. Moreover, f_i given by (29) is a linear thermodynamic stable solution of the Boltzmann equation.

Next we consider the nonlinear equation in vector form

$$\frac{d\phi}{dt} = \Lambda(\phi) + Z. \tag{34b}$$

Let $f(\phi) = \Lambda(\phi) - R\phi$. Then $f(\phi)$ does not contain the constant nor the first-order terms. Consider the homogeneous equation of (34b)

$$\frac{d\psi}{dt} = R\psi + f(\psi).$$

We notice that $\psi = 0$ is a trivial solution. Furthermore, there exist positive constants C_1 and C_2 such that $|f(\psi) - f(\bar{\psi})| \leqslant C_1|\psi - \bar{\psi}|$, where $C_1 \to 0$ as $C_2 \to 0$, and $|\psi| < C_2$, $|\bar{\psi}| < C_2$. Since $\int_{t_0}^\infty ||R_1(t)||dt < \infty$, given $|\psi(t_0)|$ sufficiently small, it can easily be proved[20] that $\psi = 0$ is an asymptotically stable solution of the homogeneous equation if all characteristic roots of A have negative real parts. As a matter of fact, $|\psi| \to 0$ exponentially. The general solution of Eq. (34b) can be written as

$$\phi = \psi + \int_{t_0}^t Y(t)Y^{-1}(t_1)[f(\phi(t_1)) + Z(t_1)]dt_1.$$

If $Z(t) \to 0$ at least as fast as t^{-2} for large t, then $\phi(t) \to 0$ as $t \to \infty$, and the solution of Eq. (34b) is asymptotically stable. Consequently the solution of Eq. (34) given by Eq. (27) is also asymptotically stable. Moreover, f_i approaches the absolute Maxwellian distribution function as $t \to \infty$.

It would be interesting to generalize the thermodynamic stability and the dynamical stability theory of Glansdorff and Prigogine to the nonlinear thermodynamic solution given by Eq. (27) and the generalized Gibbs formula (25) together with the evolution equations (2)–(8) under appropriate boundary conditions. This will be discussed in a subsequent paper.

ACKNOWLEDGMENTS

The author wishes to thank Professor Eu for many valuable discussions throughout the course of this work and the referee for some valuable comments.

He also wishes to acknowledge the financial support of this work by the National Science and Engineering Research Council of Canada, and the FCAC of the Ministry of Education of the Quebec Government.

APPENDIX A: NOTATION

Vector: $\mathbf{A}_1$.

Tensor: $\vec{\vec{\mathbf{B}}}_1$.

Scalar product of vectors: $\mathbf{A}_1 \cdot \mathbf{A}_2$.

Tensor contraction: $\overleftrightarrow{\mathbf{B}}_1 : \overleftrightarrow{\mathbf{B}}_2$.

Local average of A: $\langle A \rangle = \int d\mathbf{u}_i \, A f(\mathbf{r}, \mathbf{u}_i, t)$.

Density: $\rho_i = m_i n_i = \langle m_i \rangle$,

$$\rho = \sum_i \rho_i .$$

Concentration: $c_i = \rho_i / \rho$.

Hydrodynamic velocity: $\rho \mathbf{v} = \sum_i \langle m_i \mathbf{u}_i \rangle$.

Internal energy density:

$$\rho E = \sum_i \left\langle \frac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) \right\rangle .$$

Mass flux: $\mathbf{J}_i = \langle m_i (\mathbf{u}_i - \mathbf{v}) \rangle$.

Stress tensor:

$$\overleftrightarrow{\mathbf{P}} = \sum_i \overleftrightarrow{\mathbf{P}}_i = \sum_i \langle m_i (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle ,$$

$$\overleftrightarrow{\pi} = \sum_i \overleftrightarrow{\pi}_i = \sum_i [\overleftrightarrow{\mathbf{P}}_i]^{(2)}$$

$$= \sum_i \left\{ \frac{1}{2} (\overleftrightarrow{\mathbf{P}}_i + \overleftrightarrow{\mathbf{P}}_i^t) - \frac{1}{3} (\overleftrightarrow{\mathbf{P}}_i : \overleftrightarrow{\mathbf{I}}) \overleftrightarrow{\mathbf{I}} \right\}$$

$$= \sum_i \langle m_i [(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle ,$$

$$P_i = \frac{1}{3} (\overleftrightarrow{\mathbf{P}}_i : \overleftrightarrow{\mathbf{I}}) \quad \overleftrightarrow{\mathbf{I}} \text{: unit tensor.}$$

Traceless symmetric part of second rank tensor $\overleftrightarrow{\mathbf{A}} : [\overleftrightarrow{\mathbf{A}}]^{(2)}$.

Heat flux:

$$\mathbf{Q} = \sum_i \mathbf{Q}_i = \sum_i \left\langle \frac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \right\rangle ,$$

$$\mathbf{Q}_i' = \mathbf{Q}_i - \frac{5}{2} \left(\frac{kT}{m_i} \right) \mathbf{J}_i .$$

Third moment: $\overleftrightarrow{\psi}_i^{(p)} = \langle m_i (\mathbf{u}_i - \mathbf{v}) [(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle$,

$$\overleftrightarrow{\phi}_i^{(3)} = \langle m_i (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle .$$

Fourth moment:

$$\overleftrightarrow{\psi}_i^{(h)} = \langle \tfrac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle .$$

Boltzmann collision integral:

$$C(f_i, f_j) = C_{ij} = \int d\mathbf{u}_j \, d\Phi \, db \, b g_{ij} (f_i' f_j' - f_i f_j) .$$

Collisional average:

$$\langle AC(f_i, f_j) \rangle = \langle AC_{ij} \rangle = \langle A \rangle_{c_{ij}} ,$$

$$\overleftrightarrow{\Lambda}_{ij}^{(p)} = \langle m_i [(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle_{c_{ij}} ,$$

$$\Lambda_{ij}^{(h)} = \langle [\tfrac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) - \tfrac{5}{2} kT] (\mathbf{u}_i - \mathbf{v}) \rangle_{c_{ij}} ,$$

$$\Lambda_{ij}^{(b)} = \langle \tfrac{1}{3} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) \rangle_{c_{ij}} ,$$

$$\Lambda_{ij}^{(f)} = \langle m_i (\mathbf{u}_i - \mathbf{v}) \rangle_{c_{ij}} .$$

Other definitions:

$$\overleftrightarrow{\mathbf{Z}}_i^{(p)} = \nabla \cdot \overleftrightarrow{\psi}_i^{(p)} - \overleftrightarrow{\pi}_i \nabla \cdot \mathbf{v} - 2[\overleftrightarrow{\pi}_i \cdot \nabla \mathbf{v}]^{(2)} - 2p_i [\nabla \mathbf{v}]^{(2)}$$

$$+ (2/\rho) \sum_j [\mathbf{J}_i \nabla p_j]^{(2)} + (2/\rho) [\mathbf{J}_i \nabla \cdot \overleftrightarrow{\pi}_j]^{(2)} ,$$

$$\mathbf{Z}_i^{(h)} = \mathbf{Z}_i'^{(h)} - \frac{5}{2} \frac{kT}{m_i} \mathbf{Z}_i^{(f)} - \frac{5}{2} \frac{kT}{m_i} \mathbf{J}_i \frac{d}{dt} \log T ,$$

$$\mathbf{Z}_i'^{(h)} = -\nabla \cdot \overleftrightarrow{\psi}_i^{(h)} - \mathbf{Q}_i \nabla \cdot \mathbf{v} + \rho^{-1} \nabla \cdot \overleftrightarrow{\mathbf{P}} : (\rho_i E_i \overleftrightarrow{\mathbf{I}} + \overleftrightarrow{\mathbf{P}}_i)$$

$$- (\mathbf{Q}_i \cdot \nabla) \mathbf{v} - \overleftrightarrow{\phi}_i^{(3)} : \nabla \mathbf{v} ,$$

$$\mathbf{Z}_i^{(f)} = -\nabla \cdot (\overleftrightarrow{\mathbf{P}}_i - c_i \overleftrightarrow{\mathbf{P}}) - \overleftrightarrow{\mathbf{P}} \cdot \nabla c_i - \mathbf{J}_i \nabla \cdot \mathbf{v} - \mathbf{J}_i ,$$

$$Z_i^{(b)} = -\tfrac{5}{3} P_i \nabla \cdot \mathbf{Q}_i + (2/3\rho) \mathbf{J}_i \nabla : \overleftrightarrow{\mathbf{P}} - \tfrac{2}{3} \overleftrightarrow{\pi}_i : \nabla \mathbf{v} .$$

APPENDIX B: REDERIVATION OF THE BOLTZMANN EQUATION FROM THE MOMENT EQUATION

In this Appendix we rederive the Boltzmann equation from the set of evolution equations for the various moments. For simplicity we consider a single component system. The evolution equations become

$$\frac{d}{dt} \rho = -\rho \nabla \cdot \mathbf{v} , \tag{B1}$$

$$\rho \frac{d}{dt} \mathbf{v} = -\nabla \cdot \overleftrightarrow{\mathbf{P}} , \tag{B2}$$

$$\rho \frac{d}{dt} E = -\nabla \cdot Q - \mathbf{P} : \nabla \mathbf{v} , \tag{B3}$$

$$\frac{d}{dt} \overleftrightarrow{\pi} = -\nabla \cdot \overleftrightarrow{\Psi}^{(p)} - \overleftrightarrow{\pi} \nabla \cdot \mathbf{v} - 2[\overleftrightarrow{\pi} : \nabla \mathbf{v}]^{(2)}$$

$$- 2p[\nabla \mathbf{v}]^{(2)} + \overleftrightarrow{\Lambda}^{(p)} , \tag{B4}$$

$$\frac{d}{dt} \mathbf{Q} = -\nabla \cdot \overleftrightarrow{\Psi}^{(h)} - \mathbf{Q} \nabla \cdot \mathbf{v} + \rho^{-1} [\nabla \cdot \overleftrightarrow{\mathbf{P}} \cdot \overleftrightarrow{\mathbf{P}}$$

$$+ \rho E \nabla \cdot \overleftrightarrow{\mathbf{P}}] - \overleftrightarrow{\psi}^{(3)} : \nabla \mathbf{v} - (\mathbf{Q} \cdot \nabla) \mathbf{v} + \Lambda^{(h)} , \tag{B5}$$

where

$$\overleftrightarrow{\Psi}^{(p)} = \int d\mathbf{u} \, m(\mathbf{u} - \mathbf{v})[\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v})]^{(2)} f ,$$

$$\overleftrightarrow{\Psi}^{(h)} = \int d\mathbf{u} \, \frac{1}{2} m(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v}) f ,$$

$$\overleftrightarrow{\Psi}^{(3)} = \int d\mathbf{u} \, m(\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v}) f ,$$

$$\overleftrightarrow{\Lambda}^{(P)} = \int d\mathbf{u} \, m[(\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v})]^{(2)} J(f, f) ,$$

$$\overleftrightarrow{\Lambda}^{(h)} = \int d\mathbf{u} \, \frac{1}{2} m(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})(\mathbf{u} - \mathbf{v}) J(f, f) ,$$

$J(ff)$ is the usual Boltzmann binary collision integral, and $\overleftrightarrow{\pi}$ is the traceless symmetric part of $\overleftrightarrow{\mathbf{P}}$.

Let

$$P_{ij} = \int d\mathbf{u} \, m(u_i - v_i)(u_j - v_j) f ,$$

$$S_{ijr} = \int d\mathbf{u} \, m(u_i - v_i)(u_j - v_j)(u_r - v_r) f ,$$

$$Q_r = \int d\mathbf{u} \, \frac{1}{2} m(\mathbf{u} - \mathbf{v} \cdot (\mathbf{u} - \mathbf{v})(u_r - v_r) f .$$

Then

$$\frac{dp_{ij}}{dt} = \int d\mathbf{u}\, m(u_i - v_i)(u_j - v_j)J(f,f)$$
$$- \frac{1}{3}\delta_{ij}\int d\mathbf{u}\, J(f,f) m(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})$$
$$- \sum_{r=1}^{3}\frac{\partial}{\partial x_r}\left[S_{ijr} - \frac{2}{3}Q_r\delta_{ij}\right]$$
$$- p_{ij}\nabla\cdot\mathbf{v} + \frac{2}{3}\ddot{\pi}:\nabla\mathbf{v}$$
$$- \sum_{r=1}^{3}\left[\frac{\partial v_i}{\partial x_r}P_{jr} + \frac{\partial v_j}{\partial x_r}P_{ir}\right]$$
$$- p\left[\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\nabla\cdot\mathbf{v}\right]$$
$$= \int d\mathbf{u}\, m(u_i - v_i)(u_j - v_j)J(f,f)$$
$$- \sum_{r=1}^{3}\frac{\partial}{\partial x_r}\left[S_{ijr} - \frac{2}{3}Q_r\delta_{ij}\right]$$
$$- p_{ij}\nabla\cdot\mathbf{v} - \delta_{ij}\Lambda^{(b)} + \frac{2}{3}\ddot{\pi}:\nabla\mathbf{v}$$
$$- \sum_{r=1}^{3}\left[\frac{\partial v_i}{\partial x_r}p_{jr} + \frac{\partial v_j}{\partial x_r}p_{ir}\right]$$
$$- p\left[\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\nabla\cdot\mathbf{v}\right].$$

But

$$\frac{dp}{dt} = \Lambda^{(b)} - \frac{5}{3}p\nabla\cdot\mathbf{v} - \frac{2}{3}\nabla\cdot\mathbf{Q} - \frac{2}{3}\ddot{\pi}:\nabla\mathbf{v},$$
$$\Lambda^{(b)} = \frac{1}{3}\int d\mathbf{u}\, m(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})J(f,f).$$

Hence,

$$\frac{dP_{ij}}{dt} = \int d\mathbf{u}\, m(u_i - v_i)(u_j - v_j)J(f,f)$$
$$- \sum_{r=1}^{3}\frac{\partial}{\partial x_r}S_{ijr} - P_{ij}\nabla\cdot\mathbf{v}$$
$$- \sum_{r=1}^{3}\left[\frac{\partial v_i}{\partial x_r}P_{jr} + \frac{\partial v_j}{\partial x_r}P_{ir}\right],$$

or

$$\frac{\partial P_{ij}}{\partial t} = \int d\mathbf{u}\, m(u_i - v_i)(u_j - v_j)J(f,f)$$
$$- \sum_{r=1}^{3}\frac{\partial}{\partial x_r}\left[S_{ijr} + v_r P_{ij}\right]$$
$$- \sum_{r=1}^{3}\left[\frac{\partial v_i}{\partial x_r}P_{jr} + \frac{\partial v_j}{\partial x_r}P_{ir}\right].$$

Consequently we have

$$\int d\mathbf{u}\, m(u_i - v_i)(u_j - v_j)\left\{\frac{\partial f}{\partial t} - \mathbf{u}\cdot\nabla f - J(f,f)\right\} = 0.$$

Similarly, by Eq. (6) we can obtain

$$\int d\mathbf{u}\,\frac{1}{2}m(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})(\mathbf{u}-\mathbf{v})$$
$$\times\left\{\frac{\partial f}{\partial t} - \mathbf{u}\cdot\nabla f - J(f,f)\right\} = 0.$$

Except for the conserved equations (1)–(3), in general, by the evolution equations we can obtain

$$\int d\mathbf{u}\, g(t,\mathbf{u},\mathbf{r})\left\{\frac{\partial f}{\partial t} - \mathbf{u}\cdot\nabla f - J(f,f)\right\} = 0,$$

where g is any tensorial polynomial in $\mathbf{u}-\mathbf{v}$. Suppose f is a class $C^{(1)}$ function. Then $h = \partial f/\partial t - \mathbf{u}\cdot\nabla f - J(f,f)$ is a $C^{(0)}$ function. Thus $\int d\mathbf{u}\, gh = 0$ for any g implies $h = 0$. Therefore the entire set of the evolution equations of the moments of f is equivalent to the Boltzmann equation.

The references given in the following are not complete. We only mention those papers or monographs that are most directly related to our work. For more information, the reader should consult the references given in Eu's papers.

[1] H. B. Callen, *Thermodynamics* (Wiley, New York, 1969).
[2] I. Grigogine, Physica **15**, 272 (1949).
[3] (a) S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases* (Cambridge U. P., London, 1961); (b) J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases* (North-Holland, Amsterdam, 1972).
[4] L. Boltzmann, *Lectures on Gas Theory*, translated by S. G. Brush (Univ. California, Berkeley, 1964).
[5] S. R. De Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962), pp. 174–184.
[6] B. C. Eu, J. Chem. Phys. **73**, 2958 (1980).
[7] H. Grad, Comm. Pure Appl. Math. **2**, 311 (1949).
[8] B. C. Eu, Ann. Phys. N. Y. **140**, 341 (1982); J. Chem. Phys. **74**, 2298, 6376 (1981) **75**, 4031, (1981).
[9] C. Cercignani, *Mathematical Methods in Kinetic Theory* (Plenum, New York, 1969), p. 51.
[10] Reference 3(b), p. 77.
[11] The solution of the Boltzmann equation has been a mathematically challenging problem extensively discussed in the literature. It is impossible to give a detailed list of references in this paper. We only mention the work of L. Arkeryd, Arch. Rational Mech. Anal. **45**, 1 (1972); **77**, 11 (1981); Bull. Am. Math. Soc. **78**, 610 (1972); Commun. Math. Phys. **86**, 475 (1982). The interested readers should consult the references given therein.
[12] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1955), p. 139.
[13] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[14] L. Cesari, *Asymptotic Behavior and Stability Problems in Ordinary Differential Equations* (Springer, New York, 1971), p. 42.
[15] N. Levinson, Duke. Math. J. **15**, 14 (1948).
[16] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations* (Heath, Boston, 1965), p. 54.
[17] See Ref. 14, p. 21.
[18] R. Bellman, *Stability Theory of Differential Equations* (McGraw-Hill, New York, 1953), p. 35.
[19] P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971).
[20] See Ref. 18, p. 82.

Fermion excitations of the nonlinear Schrödinger field in the attractive case

Piotr Garbaczewski
Institute of Theoretical Physics, University of Wroclaw, 50–205 Wroclaw, Poland[a] *and Indian Statistical Institute, Calcutta 700 035, India*

(Received 4 March 1985; accepted for publication 22 August 1985)

The recent inverse scattering method analysis by L. Martínez Alonso [J. Math. Phys. **25**, 1935 (1984)] is extended to demonstrate that the Bose quantized (attractive) nonlinear Schrödinger field in 1 + 1 dimensions, admits fermion excitations in its (quantum soliton) spectrum.

I. QUANTUM SOLITON EXCITATIONS

The nonlinear Schrödinger field in 1 + 1 dimensions

$$i\psi_t = -\psi_{xx} + 2c\psi^*\psi\psi \tag{1}$$

is quantized according to Bose statistics

$$[\psi(x), \psi^*(y)]_- = \delta(x-y),$$
$$[\psi(x), \psi(y)]_- = 0 = [\psi^*(x), \psi^*(y)]_-, \tag{2}$$

since the choice of Fermi statistics would cancel the interaction term. Hence, *a priori* there is no room for fermions in this model, except for the specialized $c \to \infty$ regime in the repulsive ($c > 0$) case. Then, indeed, the Bose model exhibits a metamorphosis into the free Fermi model, see, e.g., Ref. 1, which is accompanied by the collapse of the (Bose) Fock space $\mathscr{H}_B$ into its proper subspace (of Fermi states) $\mathscr{H}_F \subset \mathscr{H}_B$.

The state space structure in the attractive ($c < 0$) case is much more complicated[2] and does not reveal any apparent fermion (Fermi states of Bose systems[3]) content. The inverse scattering method involves here a passage from the Fock representation of the canonical commutation relations $\{\psi, \psi^*, |0)\}$ to a countable family of independent Bose fields $\{\phi_n, \phi_n^*, |0), n \geqslant 1\}$ such that $|0)$ is a common (cyclic vacuum) vector for both ψ, ψ^* and $\{\phi_n, \phi_n^*, n \geqslant 1\}$, while

$$[\phi_n(p), \phi_m^*(q)]_- = \delta_{nm}\,\delta(p-q),$$
$$[\phi_n(p), \phi_m(q)]_- = 0, \tag{3}$$

so that the extended Galilei group generators acquire the following form:[2]

$$M = \frac{1}{2}\int_{-\infty}^{+\infty} dx\,\psi^*\psi = \sum_{n\geqslant 1}\frac{n}{2}\int_{-\infty}^{+\infty} dp\,\phi_n^*(p)\,\phi_n(p),$$
$$H = \int_{-\infty}^{+\infty} dx(\psi_x^*\psi_x + c\psi^{*2}\psi^2)$$
$$= \sum_{n\geqslant 1}\int_{-\infty}^{+\infty} dp\left[\frac{p^2}{n} - \frac{c^2}{12}(n^3-n)\right]\phi_n^*(p)\,\phi_n(p), \tag{4}$$
$$P = \int_{-\infty}^{+\infty} dx\,\psi^*(-i\psi_x) = \sum_{n\geqslant 1}\int_{-\infty}^{+\infty} dp\,p\,\phi_n^*(p)\,\phi_n(p),$$
$$K = -\frac{1}{2}\int_{-\infty}^{+\infty} dx\,x\psi^*\psi$$
$$= -\sum_{n\geqslant 1}\frac{in}{2}\int_{-\infty}^{+\infty} dp\,\phi_n^*(p)\frac{\partial}{\partial p}\phi_n(p).$$

[a] Permanent address.

The eigenvectors of H due to $[H, N]_- = 0$, $N = \int_{-\infty}^{+\infty} dx\,\psi^*(x)\psi(x)$ in each n-particle sector have a standard (Bethe ansatz) form

$$|f) = \int dx_1 \cdots \int dx_n\,\xi(x_1,\ldots,x_n)\,\psi^*(x_1)\cdots\psi^*(x_n)|0).$$

Nevertheless, as follows from (4) instead of the $\{\Pi_{i=1}^n\,\psi_i^*(x_i)|0)\}$ basis, another one can be used to generate the underlying state space. Namely

$$|p_1,n_1;p_2n_2,\ldots,p_r\,n_r) = \phi_{n_1}(p_1)\cdots\phi_{n_r}^*(p_r)|0), \quad n_l \geqslant 1, \quad \forall l. \tag{5}$$

Since we have

$$[\phi_n^*(p)\,\phi_n(p), \mathscr{N}_k]_- = 0, \quad \forall\, n, k, \quad \forall\, p,$$
$$\mathscr{N}_k = \int_{-\infty}^{+\infty} dq\,\phi_k^*(q)\,\phi_k(q), \tag{6}$$

each operator $\mathscr{N}_k$ commutes with the generators M, K, P, and H of (4). Hence the single interacting Galilean (Bose) field ψ^*, ψ gives rise to a countable set of independent (free) Galilean bosons ϕ_n^* and ϕ_n with $H_n = \int dp\,\omega_n(p)\,\phi_n^*(p) \times\phi_n(p)$ and $\omega_n(p) = p^2/n - (c^2/12)(n^3-n)$.

II. QUANTUM SOLITONS AS FERMIONS

Despite the fact that in the above we deal with bosons only, the diagonal (with respect to ϕ_n^*, ϕ_n) structure of generators (4) of the extended Galilei group, together with (6), suggests the existence of state space vectors which respect the Pauli principle. After accounting for the analysis of Refs. 1, 3, and 4 it would indicate that the nonlinear Schrödinger field has Fermi states, and consequently gives rise to fermion excitations (paralleling the boson ones).

For this purpose, let us consider the following sequence $\{\Pi_n, n \geqslant 1\}$ of projection operators in the state space of our Bose system (compare, e.g., in this connection the general construction of Ref. 5):

$$\Pi_n = \sum_{s=0}^{\infty}\frac{1}{s!}\sum_{\alpha_1=1}^{n}\int dq_1\cdots\sum_{\alpha_s=1}^{n}\int dq_s\,[\sigma(\alpha_1 q_1;\ldots;\alpha_s q_s)]^2$$
$$\times\phi_{\alpha_1}^*(q_1)\cdots\phi_{\alpha_s}^*(q_s):\exp\left[-\sum_{\beta=1}^{n}\int dp\,\phi_\beta^*(p)\,\phi_\beta(p)\right]:$$
$$\times\phi_{\alpha_1}(q_1)\cdots\phi_{\alpha_s}(q_s), \tag{7}$$

where the (alternating) function $\sigma(\alpha_1 q_1;\ldots;\alpha_s q_s)$ is defined as follows:

 0022-2488/86/020612-03$02.50 © 1986 American Institute of Physics

$$\sigma(\alpha_1 q_1;\ldots;\alpha_s q_s) = \prod_{1\le j<k\le s} p_{jk},$$

$$p_{jk} = \delta_{\alpha_j\alpha_k}[\Theta(q_j - q_k) - \Theta(q_k - q_j)] + (1-\delta_{\alpha_j\alpha_k})(-1)^{1+\Theta(q_j-q_k)}, \tag{8}$$

provided $\Theta(q-p) = 1$, $q \geqslant p$, 0 otherwise.

Since $\sigma^3 = \sigma$, $\sigma = \pm 1$ depending on permutations of pairs (αq) of indices, and if coinciding pairs appear in the sequence, then $\sigma = 0$, and the analysis of Ref. 5 proves that $\forall\, n$, Π_n is a projection indeed. Moreover, if to denote $\mathscr{H}_{\mathrm{B}}$ the Hilbert space of the nonlinear Schrödinger system (3) and (4), then on its proper subspace $\mathscr{H}_{\mathrm{F}}^n = \Pi_n \mathscr{H}_{\mathrm{B}}$, the following fermion field operators (Fock representation of the CAR algebra) do automatically exist[5]:

$$b_\beta(p) = \sum_{s=0}^{\infty} \frac{(1+s)^{1/2}}{s!} \sum_{\alpha_1=1}^{n} \int dq_1 \cdots \sum_{\alpha_s=1}^{n} \int dq_s \times \sigma(\alpha_1 q_1;\ldots;\alpha_s q_s) \times \sigma(\beta p;\alpha_1 q_1;\ldots;\alpha_s q_s)\, \phi^*_{\alpha_1}(q_1)\cdots\phi^*_{\alpha_s}(q_s) \times :\exp\Big[-\sum_{\gamma=1}^{n}\int dr\, \phi^*_\gamma(r)\,\phi_\gamma(r)\Big]: \times \phi_\beta(p)\,\phi_{\alpha_1}(q_1)\cdots\phi_{\alpha_s}(q_s), \tag{9}$$

where $1 \leqslant \beta \leqslant n$ and

$$[b_\alpha(p), b^*_\beta(q)]_+ = \delta_{\alpha\beta}\,\delta(p-q)\,\Pi_n,$$
$$[b_\alpha(p), b_\beta(q)]_+ = 0, \quad 1\leqslant\alpha, \quad \beta\leqslant n, \tag{10}$$

while $b_\alpha(p)|0) = 0$, $b^*_\alpha(p)|0) = \phi^*_\alpha(p)|0)$, $\forall\, \alpha, p$.

One should realize that each projection Π_n selects in $\mathscr{H}_{\mathrm{B}}$, its proper subspace $\mathscr{H}^n_{\mathrm{F}}$, on which the respective Bose variables (i.e., ϕ^*_α, ϕ_α, $1\leqslant\alpha\leqslant n$) respect the Pauli principle. It means that the operator

$$\mathscr{P}_n = \sum_{\alpha=1}^{n} \mathscr{N}_\alpha(\mathscr{N}_\alpha - 1) \tag{11}$$

has the eigenvalue 0 on the whole of $\mathscr{H}^n_{\mathrm{F}}$. Because of (6), these Pauli-principle-saving subspaces, are the Galilei invariant sectors in $\mathscr{H}_{\mathrm{B}}$, thus giving rise to *the Galilean fermion excitations* in $\mathscr{H}_{\mathrm{B}}$.

Moreover, projections $\{\Pi_n, n\geqslant 1\}$ form a decreasing sequence

$$\Pi_n\,\Pi_{n+1} = \Pi_{n+1}. \tag{12}$$

But then, according to the standard knowledge: (1) there exists a strong limit $\Pi = \text{s-lim}\,\Pi_n$, which is a projection on $\mathscr{H}_{\mathrm{B}}$, (2) the property $\Pi_n\,\Pi = \Pi$ holds true for all n, and (3) for any vector $|f) \in \mathscr{H}_{\mathrm{B}}$ for which $\lim \Pi_n\,|f) \neq 0$, upon setting $|\psi) = \lim \Pi_n\,|f)$ we have $|\psi) \neq 0$ and $\Pi_n\,|\psi) = |\psi)$, $\forall\, n$. On the respective subspace $\Pi\,\mathscr{H}_{\mathrm{B}} = \mathscr{H}_{\mathrm{F}}$ of $\mathscr{H}_{\mathrm{B}}$ the operator

$$\mathscr{P} = \sum_{\alpha=1}^{\infty} \mathscr{N}_\alpha(\mathscr{N}_\alpha - 1)$$

has the eigenvalue 0, and the generalization of the formula (9) to $n\to\infty$ is possible. Then, however, we arrive at the conclusion that the Bose quantized nonlinear Schrödinger field with attractive coupling, in addition to bearing the infinite set of Galilean bosons, gives rise as well to the infinite set of Galilean fermions

$$\Pi M \Pi = \sum_{n\geq 1} \frac{n}{2} \int_{-\infty}^{+\infty} dp\, b^*_n(p)\, b_n(p),$$

$$\Pi H \Pi = \sum_{n\geq 1} \int_{-\infty}^{+\infty} dp \left[\frac{p^2}{n} - \frac{c^2}{12}(n^3 - n)\right] \times b^*_n(p)\, b_n(p),$$

$$\Pi P \Pi = \sum_{n\geq 1} \int_{-\infty}^{+\infty} dp\, p\, b^*_n(p)\, b_n(p),$$

$$\Pi K \Pi = -\sum_{n\geq 1} \frac{in}{2} \int_{-\infty}^{+\infty} dp\, b^*_n(p) \frac{\partial}{\partial p} b_n(p), \tag{13}$$

which live in the Hilbert space of our Bose system.

Since, *a priori*, each field $\phi_n(p)$ can be given as a function of the primary interacting fields $\psi^*(x)$, $\psi(x)$, it happens so in the case of fermions $b^*_n(p)$, $b_n(p)$. However we cannot present the corresponding formulas. As well, we do not know how the primary fields $\psi(x)$, $\psi^*(x)$ act on the Pauli-principle-saving domain $\Pi\,\mathscr{H}_{\mathrm{B}} = \mathscr{H}_{\mathrm{F}}$. Nevertheless, since

$$|\alpha_1 q_1;\ldots,\alpha_s q_s)_{\mathrm{F}} \doteq b^*_{\alpha_1}(q_1)\cdots b^*_{\alpha_s}(q_s|0) = \sigma(\alpha_1 q_1;\ldots;\alpha_s q_s)\,\phi^*_{\alpha_1}(q_1)\cdots\phi^*_{\alpha_s}(q_s|0), \tag{14}$$

the analysis of Ref. 2 apparently can be applied to determine the scalar products

$$\frac{1}{\sqrt{n!}}(x_1,\ldots,x_n|\alpha_1 q_1;\ldots;\alpha_s q_s)_{\mathrm{F}}, \quad n = n_1 + \cdots + n_s, \quad |x_1,\ldots,x_n) = \psi^*(x_1)\cdots\psi^*(x_n)|0). \tag{15}$$

It is, however, quite transparent that unlike our previous investigations[3,4] the property $[H, \Pi]_- = 0$ does not suffice to convert the Bose Hamiltonian $H = H(\psi^*, \psi)$ of (4) into the (Fermi) Hamiltonian $H_{\mathrm{F}} = \Pi H \Pi$, where the primary bosons ψ^*, ψ are simply replaced by the respective fermions. In the present case, the fermion content of the model becomes manifest on another level of the theory. Albeit, the basic (boson–fermion unduality) mechanism $H_{\mathrm{B}} = P H_{\mathrm{B}} P + (1-P) H_{\mathrm{B}} (1-P)$, $\mathrm{P} H_{\mathrm{B}} P = H_{\mathrm{F}}$ is still the same as previously, see Refs. 1, 3, and 4. A more detailed study of the issue in connection with the boson and fermion Fock space unification can be found in Refs. 6 and 7.

ACKNOWLEDGMENT

The present paper has been written during my stay at the Indian Statistical Institute in Calcutta, and I would like to thank Professor Pratul Bandyopadhyay for his hospitality.

[1]P. Garbaczewski, J. Math. Phys. **24**, 341 (1983).
[2]L. Martínez Alonso, J. Math. Phys. **25**, 1935 (1984).
[3]P. Garbaczewski, J. Math. Phys. **25**, 862 (1984).
[4]P. Garbaczewski, J. Math. Phys. **26**, 2039 (1985).
[5]P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[6]R. L. Hudson and K. R. Parthasarathy, "Unification of the fermion and boson stochastic calculus," submitted for publication.
[7]P. Garbaczewski, "Some aspects of the boson–fermion (in)equivalence: A remark on the paper by Hudson and Parthasarathy," BiBOS, Bielefeld preprint, 1985.

Soliton solutions for self-dual SU(N) gauge fields on Euclidean space

Patricio S. Letelier
Departmento de Física, Universidade de Brasília, 70.910, Brasília, DF, Brazil

(Received 13 June 1985; accepted for publication 19 September 1985)

Axially symmetric soliton solutions for self-dual SU(N) gauge fields on Euclidean four-dimensional flat space are found using an extension of the Belinsky–Zakharov solution generating technique. The new solutions depend at most on $N-1$ arbitrary solutions of the usual Laplace equations in cylindrical coordinates. The inverse scattering method using the chosen seen solution reduces to the computation of at most $N-1$ quadratures. The n-soliton solution is written in a determinantal form. Three particular cases of one-soliton solutions for the SU(5) group are exhibited.

I. INTRODUCTION

The self-dual Yang–Mills equations are more conveniently described in the R gauge, first introduced by Yang[1] for the SU(2) case and subsequently extended by Prasad,[2] Ardalan,[3] and Brihaye *et al.*[4] to the SU(N) group. For this group the self-duality equations are

$$\partial_{\bar{\xi}}(g_{\xi}\, g^{-1}) + \partial_{\bar{\zeta}}(g_{\zeta}\, g^{-1}) = 0\,, \tag{1.1}$$

where the subscripts ξ and ζ denote partial differentiation, g is an $N\times N$ Hermitian matrix with unit determinant, and ξ and ζ are complex coordinates related to the four-dimensional Euclidean-space Cartesian coordinates by

$$\xi = (1/\sqrt{2})(x+iy),\quad \zeta = (1/\sqrt{2})(z-ix_4)\,. \tag{1.2}$$

Here and in the sequel the bar operation denotes complex conjugation.

A number of authors[5] have studied the self-dual equations (1.1) using different methods. In particular, the present author[6] extended the Belinsky–Zakharov solution generating technique[7,8] (BZSGT) used in general relativity to include the axially symmetric SU(2) case. In a similar manner Papadopoulos[9] studied the SU(3) case.

The purpose of this paper is to generalize the BZSGT in order to include the SU(N) case and to present explicit pure soliton solutions (1.1). This generalization is studied in Sec. II. In Sec. III we study the equation for the "wave function" ψ_0 associated to the particular "seed solution" g_0 that is a solution to (1.1) built with solutions of Laplace equation in cylindrical coordinates. For this particular seed solution the solution of the inverse scattering problem reduces to quadratures. In Sec. IV we present a determinantal form of the n-soliton solution associated to the particular g_0 previously described. Finally, in order to visualize the solution, we exhibit three different one-soliton solutions of the SU(5) gauge theory (Sec. V).

II. THE SOLUTION GENERATING ALGORITHM

If we restrict the matrix g to be a function only of $r=(2\xi\bar{\xi})^{1/2}$ and $z=(\zeta+\bar{\zeta})/2^{1/2}$, we find that (1.1) reduces to

$$\partial_r(rg_r g^{-1}) + \partial_z(rg_z g^{-1})\,, \tag{2.1}$$

Also, we have

$$g = g^{\dagger},\quad \det g = 1\,. \tag{2.2}$$

The extension of the BZ method presented in Ref. 6 can be generalized in a straightforward manner to include the SU(N) case. Thus, we shall only present the results. The BZ method for solving Eq. (2.1) is based on the fact that the condition of integrability for the system of equations

$$D_r\psi = [(rU+\lambda V)/(\lambda^2+r^2)]\psi\,, \tag{2.3a}$$

$$D_z\psi = [(rV-\lambda U)/(\lambda^2+r^2)]\psi\,, \tag{2.3b}$$

where

$$D_r \equiv \partial_r + [2\lambda r/(\lambda^2+r^2)]\partial_{\lambda}\,, \tag{2.4a}$$

$$D_z \equiv \partial_z - [2\lambda^2/(\lambda^2+r^2)]\partial_{\lambda}\,, \tag{2.4b}$$

$$U \equiv rg_r g^{-1},\quad V \equiv rg_z g^{-1} \tag{2.5}$$

is just the same as Eq. (2.1). Here ψ is an $N\times N$ complex matrix function of r, z, and the spectral parameter λ. Putting $\lambda = 0$ in (2.3), we have that $\psi(\lambda=0)=g$. Solutions with pure soliton character are associated with solutions of Eqs. (2.3) of the form

$$\psi = \chi\psi_0\,, \tag{2.6}$$

$$\chi \equiv 1 + \sum_{k=1}^{n}\frac{R_k}{\lambda-\mu_k}\,, \tag{2.7}$$

where ψ_0 is a solution to Eqs. (2.3) for a known g, say g_0, the R_k are complex matrix functions of r and z only, and the μ_k are scalar complex functions of r and z only. The pure soliton character of the solution is associated with the particular form of χ given by (2.7), i.e., with the existence of simple poles[10] in the matrix χ. The number of poles will tell us the number of solitons appearing in the solution. Note that letting $\lambda = 0$ in (2.6), we get

$$g = (\chi|_{\lambda=0})g_0\,. \tag{2.8}$$

A condition that guarantees the fact that $g = g^{\dagger}$ is

$$g = \chi(-r^2/\bar{\lambda}, r, z)g_0[\chi(\lambda, r, z)]^{\dagger}\,. \tag{2.9}$$

From (2.3)–(2.9) we find

$$g_{ab} = (g_0)_{ab} - \sum_{kl}\frac{\bar{N}_a^{(l)}(\Gamma^{-1})_{lk}N_b^{(k)}}{\mu_k\bar{\mu}_l}\,, \tag{2.10}$$

$$\Gamma_{kl} \equiv \frac{m^{(k)}\cdot\bar{m}^{(l)}}{r^2-\mu_k\bar{\mu}_l} = \bar{\Gamma}_{lk}\,, \tag{2.11}$$

$$m^{(k)}\cdot\bar{m}^{(l)} \equiv m_a^{(k)}(g_0)_{ab}\bar{m}_b^{(l)}\,, \tag{2.12}$$

$$N_a^{(k)} \equiv m_{(k)}^b(g_0)_{ba}\,, \tag{2.13}$$

 0022-2488/86/020615-05$02.50 © 1986 American Institute of Physics

$$m_a^{(k)} \equiv m_{ob}^{(k)} M_{ba}^{(k)}, \tag{2.14}$$

$$M^{(k)} = \psi_0^{-1}|_\lambda = \mu_k, \tag{2.15}$$

$$\mu_k = \alpha_k - z \pm [(\alpha_k - z)^2 + r^2]^{1/2}. \tag{2.16}$$

The sum convention on the indices a and b is assumed; a and b run from 1 to N. The $m_{ob}^{(k)}$ and α_k are sets of arbitrary complex constants. Note that the solution (2.10) is completely determined by g_0, ψ_0, and these sets of constants. Regardless of the fact that the matrix whose elements are (2.10) is Hermitian, we have, in general, that $\det g \neq 1$. To remedy this problem, we can define a new matrix

$$g^{Ph} = g/(\det g)^{1/N} \tag{2.17}$$

that satisfies both conditions (2.2). Taking the trace of Eq. (2.1) one can prove that the new g^{Ph} is also a solution to (2.1) whenever g is a solution. The determinant of (2.10) can be explicitly computed:

$$\det g_{(n)} = (-1)^n r^{2n} \left(\prod_{l=1}^{n} |\mu_l|^{-2} \right) \det g_0 . \tag{2.18}$$

Since $\det g_0 = 1$ we conclude that we can only have an even number soliton solution associated to g_0. We can have an odd number of solitons[6] by defining a new seed solution g_0' that satisfies (2.1) and

$$(g_0')^\dagger = g_0', \quad \det g_0' = -1 . \tag{2.2'}$$

Now the odd number soliton solutions constructed with g_0', i.e., $g_n^{Ph\prime}$, satisfies (2.1) and (2.2).

Finally, we want to point out that in Refs. 6 and 7 a different condition for the determinant of g_0 and g_0' is used. One can do so due to the fact that the equation satisfied by g_0, i.e., (2.1), is also satisfied by $\hat{g}_0 = r^a g_0$ whenever a is a constant.

III. THE FUNCTIONS g_0 AND ψ_0

We shall take as our seed solution the particular solution to (2.1) given by

$$(g_0)_{ab} = \begin{cases} \eta_a \exp \phi_a, & a = b = 1,2,\ldots,s, \\ \phi_a, & a = b = s+1, s+3, \ldots, N-1, \\ \exp(ic_a), & a = b-1 = s+1, s+3, \ldots, N-1, \\ \exp(-ic_{a-1}), & a = b+1 = s+2, s+4, \ldots, N, \\ 0, & \text{otherwise}, \end{cases} \tag{3.1}$$

where the η_a are indicators that can take the values ± 1, c_a is a set of real constants, s is a number such that (i) $0 \leqslant s \leqslant N$ and (ii) $(N-s)/2$ is an integer, and the ϕ_a are functions that satisfy the usual Laplace equation in cylindrical coordinates,

$$\phi_{a,rr} + \phi_{a,r}/r + \phi_{a,zz} = 0, \tag{3.2}$$

and

$$\sum_{a=1}^{s} \phi_a = 0 . \tag{3.3}$$

The determinant associated to (3.1) is

$$\det g_0 = (-1)^{(N-s)/2} \prod_{a=1}^{s} \eta_a . \tag{3.4}$$

In order to visualize (3.1) we present the SU(5) case. For $N = 5$ we have three possible s:$s = 5$, 3, and 1. The corresponding matrices g_0 are

$$g_0 = \operatorname{diag}(\eta_1 \exp \phi_1, \ \eta_2 \exp \phi_2, \ \eta_3 \exp \phi_3, \ \eta_4 \exp \phi_4, \ \eta_5 \exp \phi_5), \tag{3.5}$$

$$g_0 = \begin{pmatrix} \eta_1 \exp \phi_1 & 0 & 0 & 0 & 0 \\ 0 & \eta_2 \exp \phi_2 & 0 & 0 & 0 \\ 0 & 0 & \eta_3 \exp \phi_3 & 0 & 0 \\ 0 & 0 & 0 & \phi_4 & \exp ic_4 \\ 0 & 0 & 0 & \exp(-ic_4) & 0 \end{pmatrix}, \tag{3.6}$$

$$g_0 = \begin{pmatrix} \eta_1 & 0 & 0 & 0 & 0 \\ 0 & \phi_2 & \exp ic_2 & 0 & 0 \\ 0 & \exp(-ic_2) & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_4 & \exp(-ic_4) \\ 0 & 0 & 0 & \exp(-ic_4) & 0 \end{pmatrix}. \tag{3.7}$$

The function ψ_0 obeys the differential equations (2.3) with g replaced by g_0, i.e.,

$$D_r \psi_0 = [(rU_0 + \lambda V_0)/(\lambda^2 + r^2)] \psi_0 , \tag{3.8a}$$

$$D_z \psi_0 = [(rV_0 - \lambda U_0)/(\lambda^2 + r^2)] \psi_0 , \tag{3.8b}$$

where $U_0 = r(g_0)_r g_0^{-1}$ and $V_0 = r(g_0)_z g_0^{-1}$. Furthermore ψ_0 must satisfy the initial condition

$$\psi_0|_{\lambda=0} = g_0 . \tag{3.9}$$

A direct verification shows that the matrix whose elements are

$$(\psi_0)_{ab} = \begin{cases} \eta_a \exp F_a, & a = b = 1,2,\ldots,s, \\ F_a, & a = b = s+1, s+3, \ldots, N-1, \\ \exp(ic_a), & a = b-1 = s+1, \\ & \quad s+3, \ldots, N-1, \\ \exp(-ic_{a-1}), & a = b+1 = s+2, s+4, \ldots, N, \\ 0, & \text{otherwise}, \end{cases} \tag{3.10}$$

is the solution to (3.8) and (3.9) associated to the particular seed solution (3.1) whenever the functions $F_a = F_a(r,z,\lambda)$

satisfy the system of equations

$$D_r F_a = (r^2\phi_{a,r} + \lambda r\phi_{a,z})/(\lambda^2 + r^2), \tag{3.11a}$$

$$D_z F_a = (r^2\phi_{a,z} - \lambda r\phi_{a,r})/(\lambda^2 + r^2), \tag{3.11b}$$

together with the initial condition

$$F_a|_{\lambda=0} = \phi_a . \tag{3.12}$$

In the final formulas (2.10)–(2.14) the matrix ψ_0 appears in the form $\psi_0|_{\lambda=\mu_k}$. Thus, to construct the soliton solutions we only need

$$F_a^{(k)} \equiv F_a|_{\lambda=\mu_k}, \tag{3.13}$$

i.e., the functions F_a along the poles' trajectories. These trajectories obey the equations[8]

$$\mu_{k,r} = 2r\mu_k/(\mu_k^2 + r^2), \quad \mu_{k,z} = -2\mu_k^2/(\mu_k^2 + r^2). \tag{3.14}$$

From (3.11)–(3.14) we get

$$r\,\partial_r F_a^{(k)} - \mu_k\,\partial_z F_a^{(k)} = r\phi_{a,r}, \tag{3.15a}$$

$$\mu_k\,\partial_r F_a^{(k)} + r\,\partial_z F_a^{(k)} = r\phi_{a,z} . \tag{3.15b}$$

Thus

$$F_a^{(k)}[\phi_a] = \frac{1}{2}\int \frac{r}{\mu_k}[(\mu_{k,r}\phi_{a,r} - \mu_{k,z}\phi_{a,z})dr + (\mu_{k,r}\phi_{a,z} + \mu_{k,z}\phi_{a,r})dz] . \tag{3.16}$$

The existence of (3.16) is guaranteed by Eq. (3.2) and the fact that $\ln\mu_k$ is also a solution to (3.2). Note that

$$(\mu_{k,r}/\mu_k)|_{\mu_k\to 0} = 2/r, \quad (\mu_{k,z}/\mu_k)|_{\mu_k\to 0} = 0 . \tag{3.17}$$

Thus, (3.16) is compatible with the initial condition (3.12). In other words, the overdetermined system of equations (3.8) for the solution (3.1) is completely determined along the poles' trajectories; its solution reduces to a single quadrature.[11] Note that (3.3) and (3.16) imply that

$$\sum_{a=1}^{s} F_a^{(k)} = 0 , \tag{3.18}$$

i.e., that the number of linearly independent functions $F_a^{(k)}$ is equal to the number of linearly independent ϕ_a.

IV. THE *n*-SOLITON SOLUTION

In the general case the elements of g^{Ph} for n simple poles can be cast in the following determinantal form:

$$g_{ab}^{Ph} = r^{-2n/N}\left(\prod_{m=1}^{n}|\mu_m|^{2(1-N)/N}\right)\frac{\det[\bar{\mu}_l\,\mu_k\Gamma_{lk} - (g_0)_{ab}^{-1}\bar{N}_a^{(l)}N_b^{(k)}]}{\det\Gamma}(g_0)_{ab}, \tag{4.1a}$$

for $(g_0)_{ab} \neq 0$, and for $(g_0)_{ab} = 0$,

$$g_{ab}^{Ph} = r^{-2n/N}\left(\prod_{m=1}^{n}|\mu_m|^{2(1-N)/N}\right)\frac{\Pi_{m'=1}^{n}|\mu_{m'}|^2\det\Gamma - \det[\bar{\mu}_l\,\mu_k\Gamma_{lk} + N_a^{(l)}N_b^{(k)}]}{\det\Gamma} . \tag{4.1b}$$

In deriving (4.1) we have made use of (2.10), (2.17), and the identities

$$\det(k_i k_j + D_{ij}) = \left[1 + \sum_{i,j=1}^{M} k_i k_j (D^{-1})_{ij}\right]\det D, \tag{4.2}$$

$$\det(k_i k_j D_{ij}) = (k_1 k_2 \cdots k_M)^2 \det D, \tag{4.3}$$

which are valid for a nonsingular $M\times M$ matrix D and an arbitrary vector k_i.

The expressions (4.1) are particularly useful in the study of the general properties of the multisoliton solutions.[12] To compute the different quantities that appear in (4.1) we need first to compute $M^{(k)}$ and $m_a^{(k)}$ as (2.10) and (2.11) indicate, we find

$$(M^{(k)})_{ab} = \begin{cases} \eta_a \exp(-F_a^{(k)}), & a = b = 1,2,\ldots,s, \\ \exp(ic_a), & a = b-1 = s+1, s+3,\ldots,N-1, \\ -F_{a-1}^{(k)}, & a = b = s+2, s+4,\ldots,N, \\ \exp(-ic_a), & a = b+1 = s+2, s+4,\ldots,N, \\ 0, & \text{otherwise}, \end{cases} \tag{4.4}$$

and

$$m_b^{(k)} = \begin{cases} \eta_b m_{ob}^{(k)}\exp(-F_b^{(k)}), & b \leqslant s, \\ m_{ob+1}^{(k)}\exp(-ic_b), & b = s+1, s+3,\ldots,N-1, \\ -m_{ob}^{(k)}F_{b-1}^{(k)} + m_{ob-1}^{(k)}\exp(ic_{b-1}), & b = s+2, s+4,\ldots,N. \end{cases} \tag{4.5}$$

Then, we get

$$N_b^{(k)} = \begin{cases} \eta_b m_{ob}^{(k)}\exp(\phi_b - F_b^{(k)}), & b \leqslant s, \\ m_{ob}^{(k)} + m_{ob+1}^{(k)}(\phi_b - F_b^{(k)})\exp(-ic_b), & b = s+1, s+3,\ldots,N-1, \\ m_{ob}^{(k)}, & b = s+2, s+4,\ldots,N, \end{cases} \tag{4.6}$$

and

$$\Gamma_{kl} = (r^2 + \mu_k \bar{\mu}_l)^{-1} \left\{ \sum_{b=1}^{s} m_{ob}^{(k)} \bar{m}_{ob}^{(l)} \exp(\phi_b - F_b^{(k)} - \bar{F}_b^{(l)}) + \sum_{b=s+1}^{N-1}{}' m_{ob+1}^{(k)} \bar{m}_{ob+1}^{(l)} (\phi_b - F_b^{(k)} - \bar{F}_b^{(l)}) \right.$$
$$\left. + \sum_{b=s+1}^{N-1}{}' [m_{ob}^{(k)} \bar{m}_{ob+1}^{(l)} \exp(ic_b) + m_{ob+1}^{(k)} \bar{m}_{ob}^{(l)} \exp(-ic_b)] \right\}, \tag{4.7}$$

where Σ' indicates a sum on $b = s+1, s+3, \ldots, N-1$.

An interesting feature of the solution (4.1) is that it is equivalent to solving the stationary, axially symmetric Einstein–$(N-2)$–Maxwell field equations.[13]

V. SU(5) ONE-SOLITONS

To illustrate the previous results we shall compute the one-soliton solutions associated to the particular SU(5) seeds solutions (3.5), (3.6), and (3.7).

For the diagonal case (3.5) we get

$$g_{ab}^{Ph} = \eta_a |r/\mu_1|^{8/5} \Delta^{-1} \exp(\phi_a) [\, |\mu_1/r|^2 \Delta - \eta_a (1 + |\mu_1/r|^2) |q_a|^2 \exp(\phi_a - 2 \operatorname{Re} F_a^{(1)})], \tag{5.1a}$$

when $a = b$, and for $a \neq b$,

$$g_{ab}^{Ph} = -\eta_a \eta_b |r/\mu_1|^{8/5} (1 + |\mu_1/r|^2) \bar{q}_a q_b \times \exp(\phi_a + \phi_b - \bar{F}_a^{(1)} - F_b^{(1)}), \tag{5.1b}$$

where

$$\Delta \equiv \sum_{b=1}^{5} |q_b|^2 \exp(\phi_b - 2 \operatorname{Re} F_b^{(1)}), \tag{5.2}$$

$$q_a \equiv m_{oa}^{(1)}. \tag{5.3}$$

Note that in this case we must have

$$\prod_{a=1}^{5} \eta_a = -1. \tag{5.4}$$

For the case (3.6) we get, when $a,b \leqslant 3$, similar formulas to (5.1), but now

$$\Delta = 2 \operatorname{Re}(q_4 \bar{q}_5 \exp ic_4) + |q_5|^2 (\phi_4 - 2 \operatorname{Re} F_4^{(1)}) + \sum_{b=1}^{3} |q_b|^2 \exp(\phi_b - 2 \operatorname{Re} F_b^{(1)}) \tag{5.5}$$

and $\eta_1 \eta_2 \eta_3 = -1$. For $a \leqslant 3$ and $b = 4$ we have

$$g_{a4}^{Ph} = -\eta_a |r/\mu_1|^{8/5} (1 + |\mu_1/r|^2) \Delta^{-1} \times [\bar{q}_a q_4 \exp(\phi_a - \bar{F}_a^{(1)}) + \bar{q}_a q_5 (\phi_4 - F_4^{(1)}) \exp(\phi_a - \bar{F}_a^{(1)} - ic_4)], \tag{5.6a}$$

and for $a \leqslant 3$ and $b = 5$,

$$g_{a5}^{Ph} = -\eta_a |r/\mu_1|^{8/5} (1 + |\mu_1/r|^2) \Delta^{-1} \bar{q}_a q_5 \times \exp(\phi_a - \bar{F}_a^{(1)}). \tag{5.6b}$$

Also,

$$g_{44}^{Ph} = |r/\mu_1|^{8/5} \Delta^{-1} [\, |\mu_1/r|^2 \Delta \phi_4 - (1 + |\mu_1/r|^2) |q_4 \times \exp ic_4 + q_5 (\phi_4 - F_4^{(1)})|^2], \tag{5.6c}$$

$$g_{45}^{Ph} = |r/\mu_1|^{8/5} \Delta^{-1} \{ |\mu_1/r|^2 \Delta \exp ic_4 - (1 + |\mu_1/r|^2) [\bar{q}_4 q_5 + |q_5|^2 (\phi_4 - \bar{F}_4^{(1)}) \exp(-ic_4)] \}, \tag{5.6d}$$

$$g_{55}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} (1 + |\mu_1/r|^2) |q_5|^2. \tag{5.6e}$$

The non listed components can be obtained from the property $g = g^{\dagger}$.

And for the (3.7) seed solution we obtain

$$g_{11}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} [\, |\mu_1|^2 \Delta + (r^2 + |\mu_1|^2) |q_1|^2], \tag{5.7a}$$

$$g_{12}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} (1 + |\mu_1/r|^2) \times [\bar{q}_1 q_2 + \bar{q}_1 q_3 (\phi_2 - F_2) \exp(-ic_2)], \tag{5.7b}$$

$$g_{13}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} (1 + |\mu_1/r|^2) \bar{q}_1 q_3, \tag{5.7c}$$

$$g_{14}^{Ph} = g_{12}^{Ph} (2 \to 4, 3 \to 5), \tag{5.7d}$$

$$g_{15}^{Ph} = g_{13}^{Ph} (3 \to 5), \tag{5.7e}$$

$$g_{22b}^{Ph} = |r/\mu_1|^{8/5} \Delta^{-1} [\, |\mu_1/r| \Delta \phi_2 - (1 + |\mu_1/r|^2) |q_1 \times \exp ic_2 + q_3 (\phi_2 - F_2^{(1)})|^2], \tag{5.7f}$$

$$g_{23}^{Ph} = |r/\mu_1|^{8/5} \Delta^{-1} \{ |\mu_1/r|^2 \Delta \exp ic_a - (1 + |\mu_1/r|^2) \times [\bar{q}_2 q_3 + |q_3|^2 (\phi_2 - \bar{F}_2^{(1)}) \exp ic_2] \}, \tag{5.7g}$$

$$g_{24}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} (1 + |\mu_1/r|^2) \times [\bar{q}_2 + \bar{q}_3 (\phi_2 - \bar{F}_2^{(1)}) \exp ic_2] \times [q_4 + q_5 (\phi_4 - F_4^{(1)}) \exp(-ic_4)], \tag{5.7h}$$

$$g_{25}^{Ph} = -|r/\mu_1|^{8/5} \Delta^{-1} (1 + |\mu_1/r|^2) \times [\bar{q}_2 q_5 + \bar{q}_3 q_5 (\phi_2 - F_2^{(1)}) \exp ic_2], \tag{5.7i}$$

$$g_{33}^{Ph} = g_{13}^{Ph} (1 \to 3), \tag{5.7j}$$

$$g_{34}^{Ph} = g_{12}^{Ph} (1 \to 3, 2 \to 4, 3 \to 5), \tag{5.7k}$$

$$g_{35}^{Ph} = g_{13}^{Ph} (1 \to 3, 3 \to 5), \tag{5.7l}$$

$$g_{44}^{Ph} = g_{22}^{Ph} (2 \to 4, 3 \to 5), \tag{5.7m}$$

$$g_{45}^{Ph} = g_{23}^{Ph} (2 \to 4, 3 \to 5), \tag{5.7n}$$

$$g_{55}^{Ph} = g_{33}^{Ph} (3 \to 5), \tag{5.7o}$$

where $g_{12}^{Ph}(2 \to 4, 3 \to 5)$, etc. means that we let $2 \to 4$ and $3 \to 5$ in the expression (5.7b), etc. For this case, Δ is given by

$$\Delta = |q_1|^2 + |q_2|^2 (\phi_2 - 2 \operatorname{Re} F_2^{(1)}) + |q_4|^2 (\phi_4 - 2 \operatorname{Re} F_4^{(1)}) + \operatorname{Re}(q_2 \bar{q}_3 \exp ic_2 + q_4 \bar{q}_5 \exp ic_4). \tag{5.8}$$

ACKNOWLEDGMENT

This work was partially supported by CNPq (a Brazilian government agency).

[1] C. N. Tang, Phys. Rev. Lett. **38**, 1377 (1977).
[2] M. K. Prasad, Phys. Rev. D **17**, 3243 (1978).

[3]F. Ardalan, Phys. Rev. D **18**, 1960 (1978).
[4]Y. Brihaye, D. B. Fairlie, J. Nuyts, and R. G. Yates, J. Math. Phys. **19**, 2528 (1978).
[5]See, for instance, R. Takasaki, Commun. Math. Phys. **94**, 35 (1984); J. Harnard, Y. Saint-Aubin, and E. Kyriakopoulos, J. Math. Phys. **23**, 2467 (1982); P. Fargács, Phys. Lett. B **99**, 232 (1981); L. Witten, Phys. Rev. D **19**, 718 (1979); and references therein.
[6]P. S. Letelier, J. Math. Phys. **23**, 1175 (1982).
[7]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)].
[8]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].
[9]D. Papadopoulos, J. Math. Phys. **26**, 2990 (1985).
[10]An extension of the BZSGT to include higher-order poles in studied in M. Francaviglia and C. Sgarra, Nuovo Cimento B **80**, 223 (1984).
[11]Similar results for a real SU(2) g_0 (general relativity case) can be found in P. S. Letelier, J. Math. Phys. **26**, 469 (1985); J. Math. Phys. **27**, 564 (1986).
[12]P. S. Letelier, Class. Quantum. Grav. **2**, 419 (1985).
[13]M. Gürses, Phys. Rev. D **30**, 486 (1984).

SO(4) Reduction of the SU(N) Yang–Mills–Dirac equations

M. Légaré
Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

J. Harnad[a)]
School of Mathematics, The Institute for Advanced Study, Princeton University, Princeton, New Jersey 08540

(Received 23 July 1985; accepted for publication 16 October 1985)

The equations in compactified Minkowski space that describe SU(N) Yang–Mills fields minimally coupled to Dirac spinor fields transforming under the lowest-dimensional representation(s) of SU(N) are reduced by the action of an SO(4) subgroup of the conformal group of space-time. The reduced systems can be interpreted as Hamiltonian systems with symmetry constrained by a condition on the associated moment maps.

I. INTRODUCTION

In a previous paper,[1] we considered the reduction of the SU(n), SO(n), and Sp(n) Yang–Mills equations under the SO(4) subgroup of the conformal group of space-time $C(3,1)$. This work was based on an earlier study[2] of the geometric formulation of invariant gauge fields under smooth group actions. These methods have also been applied to derive the dimensional reduction procedure,[3] to reduce and solve gauge[4] and matter-coupled gauge systems,[5,6] and to determine invariant spinor fields with gauge freedom.[7]

Our purpose in the following is to extend these investigations to the case of coupled Yang–Mills–Dirac equations on conformally compactified Minkowski space with massless Dirac spinor fields transforming under the lowest-dimensional representation(s) of the gauge group SU(N). For $N=2$, reductions under certain compact subgroups of $C(3,1)$, including SO(4), have already been done and invariant solutions found.[6,8] Here, we consider higher-dimensional gauge groups.

In Sec. II, we give the characterization of SU(N) gauge fields and Dirac spinors invariant under the SO(4) action. Such actions are generally characterized by a homomorphism of the isotropy subgroup of SO(4) into SU(N), but for a large class of homomorphisms, either no nontrivial invariant Dirac spinor fields exist, or, as shown in the Appendix, the field equations force the invariant spinors to be zero. Thus, only particular homomorphisms allow Yang–Mills systems coupled with one lowest-dimensional multiplet of spinors, and we consider some of these in Sec. III. More precisely, we look at the reduction of the SU($2n$) Yang–Mills–Dirac equations on the manifold $S^1\times S^3$ with Lorentzian metric (diffeomorphic to the compactified Minkowski space) for typical SO(4) embeddings. In a convenient gauge, the reduced systems are seen to be interpretable as Hamiltonian systems with U(n) symmetry constrained by the condition that the SU(n) part of the moment map (i.e., the associated conserved quantities) vanishes. Assuming that either one of the two Weyl components equals zero, we further simplify the residual systems by use of this constraint to obtain a set of one-dimensional systems interacting via inverse square potentials. Finally, we present a nontrivial invariant solution to the SU(4) coupled system, the solution on the compactified space being expressible as a solution on Minkowski space by an appropriate transformation.[6,9]

II. SO(4) INVARIANT FIELDS

First, we summarize some of the notations given in Refs. 1–6 that will be used in the following. Let $\bar{M}$ be the conformally compactified Minkowski space, identified with the group U(2) and, for simplicity, let us work on the twofold covering $U(1)\times SU(2)$, identified as $S^1\times S^3$, with points $p=(e^{i\psi},v)$, $e^{i\psi}\in U(1)$, and $v\in SU(2)$.

We also consider the following natural group actions on $U(1)\times SU(2)$: (a) left action of SU(2); L_g: $(e^{i\psi},v)\mapsto(e^{\psi},gv)$, where $g\in SU(2)$; (b) right action of SU(2); R_g: $(e^{i\psi},v)\mapsto(e^{i\psi},vg)$, where $g\in SU(2)$; (c) left action of $SU(2)\times SU(2)$: $L_{(g',g)}$; $(e^{i\psi},v)\mapsto(e^{i\psi},g'vg^{-1})$, where $g,g'\in SU(2)$; and (d) left action of the diagonal subgroup $SU(2)_D\equiv(SU(2)\times SU(2))_D$; D_g: $(e^{i\psi},v)\mapsto(e^{i\psi},gvg^{-1})$, where $g\in SU(2)$

In terms of the Cartesian coordinates $\{x^\mu\}$, the injection of the Minkowski space (M) in its compactified version $\bar{M}$ is defined by

$$e^{i\psi}=u^5+iu^0 \quad \text{and} \quad u=u^4-iu^i\sigma_i \tag{2.1a}$$

($i=1,2,3$), with $\{\sigma_i\}$ representing the Pauli matrices, where

$$u^\mu=\frac{x^\mu}{\tau},\quad u^4=\frac{(1+x_\mu x^\mu)}{2\tau},\quad u^5=\frac{(1-x_\mu x^\mu)}{2\tau}, \tag{2.1b}$$

and $\tau=[x_0^2+\frac{1}{4}(1-x_\mu x^\mu)^2]^{1/2}$. Its (singular) inverse is given by

$$(e^{i\psi},v)\mapsto\{x^\mu=u^\mu/(u^4+u^5)\}\,. \tag{2.2}$$

The subgroup $SO(4)\subset SO_0(4,2)\sim SU(2,2)/Z_2$ (where the subindex 0 specifies the identity component) acts on the (u^1,u^2,u^3,u^4) subspace, and its twofold covering $SU(2)\times SU(2)$ is the corresponding subgroup of SU(2,2). The isotropy subgroup $SU(2)\times SU(2)$ at the reference point $p_0=(e^{i\psi},1_2)$ is the diagonal subgroup $SU(2)_D$ and the orbits are the S^3 corresponding to fixed ψ.

On $U(1)\times SU(2)$, there exists a natural $SO(2)\times SO(4)$ left invariant Lorentzian metric, denoted g, which is conformal to the Minkowski metric g:

$$g=\tau^2\bar{g}=\tau^2\left(\theta^0\otimes\theta^0-\sum_i\theta^i\otimes\theta^i\right). \tag{2.3}$$

[a)] Address after 1 September 1985: Department of Pure and Applied Mathematics, Stevens Institute of Technology, Castle Point, Hoboken, New Jersey 07030.

 0022-2488/86/020620-07$02.50 © 1986 American Institute of Physics

The one-forms θ^0 and θ^i are defined as left invariant forms and they constitute a global set of orthonormal coframes on $S^1\times S^3$. Explicitly, they can be expressed as

$$\theta^0 = d\psi \quad \text{and} \quad \theta^i = \tfrac{1}{2}(v^\dagger dv)^i \quad [v\in \mathrm{SU}(2)]\,. \tag{2.4}$$

We denote the associated dual frames $\{e_\mu\}$. Since these are orthonormal, the linear generators $\{\gamma^\mu\}$ of the Clifford algebra are just the ordinary Dirac matrices. As in Ref. 6, we use the following representation:

$$\gamma^0 = \begin{bmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{bmatrix}, \quad \gamma^i = \begin{bmatrix} 0 & -\sigma_i\sigma_2 \\ \sigma_2\sigma_i & 0 \end{bmatrix},$$

and (2.5)

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{bmatrix} 1_2 & 0 \\ 0 & -1_2 \end{bmatrix}.$$

A. Invariant SU(N) Yang–Mills fields

It follows from the result of Ref. 2 that each equivalence class of principal bundles $P(\bar{M}\sim S^1\times[(\mathrm{SU}(2)\times\mathrm{SU}(2))/\mathrm{SU}(2)_D]$, $\mathrm{SU}(N))$, admitting a lift of the group action of $\mathrm{SU}(2)\times\mathrm{SU}(2)$, is characterized by a conjugacy class of homomorphisms (λ) of the isotropy subgroup $\mathrm{SU}(2)_D$ into the gauge group (structural group) $\mathrm{SU}(N)$. Since $\mathrm{SU}(2)$ is a simple and simply connected group, the problem of determination, up to conjugacy, of its homomorphisms into $\mathrm{SU}(N)$ is equivalent to the classification of the su(2) subalgebras into su(N). Mal'cev[10] and Dynkin[11] have solved this for all semisimple Lie subalgebras. In the case of $\mathrm{su}(2)\subset\mathrm{su}(N)$, these classes are in one-to-one correspondence with the systems of highest weights (spins) of (nontrivial) su(2) irreducible representations constrained so that the sum of their associated dimensions is less than or equal to N.

Among these classes, we shall mainly be interested in homomorphisms, which we shall call "homogeneous." For m,N, and $N/m\in\mathbb{N}$ (natural numbers), these correspond to SU(2) subgroups formed from N/m identical irreducible SU(2) representations $D^i(g)$ of (highest) weight $j=(m-1)/2$:

$$\lambda\colon (g,g)\in\mathrm{SU}(2)_D \mapsto 1_{N/m}\otimes D^j(g)\in\mathrm{SU}(N)\,,\quad g\in\mathrm{SU}(2)\,. \tag{2.6}$$

The explicit calculation of the SO(4) invariant gauge fields for these classes of bundles is given in Ref. 1, and we recall the results below. The invariant connections $\omega(\psi)$ obtained are

$$\omega(\psi) = H'(\psi)\otimes D^j(\tau_i)\theta^i + \Gamma(\psi)\otimes 1_m\theta^0\,, \tag{2.7}$$

where $H'\in\mathscr{H}(N/m)$, the space of $N/m\times N/m$ Hermitian matrices, $\Gamma(\psi)\in\mathrm{su}(N/m)$, $\{\tau_i\equiv\sigma_i/2i\}$, and $\psi\in S^1$. The $\Gamma(\psi)$ component may be thought of as a gauge potential on S^1 with respect to the residual gauge group. This may, of course, be gauged to zero. However, for the purposes of Euler–Lagrange variations, it is preferable to retain this gauge freedom.

The "nonhomogeneous" homomorphisms may in general be written as embeddings of the type

$$\begin{aligned}\lambda\colon (g,g)\in\mathrm{SU}(2)_D &\mapsto (1_{M_0/m_0}\otimes D^{j_0}(g))\oplus(1_{M_1/m_1}\otimes D^{j_1}(g))\\ &\oplus\cdots\oplus(1_{M_n/m_n}\otimes D^{j_n}(g))\in\mathrm{SU}\Big(\sum_{k=0}^{n} M_n\Big)\,,\end{aligned} \tag{2.8}$$

where $j_k\equiv(m_k-1)/2$, $M_k\in\mathbb{N}$, $\forall k\in\{0,1,\ldots,n\}$. A calculation shows that the resulting invariant fields correspond to an uncoupled direct sum of each homogeneous invariant field except for every pair of weights $(j,j+1)$, which in addition produces "off-diagonal" contributions. For example, if the embedding λ is

$$\begin{aligned}\lambda\colon (g,g)\in\mathrm{SU}(2)_D &\mapsto (1_{M/2j+3}\otimes D^{j+1}(g))\\ &\oplus(1_{M'/2j+1}\otimes D^j(g))\in\mathrm{SU}(M+M')\,,\end{aligned} \tag{2.9}$$

the most general SO(4) invariant gauge field has the simple form

$$\omega(\psi) = \begin{bmatrix} H_M\otimes D^{j+1}(\tau_i) & K\otimes\Omega(\tau_i) \\ -K^\dagger\otimes\Omega^\dagger(\tau_i) & H_{M'}\otimes D^j(\tau_i) \end{bmatrix}\theta^i + c(\psi)\theta^0\,, \tag{2.10}$$

where $H_M\in\mathscr{H}(M/2j+3)$, $H_{M'}\in\mathscr{H}(M'/2j+1)$, K is a $M/2j+3\times M'/2j+1$ complex matrix, the $\{\Omega(\tau_i)\}$ stands for $(2j+3)\times(2j+1)$ matrices, which are expressed in terms of Clebsch–Gordan coefficients for the coupling of j and $j+1$ to 1, and the element $c(\psi)$ belongs to the centralizer of the image $\lambda_*(\mathrm{su}(2)_D)$ in the gauge Lie algebra $\mathrm{su}(M+M')$.

B. Invariant SU(N) Dirac spinor fields

Let us define $\Psi\in\mathbb{C}^4\times\mathbb{C}^N$ as a massless Dirac spinor field transforming under the fundamental representation of $\mathrm{SU}(N)$, denoted $\tilde{D}$. With respect to our choice of Dirac matrices and orthonormal coframes, the SO(4) invariance condition reads (for details, see Refs. 6 and 7)

$$\begin{aligned}\Psi(L_{(g',g)}p) = &\begin{bmatrix} D^{(1/2,0)}(g) & 0 \\ 0 & D^{(1/2,0)*}(g) \end{bmatrix}\\ &\otimes\tilde{D}(\rho^{-1}((g',g),p))\Psi(p)\,,\end{aligned} \tag{2.11}$$

$\forall(g',g)\in\mathrm{SU}(2)\times\mathrm{SU}(2)$ and $p\in\mathrm{U}(1)\times\mathrm{SU}(2)$, where $D^{(1/2,0)}$ and $D^{(0,1/2)}$ (equivalent to the complex conjugate of $D^{(1/2,0)}$) are basic representations of $\mathrm{SL}(2,\mathbb{C})$. The "transformation function" $\rho^{-1}((g',g),p)$, which characterizes the group action on the principal fiber bundle (see Ref. 2), may be chosen independent of the point p since the homomorphism λ extends smoothly to an homomorphism (Λ) of $\mathrm{SU}(2)\times\mathrm{SU}(2)$ into $\mathrm{SU}(N)$. In fact, we may define

$$\rho^{-1}((g',g),p) = \Lambda((g',g)) = \lambda((g,g))\,. \tag{2.12}$$

Consequently, the Dirac spinor field must satisfy the following linear isotropy condition at the reference point $p_0=(e^{i\psi},1_2)$ for any homogeneous homomorphism:

$$\Psi(p_0) = \begin{bmatrix} g & 0 \\ 0 & g^* \end{bmatrix}\Psi(p_0)(1_{N/m}\otimes D^j(g))^T\,, \tag{2.13}$$

$\forall g\in\mathrm{SU}(2)$, $\Psi\in\mathbb{C}^{4\times N}$. However, Schur's lemma forbids the existence of nontrivial invariant Dirac spinor fields unless the highest weight j equals $\frac{1}{2}$. In that case, the corresponding invariant spinor takes the form

$$\Psi(\psi) = \begin{bmatrix} \xi^T(\psi)\otimes\sigma_2 \\ \eta^T(\psi)\otimes 1_2 \end{bmatrix}, \tag{2.14}$$

where $\xi,\eta\in\mathbb{C}^{n\times 1}$ are functions of $\psi\in S^1$.

We remark that spinors transforming under any representation of the gauge group will have nontrivial SO(4) invariant fields if the restriction to $\lambda(\mathrm{SU}(2)_D)$ of the represen-

tation contains at least one irreducible $j=\frac{1}{2}$ component. For example, the reduction by SO(4) of simple super Yang–Mills systems for any homogeneous homomorphism is equivalent to the sourceless Yang–Mills case since these embeddings do not provide any nontrivial invariant spinor field in the adjoint representation of SU(N).

III. SU(2n) YANG–MILLS–DIRAC EQUATIONS

In the orthonormal basis $\{\theta^{\mu}\}$ defined above, the pseudo-Riemannian connection of $\bar{g}$ ($\Gamma^{\mu}_{\nu\rho}$) is derived from the Maurer–Cartan structure equations:

$$d\theta^0=0 \quad \text{and} \quad d\theta^i+\epsilon_{ijk}\theta^j\wedge\theta^k=0\,. \tag{3.1}$$

This implies that the only nonzero components of $\Gamma^{\mu}_{\rho\nu}$ are

$$\Gamma^i_{jk}=\epsilon_{ijk}\,, \tag{3.2}$$

for

$$d\theta^{\mu}+\Gamma^{\mu}_{\rho\nu}\,\theta^{\rho}\wedge\theta^{\nu}=0\,. \tag{3.3}$$

Thus, we can write the canonical spin connection[6] as

$$\tilde{\sigma}=\tilde{\sigma}_{\rho}\theta^{\rho}=\tfrac{1}{8}[\gamma^{\mu},\gamma^{\nu}]\Gamma_{\mu\rho\nu}\theta^{\rho}=\tfrac{1}{8}[\gamma^{j},\gamma^{k}]\epsilon_{ijk}\theta^{i}\,. \tag{3.4}$$

It follows that the action ($\mathscr{A}$) on U(1)×SU(2) with orthonormal basis $\{\theta^{\mu}\}$ is given for the Yang–Mills–Dirac system by

$$\mathscr{A}=\int_{\bar{M}}\left\{-\frac{1}{2k}\,\mathrm{tr}(F\wedge *F)+\frac{i}{2}\,[\mathrm{tr}(\bar{\Psi}\gamma^{\mu}(e_{\mu}+\tilde{\sigma}_{\mu})\Psi + \bar{\Psi}\gamma^{\mu}\Psi\tilde{D}(A_{\mu})^{T})-\text{Hermitian conjugate}]V\right\}. \tag{3.5}$$

Here $V=\theta^0\wedge\theta^1\wedge\theta^2\wedge\theta^3$ is the volume element, $F=D\omega=\frac{1}{2}F^{a}_{\mu\nu}T_a\theta^{\mu}\wedge\theta^{\nu}$ is the curvature associated to the gauge field $\omega=A^{a}_{\mu}T_a\theta^{\mu}$, and $*F$ represents its dual relative to the metric $\bar{g}$. The $\{T_a\}$ forms a basis of the gauge Lie algebra such that $\mathrm{tr}(T_aT_b)=k\delta_{ab}$, and k may be any negative real constant.

The Yang–Mills–Dirac equations determined from $\mathscr{A}$ are (i) Yang–Mills,

$$*D*F=J\,, \tag{3.6}$$

which possess a one-form spinor current with values in the gauge Lie algebra

$$J=i\,\mathrm{tr}(\bar{\Psi}\gamma_{\mu}\Psi T^{T}_{a})\theta^{\mu}T_a\,; \tag{3.7}$$

and (ii) Dirac,

$$\gamma^{\mu}[(e_{\mu}+\tilde{\sigma}_{\mu})\Psi+\Psi\tilde{D}(A_{\mu})^{T}]=0\,. \tag{3.8}$$

A. Reduction

Inserting the explicit forms (2.7) and (2.14) for the respective SO(4) invariant gauge and spinor fields in the action (3.5), we arrive at the reduced action

$$\mathscr{A}_R=\int_{S^1}\mathscr{L}_R\,\theta^0\,, \tag{3.9a}$$

where

$$\mathscr{L}_R=\mathrm{tr}\{\tfrac{1}{2}[(\mathscr{D}H)^2-(1-H^2)^2] -(2ik/3)[\mathscr{D}\xi\xi^{\dagger}+\mathscr{D}\eta\eta^{\dagger}-\xi(\mathscr{D}\xi)^{\dagger}-\eta(\mathscr{D}\eta)^{\dagger}] -2k\,[(\eta\eta^{\dagger}-\xi\xi^{\dagger})H]\}\,, \tag{3.9b}$$

is the reduced Lagrangian density written in terms of a residual gauge element (Γ) of zero curvature, a set of "scalar components" (H), and spinor remnants (ξ and η). The "covariant derivatives" with respect to the residual component Γ are defined by

$$\mathscr{D}H\equiv\dot{H}+[\Gamma,H]\,, \tag{3.10a}$$

$$\mathscr{D}\xi=\dot{\xi}+\Gamma\xi\,, \tag{3.10b}$$

$$\mathscr{D}\eta\equiv\dot{\eta}+\Gamma\eta\,, \tag{3.10c}$$

where the dot indicates the derivative of the variable with respect to ψ. The variational equations consist of

$$\mathscr{D}(\mathscr{D}H)-2H(1-H^2)=-2kh'\,, \tag{3.11a}$$

$$[H,\mathscr{D}H]=(4ik/3)\,h_0\,, \tag{3.11b}$$

and

$$\mathscr{D}\xi+(3i/2)H\xi=0\,, \tag{3.11c}$$

$$\mathscr{D}\eta-(3i/2)H\eta=0\,, \tag{3.11d}$$

where $\Gamma\in\mathrm{su}(n)$, $H'\equiv H+1_n\in\mathscr{H}(n)$, h_0 is identified as the traceless part of $h\equiv\eta\eta^{\dagger}+\xi\xi^{\dagger}$, and $h'\equiv\eta\eta^{\dagger}-\xi\xi^{\dagger}$. We can check that the reduced Lagrangian density ($\mathscr{L}_R$) and the corresponding equations are left invariant by the gauge transformations $U\in\mathrm{SU}(n)$:

$$\Gamma\to U^{\dagger}\Gamma U+U^{\dagger}\dot{U}\,, \tag{3.12a}$$

$$H\to U^{\dagger}HU\,, \tag{3.12b}$$

$$\xi\to U^{\dagger}\xi\,, \tag{3.12c}$$

and

$$\eta\to U^{\dagger}\eta\,. \tag{3.12d}$$

We shall now choose a gauge in which the residual component Γ vanishes, and thus the Eqs. (3.11) simplify to

$$\ddot{H}-2H(1-H^2)=2kh'\,, \tag{3.13a}$$

$$[H,\dot{H}]=(4ik/3)\,h_0\,, \tag{3.13b}$$

and

$$\dot{\xi}+(3i/2)H\xi=0\,, \tag{3.13c}$$

$$\dot{\eta}-(3i/2)H\eta=0\,. \tag{3.13d}$$

Note that if this gauge had been fixed before the Euler–Lagrange variation (in $\mathscr{A}_R$), Eq. (3.13b) would not be obtained. Substitution of the invariant fields in (3.6) and (3.8) in the gauge $\Gamma=0$ also produces (3.13a)–(3.13d), as the SO(4) reduction derived for a multiplet of spinors transforming under the contragredient fundamental representation $\tilde{D}$*.

From the equations (3.13a), (3.13c), and (3.13d), it follows that the anti-Hermitian matrix,

$$I=[H,\dot{H}]-(4ik/3)\,h\,, \tag{3.14}$$

is conserved. This constant is related to the invariance of the system under the U(n) transformations: $H\to UHU^{\dagger}$, $\xi\to U\xi$, and $\eta\to U\eta$($U\in$U(n)). The second Yang–Mills–Dirac equation, Eq. (3.13b), may be recognized as the vanishing of the traceless (su(n)) part. Moreover, taking the trace shows that the quantity $|\xi|^2+|\eta|^2$ must be a real constant. Note that I is the sum of two terms coming, respectively, from the Yang–Mills and Dirac spinor fields.

B. Hamiltonian system

In the following, we formulate the above as a Hamiltonian system with symmetry constrained by a condition on the associated moment map.

Consider $\mathscr{H}(n)$ and $\mathbb{C}^n$ as spaces with the respective Hermitian inner products $\langle H,H'\rangle = \mathrm{tr}(HH')$, $H,H'\in\mathscr{H}(n)$, and $\langle\xi,\pi\rangle = \Sigma_{i=1}^n \xi^i\pi^{i*}$, $\xi,\pi\in\mathbb{C}^n$. Correspondingly, we define on $\mathscr{H}\times\mathscr{H}\times\mathbb{C}^n\times\mathbb{C}^n$ the symplectic structure

$$\Omega(H,P,\xi,\eta) = \mathrm{tr}(dH\wedge dP) - \frac{4ik}{3}\sum_{i=1}^{n}(d\xi^i\wedge d\xi^{i*} + d\eta^i\wedge d\eta^{i*}), \tag{3.15}$$

where $(H,P,\xi,\eta)\in\mathscr{H}\times\mathscr{H}\times\mathbb{C}^n\times\mathbb{C}^n$.

On this space, the Hamiltonian

$$\mathscr{H}_R = \tfrac{1}{2}\,\mathrm{tr}(P^2 + (1-H^2)^2) + 2k(\langle H\eta,\eta\rangle - \langle H\xi,\xi\rangle) \tag{3.16}$$

gives rise to the following (Hamilton) equations:

$$\dot{H} = P, \tag{3.17a}$$

$$\dot{P} = 2H(1-H^2) - 2k(\eta\eta^\dagger - \xi\xi^\dagger), \tag{3.17b}$$

and

$$\dot{\xi} = -(3i/2)H\xi, \quad \dot{\xi}^{*T} = (3i/2)\xi^{*T}H, \tag{3.17c}$$

$$\dot{\eta} = (3i/2)H\eta, \quad \dot{\eta}^{*T} = -(3i/2)\eta^{*T}H, \tag{3.17d}$$

which are equivalent to the three reduced equations (3.13a), (3.13c), and (3.13d) of the Yang–Mills–Dirac system. The $\mathrm{U}(n)$ action defined by

$$U:\ (H,P,\xi,\xi^*,\eta,\eta^*) \mapsto (UHU^\dagger, UPU^\dagger, U\xi, U^*\xi^*, U\eta, U^*\eta^*),$$

is symplectic and preserves the Hamiltonian. Its moment map[12] is

$$I(H,P,\xi,\xi^*,\eta,\eta^*) = [H,P] - (4ik/3)(\eta\eta^\dagger + \xi\xi^\dagger)\in u^*(n)$$
(the dual $\mathrm{U}(n)$ Lie algebra). (3.18)

Adding the condition that the $\mathrm{su}(n)$ part $(I - (1/n)\mathrm{tr}\,I)$ equals zero, reproduces the entire system (3.13a) – (3.13d) interpreted as a constrained Hamiltonian system. We also have the real constant $|\xi|^2 + |\eta|^2$, which remains free.

C. Additional reduction

If we suppose that one of the two Weyl components vanishes, that is, either $\eta = 0$ or $\xi = 0$, we can further reduce the Yang–Mills–Dirac equations with the help of the relation (3.13b). Let us assume that $\eta = 0$, the trace of the moment map I then implies that

$$|\xi|^2 = \sum_{i=1}^{n}|\xi_i|^2 = \frac{Cn}{2}, \tag{3.19}$$

where C is an arbitrary real constant. However, since H is Hermitian, it can be diagonalized with a ψ-dependent transformation in $\mathrm{U}(n)$:

$$H_D(\psi) = U(\psi)HU^\dagger(\psi) = \begin{bmatrix}\lambda_1(\psi) & & 0\\ & \ddots & \\ 0 & & \lambda_n(\psi)\end{bmatrix}, \tag{3.20}$$

with $U\in\mathrm{U}(n)$. We also define

$$\tilde{P} \equiv UPU^\dagger, \tag{3.21a}$$

$$\xi' \equiv U\xi, \quad \xi'^* \equiv U^*\xi^*. \tag{3.21b}$$

The constraint I (3.18) now reads

$$(\lambda_i - \lambda_j)\tilde{P}_{ij} = (4ik/3)(\xi'_i\,\xi'^*_j - C\delta_{ij}), \tag{3.22}$$

where i,j takes the values $1,\ldots,n$. For $C\neq 0$, it follows that $\lambda_i\neq\lambda_j$ for $i\neq j$. Substituting Hamilton's equation (3.17a) in the relation (3.22), we get an expression for the element $u(\psi)\equiv\dot{U}U^\dagger$ of the Lie algebra $\mathrm{u}(n)$ of $\mathrm{U}(n)$: (i) if $i\neq j$,

$$u_{ij} = 4ik\xi'_i\,\xi'^*_j/3(\lambda_j - \lambda_i)^2, \tag{3.23}$$

and (ii) if $i = j$, the terms u_{ij} are left undetermined since they correspond to the elements of the centralizer of H_D in $\mathrm{u}(n)$, and hence can be ignored.

Upon substitution of (3.23) in the two remaining Yang–Mills–Dirac equations, we derive from the diagonal terms of (3.13a) that

$$\ddot{\lambda}_i + \sum_{j\neq i}\frac{32k^2C^2}{9(\lambda_j - \lambda_i)^3} - 2\lambda_i(1-\lambda_i^2) - 2kC = 0, \tag{3.24}$$

for every $i = 1,\ldots,n$; while the off-diagonal contributions are automatically satisfied. Finally, Eq. (3.24c) reduces to

$$\dot{\xi}'_i - \frac{4ik}{3}C\sum_{j\neq i}\frac{1}{(\lambda_j - \lambda_i)^2}\xi'_i + \frac{3i}{2}\lambda_i\xi'_i = 0, \tag{3.25}$$

for every $i = 1,\ldots,n$. Setting $\xi = 0$ instead of η leads to a similar set of equations.

We remark that the system (3.13) can be regarded as a set of n one-dimensional systems with quartic potentials and Calogero type interaction.[13,14] Once solved for the eigenvalues $\{\lambda_i\}$, ξ is determined by quadrature and the solution $U(\psi)$ from the definition of u. In the case of vanishing Fermi fields, we have $\xi = 0$, $\eta = 0$, and $C = 0$ in Eq. (3.24). It is then found that U is a constant and (3.24) decouples giving the general solution in terms of elliptic functions as in Ref. 1.

We have not been able to integrate (3.24) in general, but in the next section, a particular solution to (3.13) is presented.

D. SU(4) solution

As noted above, the moment map splits into the sum of two parts, corresponding to the gauge field and spinor field contributions. We shall derive a specific solution for the case where not just the sum, but each term separately vanishes. That is, we assume

$$[H,\dot{H}] = 0, \tag{3.26a}$$

and

$$\eta\eta^\dagger + \xi\xi^\dagger = [(|\xi|^2 + |\eta|^2)/n]\,1_n. \tag{3.26b}$$

However, Eq. (3.26b) can only be satisfied by nontrivial ξ and η if $n = 1$ or 2. Then $n = 1$ case corresponds to the SU(2) Yang–Mills–Dirac system, which has been solved in Refs. 6 and 8. For $n = 2$, (3.26b) implies that

$$\xi^\dagger\eta = 0, \tag{3.27a}$$

and

$$|\xi|^2 = |\eta|^2. \tag{3.27b}$$

Let us suppose for simplicity a normalization

$|\xi|^2 = |\eta|^2 = 1$, and initial conditions that respect the constraint (3.27a):

$$\xi_0 = \begin{bmatrix} e^{i\phi} \\ 0 \end{bmatrix} \quad \text{and} \quad \eta_0 = \begin{bmatrix} 0 \\ e^{i\phi'} \end{bmatrix}, \tag{3.28}$$

with real numbers ϕ,ϕ'. We can thus express $\xi(\psi)$ and $\eta(\psi)$ as

$$\xi(\psi) = U(\psi)\xi_0 \quad \text{and} \quad \eta(\psi) = V(\psi)\eta_0, \tag{3.29}$$

where $U,V \in U(2)$ depend on the parameter $\psi \in S^1$. Consequently, the relation (3.26b) becomes

$$U\sigma_3 U^\dagger - V\sigma_3 V^\dagger = 0, \tag{3.30}$$

and the first Yang–Mills–Dirac equation takes the form

$$\ddot{H} - 2H(1 - H^2) = 2kU\sigma_3 U^\dagger. \tag{3.31}$$

However, (3.26a) requires that $H(\psi) = h^0(\psi)1_2 + h^3(\psi)U\sigma_3 U^\dagger$, where h^0 and h^3 are real functions of ψ. From the Dirac equation, it then follows that (3.31) is solved by

$$H(\psi) = \begin{bmatrix} \lambda(\psi) & 0 \\ 0 & \omega(\psi) \end{bmatrix}, \tag{3.32}$$

where λ and ω satisfy, respectively, the equations

$$\ddot{\lambda} - 2\lambda + 2\lambda^3 - 2k = 0, \tag{3.33a}$$

and

$$\ddot{\omega} - 2\omega + 2\omega^3 + 2k = 0. \tag{3.33b}$$

The solutions to (3.33) can be expressed in terms of elliptic functions and correspondingly, the Dirac spinor solutions are given by

$$\xi(\psi) = \begin{bmatrix} e^{i\phi} \exp((3/2i)\int_{\psi_0}^{\psi} \lambda\, d\psi) \\ 0 \end{bmatrix}, \tag{3.34a}$$

and

$$\eta(\psi) = \begin{bmatrix} 0 \\ e^{i\phi'} \exp((-3/2i)\int_{\psi_0}^{\psi} \omega\, d\psi) \end{bmatrix}. \tag{3.34b}$$

For a different choice of initial condition $g\xi_0$ and $g\eta_0$, with g a constant element of U(2), the solution can be written as $g\xi(\psi)$, $g\eta(\psi)$, and $gH(\psi)g^\dagger$,

IV. SUMMARY

In this work, we have examined the reduction by SO(4) symmetry of SU(N) Yang–Mills fields minimally coupled to massless Dirac spinor fields transforming under the lowest-dimensional representation(s) of SU(N). We showed that only a restricted class of homomorphisms characterizing the SO(4) invariant SU(N) gauge fields allows nonvanishing invariant Dirac spinor fields. These homomorphisms are specified by sets of consecutive spins (highest weights) increasing by one unit and starting with value $\frac{1}{2}$. We also explicitly reduced the SU($2n$) Yang–Mills–Dirac systems corresponding to the "homogeneous" homomorphisms with spin-$\frac{1}{2}$. An interpretation of these systems in terms of Hamiltonian systems with symmetry constrained by a condition on the associated moment maps was formulated. This condition requires the vanishing of the traceless part of the conserved quantity, which is composed of the sum of contributions coming from the gauge and spinor fields. In the case where each contribution of the traceless part equals zero, the spinor fields is trivial for every $n > 2$ and the solution to the corresponding sourceless Yang–Mills systems is presented in Ref. 1. For the gauge groups SU(2) ($n = 1$) and SU(4) ($n = 2$), the coupled systems can be completely solved in terms of elliptic functions, with nontrivial spinor solutions. Finally, setting one of the two residual spinor components to zero, we were able to further reduce the SU($2n$) Yang–Mills–Dirac systems to derive a set of one-dimensional systems interacting via a Calogero-type potential.

ACKNOWLEDGMENTS

One of the authors (M.L.) gratefully acknowledges scholarships from the National Sciences and Engineering Research Council of Canada and the Université de Montréal. This work is supported in part through funds provided by the United States Department of Energy (DOE) under Contract No. DE-AC02-76ER03069.

APPENDIX: NONHOMOGENEOUS HOMOMORPHISMS

We discuss here the reduction of the Yang–Mills–Dirac equations by nonhomogeneous homomorphisms of SU(2). Specifically, we find from the field equations that the homomorphisms that may lead to nonzero SO(4) invariant spinor solutions are consisting of a sequence (2.7) with $j_0 = \frac{1}{2}, j_1 = \frac{3}{2}, \ldots, j_n = \frac{1}{2} + n$ ($n > 1$). To see this, we consider the θ^0 component (or constraint component) of the Yang–Mills equations with spinor sources.

Let us write any homomorphism λ as a sum of homogeneous parts associated with disjoint sets of consecutive highest weights:

$$\{\tfrac{1}{2},\tfrac{3}{2},\ldots,\tfrac{1}{2}+n\} \oplus \{j_0^1, j_0^1+1,\ldots,j_0^1+n^1\} \oplus \cdots \oplus \{j_0^l, j_0^l+1,\ldots j_0^l+n^l\} \quad (n^1,\ldots,n^l, l \in \mathbb{N}\cup\{0\}),$$

$$\lambda\colon (g,g) \in SU(2)_D \mapsto \sum_{a=0}^{l} \sum_{i=0}^{n^a} \oplus (1_{m_i^a} \otimes D^{j_0^a + i}(g)) \in SU\left(\sum_{a=0}^{l} \sum_{i=0}^{n^a} M_i^a\right), \tag{A1}$$

with $j_0^0 = \frac{1}{2}$, $M_i^a = [(2j_0^a + 1) + 2i]m_i^a$, $m_i^a, n^a \in \mathbb{N}\cup\{0\}$.

In a convenient gauge, we know that the θ^0 contribution to the SO(4) invariant gauge field can be made to vanish. Corresponding to the homomorphism (4.1), the invariant field ω is expressed as the direct sum of overlapping sums (associated to each disjoint set) of contributions (2.10). For instance, if $l = 0$ (i.e., one set of consecutive weights with $j_0^0 = \frac{1}{2}$, $M_i = M_i^0$, $m_i = m_i^0$, and $n = n^0$), then

$$\omega(\psi) \equiv \begin{bmatrix} H_0 \otimes \tau_i & G_0 \otimes \Omega^{(3/2,1/2)} & & & & & \theta^i, \\ -G_0^\dagger \otimes \Omega^{(3/2,1/2)\dagger}(\tau_i) & H_1 \otimes D^{3/2}(\tau_i) & & & & & \\ & & \ddots & & & & \\ & 0 & & H_{n-2} \otimes D^{n-3/2}(\tau_i) & G_{n-2} \otimes \Omega^{(n-1/2,n-3/2)}(\tau_i) & 0 & \\ & & & -G_{n-2}^\dagger \otimes \Omega^{(n-1/2,n-3/2)\dagger}(\tau_i) & H_{n-1} \otimes D^{n-1/2}(\tau_i) & G_{n-1} \otimes \Omega^{(n-1/2,n-1/2)}(\tau_i) & \\ & & & 0 & -G_{n-1}^\dagger \otimes \Omega^{(n+1/2,n-1/2)\dagger}(\tau_i) & H_n \otimes D^{n+1/2}(\tau_i) & \end{bmatrix} \tag{A2}$$

where $H_i \in \mathcal{H}(m_i)$, G_i are complex matrices, and the matrices $\Omega^{(j+1,j)}(\tau_i)$ are expressed in terms of Clebsch–Gordan coefficients coupling $j+1$ and j to 1. We add (direct sum) to the right-hand side of (A2) a similar matrix expression for each supplementary disjoint sequence.

As we recall from Sec. II, only the $j=\frac{1}{2}$ homogeneous part of the embedding allows a nontrivial contribution to the SO(4) invariant spinor field transforming under the fundamental representation of the gauge group. Explicitly, we get, for $\Psi \in \mathbb{C}^{4\times \Sigma_{il} M_i^l}$,

$$\Psi(\psi) = \begin{bmatrix} \xi^T \otimes \sigma_2 & A=0 \\ \eta^T \otimes 1_2 & \end{bmatrix}, \tag{A3}$$

where $A \in \mathbb{C}^{4\times M'}, M' = \Sigma_{i=1}^n M_i + \Sigma_{a=1}^l \Sigma_{i=0}^{n^l} M_i^a$, and $\xi, \eta \in \mathbb{C}^{m_0 \times 1}$.

We now insert the invariant fields into Eq. (3.6) and evaluate the θ^0 component of each member. The θ^0 part of the spinor current (J) is found to be

$$J|_{\theta^0} = i\alpha(|\eta|^2 + |\xi|^2)(-(1/M_0)1_{M_0} \oplus (1/M')1_{M'})\,\theta^0, \tag{A4}$$

where $\alpha = (-kM_0M'/(M_0+M'))$ is a constant.

We also compute the part $*D*F|_{\theta^0}$ of the Yang–Mills–Dirac equation (3.6). However, each disjoint sequence is associated with a different element [i.e., matrix of type (A2)] of the direct sum in the expression for the invariant ω and the field equations do not mix the components of different elements. Thus, let us consider the element corresponding to the sequence of highest weights beginning with $\frac{1}{2}$: $\{\frac{1}{2},\frac{3}{2},\ldots,j_n=\frac{1}{2}+n\}$. Taking the trace on each side of the θ^0 contribution of (3.6), we derive that

$$\begin{aligned} (1)\quad & [3/(2j_n+1)]\,\mathcal{G}_{n-1} = i(\alpha/M')(|\eta|^2+|\xi|^2)m_n\,, \\ (2\leqslant i\leqslant n)\quad & [3/(2j_{n+1-i}+1)](-\mathcal{G}_{n+1-i}+\mathcal{G}_{n-i}) \\ & \quad = i(\alpha/M')(|\eta|^2+|\xi|^2)m_{n+1-i}\,, \\ (n+1)\quad & 3\,\mathcal{G}_0 = i\alpha(|\eta|^2+|\xi|^2)\,, \end{aligned} \tag{A5}$$

where by definition

$$\mathcal{G}_i \equiv \mathrm{tr}(G_i\dot{G}_i^+ - \dot{G}_iG_i^+)\,, \tag{A6}$$

for $i=1,\ldots,n+1$.

From these $(n+1)$ relations, it follows that

$$\frac{M_i}{M'}\sum_{i=1}^{n}(|\eta|^2+|\xi|^2) = (|\eta|^2+|\xi|^2)\,. \tag{A7}$$

But this is only satisfied if either (1) $m_i^a = 0$ for every $a \geqslant 1$ and every $i \in \{0,\ldots,n^a\}$, or (2) $|\eta|^2+|\xi|^2=0$, which implies that $\xi=\eta=0$.

We thus conclude from the reduced field equations (θ^0 part) that only those "nonhomogeneous" homomorphisms associated with the sequence of spins of SU(2): $\{\frac{1}{2},\frac{3}{2},\ldots,\frac{1}{2}+n\}$, may lead to a coupling with a nontrivial SO(4) invariant lowest-dimensional multiplet of spinors. All the other nonhomogeneous homomorphisms require that $\xi=\eta=0$ and hence reduce the problem to Yang–Mills systems without sources.

[1]M. Légaré and J. Harnard, J. Math. Phys. **25**, 1542 (1984).
[2]J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980).
[3]J. Harnad, S. Shnider, and J. Tafel, Lett. Math. Phys. **4**, 107 (1980).
[4]J.-P. Antoine and M. Jacques, Class. Quantum Gravit. **1**, 431 (1984).
[5]L. Vinet, Phys. Rev. D **24**, 3179 (1981).
[6]J. Doneux, Y. Saint-Aubin, and L. Vinet, Phys. Rev. D **25**, 484 (1982).
[7]M. Légaré, J. Math. Phys. **24**, 1219 (1983).
[8]K. Meetz, Z. Phys. C **6**, 41 (1980).
[9]Y. Choquet-Bruhat and D. Christodoulou, Ann. Scient. École Norm. Sup., Sér. 4 **14**, 481 (1981).
[10]A. I. Mal'cev, Am. Math. Soc. Transl., Ser. 1 **9**, 173 (1962).
[11]E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2 **6**, 111 (1965).
[12]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), Chap. 4.
[13]F. Calogero, J. Math. Phys. **12**, 419 (1971).
[14]D. Kazhdan, B. Kostant, and S. Sternberg, Comm. Pure Appl. Math. **XXXI**, 481 (1978).

Two-loop Feynman integrals in the physical light-cone gauge

George Leibbrandt
Department of Mathematics and Statistics, University of Guelph, Ontario, N1G 2W1, Canada

Su-Long Nyeo
Department of Physics, University of Guelph, Ontario, N1G 2W1, Canada

(Received 20 March 1985, accepted for publication 3 October 1985)

The feasibility of doing multiloop calculations in the noncovariant light-cone gauge $n_\mu A^a_\mu = 0$, $n^2 = 0$, is investigated by evaluating various Feynman integrals arising in a two-loop Yang–Mills self-energy. Application of a consistent prescription for $(q \cdot n)^{-1}$ is essential.

I. INTRODUCTION

Old-fashioned perturbation theory, with its emphasis on Feynman diagrams, continues to play an essential role in quantum field theory. The success of the perturbative approach depends to a large measure on the accurate evaluation of multiloop Feynman integrals.[1] There exist, of course, many varieties of these integrals, but here we shall only be concerned with the distinction between covariant-gauge and noncovariant-gauge Feynman integrals. As the name implies, covariant-gauge Feynman integrals occur in theories that are quantized in a covariant gauge, such as the Landau or Feynman gauge, while noncovariant-gauge integrals arise whenever a noncovariant gauge is employed, such as the axial and planar gauges,[2] or the light-cone gauge.[3–5] It is no secret that noncovariant-gauge integrals are much trickier to handle than their covariant counterparts, which may explain, in turn, the absence of detailed calculations beyond the one-loop level, both in quantum chromodynamics (QCD) and supersymmetric theories. (For multiloop covariant-gauge Feynman integrals, we refer to Ref. 1).

The purpose of this article is (1) to demonstrate the feasibility of performing accurate two-loop calculations in the noncovariant light-cone gauge and (2) to highlight some of the technical difficulties symptomatic of this peculiar gauge. Specifically we shall evaluate the two-loop Yang–Mills (YM) self-energy in Fig. 1, having overlapping divergences.

II. REVIEW OF LIGHT-CONE GAUGE

The light-cone gauge is specified by $n_\mu A^a_\mu = 0$, $\mu = 0,1,2,3$, where n_μ is a constant vector with $n^2 = 0$ and A^a_μ is the gauge field with a the group index. The propagator in this gauge reads

$$G^{ab}_{\mu\nu}(q) = \frac{-i\delta^{ab}}{(q^2 + i\epsilon)}\left(\delta_{\mu\nu} - \frac{n_\mu q_\nu + n_\nu q_\mu}{n \cdot q}\right), \quad \epsilon > 0; \tag{1}$$

we use a $(+,-,-,-)$ metric and employ dimensional regularization in a space-time of 2ω dimensions. Our prescription for the $(q \cdot n)^{-1}$ term in the resulting Feynman integrals is[4]

$$\frac{1}{q \cdot n} \to \lim_{\epsilon \to 0} \frac{q \cdot n^*}{q \cdot n\, q \cdot n^* + i\epsilon}, \quad \epsilon > 0,$$
$$n_\mu = (n_0, \mathbf{n}), \quad n^*_\mu = (n_0, -\mathbf{n}), \tag{2}$$

leading to well-defined, local momentum integrals that respect naive power counting.

III. TWO-LOOP YM SELF-ENERGY

The amplitude for the two-loop self-energy shown in Fig. 1 reads

$$\Pi^{af}_{\mu\nu}(p) = \frac{1}{3!}\int\int \frac{dq\,dk}{(2\pi)^{4\omega}} V^{abcd}_{\mu\lambda\sigma\rho} G^{dg}_{\rho\alpha}(q) \times G^{ch}_{\sigma\beta}(p-k-q) G^{be}_{\lambda\xi}(k) V^{ghef}_{\alpha\beta\xi\nu}, \tag{3}$$
$$dq \equiv d^{2\omega}q, \quad dk \equiv d^{2\omega}k,$$

where

$$V^{abcd}_{\mu\lambda\sigma\rho} = -ig^2[f^{abe}f^{cde}(\delta_{\mu\sigma}\delta_{\lambda\rho} - \delta_{\mu\rho}\delta_{\lambda\sigma}) + f^{ace}f^{bde}(\delta_{\mu\lambda}\delta_{\sigma\rho} - \delta_{\mu\rho}\delta_{\lambda\sigma}) + f^{ade}f^{cbe}(\delta_{\mu\sigma}\delta_{\lambda\rho} - \delta_{\mu\lambda}\delta_{\sigma\rho})] \tag{4}$$

denotes the four-gluon vertex of zero-loop order. Substituting (1) and (4) into (3), shifting the variables of integration, and simplifying the resulting expressions, we get

$$\Pi^{af}_{\mu\nu}(p) = \frac{-3ig^4N^2(2\omega-3)}{2(2\pi)^{4\omega}}\delta^{af}[A(p)\delta_{\mu\nu} - (B_\mu(p)n_\nu + B_\nu(p)n_\mu) - n_\mu n_\nu C(p)], \tag{5}$$

where $N = 3$ for SU(3), and

$$A(p) = \int\int \frac{dq\,dk}{q^2k^2(p-k-q)^2}, \tag{6a}$$

$$B_\mu(p) = \int\int \frac{dq\,dk\,q_\mu}{q^2k^2(p-k-q)^2 q \cdot n}, \tag{6b}$$

$$C(p) = \int\int \frac{dq\,dk\,k \cdot q}{q^2k^2(p-k-q)^2 q \cdot n\, k \cdot n}. \tag{6c}$$

The remaining portion of this paper is devoted to a systematic study of integrals $A(p)$, $B_\mu(p)$, and $C(p)$.

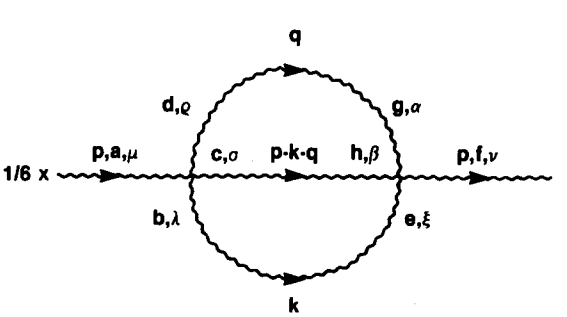


FIG. 1. Two-loop Yang–Mills self-energy.

 0022-2488/86/020627-06$02.50 © 1986 American Institute of Physics

A. The integral $A(p)$

Evaluation of the integral (6a) is unproblematic and leads to[1]

$$A(p) = i(-\pi)^{\omega}\Gamma(2-\omega)B(\omega-1,\omega-1) \times\int \frac{dq}{q^2[(q-p)^2]^{2-\omega}}$$
$$= -(-\pi)^{2\omega}\Gamma(3-2\omega)B(\omega-1,\omega-1) \times B(\omega-1,2\omega-2)(p^2)^{2\omega-3},$$
$$A(p) = -(-\pi)^{2\omega}\Gamma(\epsilon)R, \quad \epsilon\equiv 4-2\omega, \tag{7}$$
$$R = (3-2\omega)^{-1}B(\omega-1,\omega-1) \times B(\omega-1,2\omega-2)(p^2)^{2\omega-3}.$$

B. The integral $B_\mu(p)$

This is a typical noncovariant-gauge integral. Since the k-integral is covariant, it can easily be integrated to give

$$B_\mu(p) = i(-\pi)^{\omega}\Gamma(2-\omega)B(\omega-1,\omega-1) \times\int \frac{dq\, q_\mu}{q^2[(q-p)^2]^{2-\omega}q\cdot n}. \tag{8}$$

To evaluate the q-integral we proceed as follows.

(1) Perform a Wick rotation to Euclidean space and use prescription (2), replacing $(q\cdot n)^{-1}$ by[4]

$$\frac{1}{q\cdot n} \to \frac{-(\mathbf{q}\cdot\mathbf{n}+in_0q_4)}{(\mathbf{q}\cdot\mathbf{n})^2+n_0^2q_4^2}. \tag{9}$$

(2) Exponentiate all propagators via

$$\frac{1}{A^N} = \frac{1}{\Gamma(N)}\int_0^\infty d\alpha\,\alpha^{N-1}\exp(-\alpha A), \quad A>0. \tag{10}$$

(3) Write $d^{2\omega}q = d^{2\omega-1}\mathbf{q}\,dq_4$; this separation is necessary since (9) is *not* Lorentz invariant; finally, integrate over q_4- and $\mathbf{q}$-space by employing the basic formulas[4]

$$\int_{-\infty}^{+\infty} dq_4\exp(-aq_4^2+2bq_4p_4) = \left(\frac{\pi}{a}\right)^{1/2}\exp\left(\frac{b^2p_4^2}{a}\right), \tag{11a}$$

$$\int d^{2\omega-1}\mathbf{q}\exp[-\alpha\mathbf{q}^2+2\beta\mathbf{q}\cdot\mathbf{p}-\gamma(\mathbf{q}\cdot\mathbf{n})^2] = \frac{(\pi/\alpha)^{\omega-1/2}\alpha^{1/2}}{(\alpha+\gamma\mathbf{n}^2)^{1/2}}\exp\left[\frac{\beta^2\mathbf{p}^2}{\alpha}-\frac{\gamma\beta^2(\mathbf{p}\cdot\mathbf{n})^2}{\alpha(\alpha+\gamma\mathbf{n}^2)}\right], \tag{11b}$$

with additional formulas listed in Appendix B.

These three steps enable us to write the q-integral (in Minkowski space) as[6]

$$\int \frac{dq\,q_\mu}{q^2[(q-p)^2]^\sigma q\cdot n} = \frac{i(-\pi)^{\omega}\Gamma(\sigma+1-\omega)n^*_\mu}{\Gamma(\sigma)n\cdot n^*}\int_0^1 dx\,dy\,y^{\omega-2}H^{\omega-\sigma-1}$$
$$+\frac{2i(-\pi)^{\omega}\Gamma(\sigma+2-\omega)p\cdot n^*p_\mu}{\Gamma(\sigma)n\cdot n^*}\int_0^1 dx\,dy\,y^{\omega-1}H^{\omega-\sigma-2}$$
$$-\frac{2i(-\pi)^{\omega}\Gamma(\sigma+2-\omega)p\cdot n\,p\cdot n^*n^*_\mu}{\Gamma(\sigma)(n\cdot n^*)^2}\int_0^1 dx\,dy\,x\,y^{\omega-1}H^{\omega-\sigma-2}$$
$$-\frac{2i(-\pi)^{\omega}\Gamma(\sigma+2-\omega)(p\cdot n^*)^2n_\mu}{\Gamma(\sigma)(n\cdot n^*)^2}\int_0^1 dx\,dy\,x\,y^{\omega-1}H^{\omega-\sigma-2}, \tag{12}$$

with $H=(1-y)\,p^2+2xy\,p\cdot n\,p\cdot n^*/n\cdot n^*$ and σ a complex number. Substitution of (12) with $\sigma=2-\omega$ into (8) yields the following answer for the *divergent* part of the double integral:

$$B_\mu(p) = \frac{(-\pi)^{2\omega}\Gamma(\epsilon)}{2n\cdot n^*}\left(p^2n^*_\mu-2p\cdot n^*p_\mu+\frac{2p\cdot n\,p\cdot n^*}{n\cdot n^*}n^*_\mu+\frac{(p\cdot n^*)^2}{n\cdot n^*}n_\mu\right), \quad \omega\to 2, \tag{13}$$

which is seen to possess only a *simple* pole.

C. The overlapping integral $C(p)$

The integral (6c) is particularly challenging, due to the propagator $[(p-k-q)^2]^{-1}$ and the light-cone gauge-related factors $(q\cdot n)^{-1}$ and $(k\cdot n)^{-1}$. There are several ways of attacking this difficult integral, but here we shall only concentrate on two distinct methods. The first method, presented below, is probably the most direct approach; a second method is summarized, for comparison, in Appendix A. Consider

$$C(p) = \int\int \frac{dk\,dq\,k\cdot q}{k^2q^2(p-k-q)^2k\cdot n\,q\cdot n}, \tag{6c}$$

$$= \int \frac{dk\,k_\mu}{k^2\,k\cdot n}\int \frac{dq\,q_\mu}{q^2(p-k-q)^2q\cdot n}, \tag{14}$$

$$C(p) = \frac{i(-\pi)^{\omega}\Gamma(2-\omega)}{n\cdot n^*}\int\frac{dk\,k\cdot n^*}{k^2\,k\cdot n}\int_0^1 dx\,dy\,y^{\omega-2}H^{\omega-2}+\frac{2i(-\pi)^{\omega}\Gamma(3-\omega)}{n\cdot n^*}\int\frac{dk(p-k)\cdot n^*\,k\cdot(p-k)}{k^2\,k\cdot n}$$

$$\times\int_0^1 dx\,dy\,y^{\omega-1}H^{\omega-3}-\frac{2i(-\pi)^{\omega}\Gamma(3-\omega)}{(n\cdot n^*)^2}\int\frac{dk(p-k)\cdot n\,(p-k)\cdot n^*\,k\cdot n^*}{k^2\,k\cdot n}\int_0^1 dx\,dy\,x\,y^{\omega-1}H^{\omega-3}$$

$$-\frac{2i(-\pi)^{\omega}\Gamma(3-\omega)}{(n\cdot n^*)^2}\int\frac{dk\,[(p-k)\cdot n^*]^2}{k^2}\int_0^1 dx\,dy\,x\,y^{\omega-1}H^{\omega-3}, \tag{15}$$

with

$$H\equiv(1-y)(p-k)^2+2xy(p-k)\cdot n\,(p-k)\cdot n^*/n\cdot n^*.$$

Thus

$$C(p)=\sum_{i=1}^{4}I_i, \tag{16}$$

where

$$I_1=\frac{i(-\pi)^{\omega}\Gamma(2-\omega)}{2(\omega-1)}\int\frac{dk(p-k)\cdot n^*}{(p-k)^2(p-k)\cdot n\,k\cdot n\,k\cdot n^*}\int_0^1 dy\,y^{\omega-3}\{K^{\omega-1}-[(1-y)k^2]^{\omega-1}\}, \tag{17a}$$

$$I_2=\frac{i(-\pi)^{\omega}\Gamma(3-\omega)}{\omega-2}\int\frac{dk\,k\cdot(p-k)}{(p-k)^2k\cdot n\,(p-k)\cdot n}\int_0^1 dy\,y^{\omega-2}\{K^{\omega-2}-[(1-y)k^2]^{\omega-2}\}, \tag{17b}$$

$$I_3=\frac{i(-\pi)^{\omega}\Gamma(2-\omega)}{n\cdot n^*}\int\frac{dk(p-k)\cdot n^*}{(p-k)^2(p-k)\cdot n}\int_0^1 dy\,y^{\omega-2}K^{\omega-2}-I_1, \tag{17c}$$

$$I_4=\frac{i(-\pi)^{\omega}\Gamma(2-\omega)}{n\cdot n^*}\int\frac{dk\,k\cdot n^*}{(p-k)^2k\cdot n}\int_0^1 dy\,y^{\omega-2}K^{\omega-2}+\frac{i(-\pi)^{\omega}\Gamma(1-\omega)}{2}\int\frac{dk}{(p-k)^2(k\cdot n)^2}\int_0^1 dy\,y^{\omega-3}K^{\omega-1}$$

$$-\frac{i(-\pi)^{\omega}\Gamma(1-\omega)\Gamma(\omega-2)\Gamma(\omega)}{2\Gamma(2\omega-2)}\int\frac{dk(k^2)^{\omega-1}}{(p-k)^2(k\cdot n)^2}, \tag{17d}$$

where

$$K(k,y)\equiv(1-y)k^2+2y\,k\cdot n\,k\cdot n^*/n\cdot n^*.$$

Adding the integrals (17a)–(17d), we obtain

$$C(p)=i(-\pi)^{\omega}\Big\{\frac{\Gamma(2-\omega)}{n\cdot n^*}\int_0^1 dy\,y^{\omega-2}\int\frac{dk(p-k)\cdot n^*K^{\omega-2}}{(p-k)^2(p-k)\cdot n}+\frac{\Gamma(3-\omega)}{(\omega-2)}\int_0^1 dy\,y^{\omega-2}\int\frac{dk\,k\cdot(p-k)}{(p-k)^2k\cdot n\,(p-k)\cdot n}$$

$$\times\{K^{\omega-2}-[(1-y)k^2]^{\omega-2}\}+\frac{\Gamma(2-\omega)}{n\cdot n^*}\int_0^1 dy\,y^{\omega-2}\int\frac{dk\,k\cdot n^*K^{\omega-2}}{(p-k)^2k\cdot n}+\frac{\Gamma(1-\omega)}{2}$$

$$\times\int_0^1 dy\,y^{\omega-3}\int\frac{dk\,K^{\omega-1}}{(p-k)^2(k\cdot n)^2}-\frac{\Gamma(1-\omega)\Gamma(\omega-2)\Gamma(\omega)}{2\Gamma(2\omega-2)}\int\frac{dk(k^2)^{\omega-1}}{(p-k)^2(k\cdot n)^2}\Big\}. \tag{18}$$

The remaining k-integrals can all be evaluated by following the procedure outlined between Eqs. (9)–(11) and using the identity[7]

$$\frac{1}{k\cdot n\,(p-k)\cdot n}=\frac{1}{p\cdot n}\left[\frac{1}{(p-k)\cdot n}+\frac{1}{k\cdot n}\right]. \tag{19}$$

The most important of these are listed in Appendix C; other, related integrals can be found in Ref. 8. We shall refrain from showing here the various substitutions and reductions. Suffice it to say that $C(p)$ reduces eventually to the form

$$C(p)=-(-\pi)^{2\omega}(a_2/\epsilon^2+a_1/\epsilon+a_0), \tag{20}$$

where the coefficients $a_i=a_i(p,n,n^*)$, $i=0,1,2$, are finite. A somewhat less direct method of evaluating $C(p)$ is discussed in Appendix A. This completes our analysis of the overlapping integral $C(p)$.

IV. DISCUSSION

We have demonstrated the feasibility of performing higher-loop calculations in the noncovariant light-cone gauge $n_\mu A^a_\mu=0$, $n^2=0$, by evaluating in detail various Feynman integrals arising in a two-loop Yang–Mills self-energy $\Pi^{af}_{\mu\nu}$. We find that use of a consistent prescription for $(q\cdot n)^{-1}$, such as formula (2), enables us to compute all *momentum* integrals unambiguously. The computation of $\Pi^{af}_{\mu\nu}$ is aggravated by the appearance of the vector n^*_μ and by the presence, in one of the double integrals, of an overlapping divergence. The above results may not sound particularly impressive, but it is worth remembering that *none* of the multiloop techniques, developed during the past dozen years for covariant-gauge integrals, are applicable in the case of the light-cone gauge, the trickiest of all ghost-free gauges. This work is a very modest attempt at analyzing and systematizing the computation of multiloop Feynman integrals in noncovariant gauges.

ACKNOWLEDGMENT

We should like to thank the Natural Sciences and Engineering Research Council of Canada for support under Grant No. A8063.

APPENDIX A: ALTERNATIVE METHOD FOR $C(p)$

In view of the complexity of $C(p)$ in Eq. (6c), we discuss here a second method of evaluation. We begin by making a change of variables in (6c), so that $X(p)$ becomes

$$C(p)=\int\int \frac{dq\,dk\,q\cdot(p-k-q)}{q^2k^2(p-k-q)^2q\cdot n\,(p-k-q)\cdot n}$$
$$=\int\int \frac{dq\,dk\,q\cdot(p-k-q)}{q^2k^2(p-k-q)^2(p-k)\cdot n}$$
$$\times\left[\frac{1}{q\cdot n}+\frac{1}{(p-k-q)\cdot n}\right], \tag{A1}$$

or

$$C(p)\equiv 2C_1(p)-2C_2(p)\,, \tag{A2}$$

where

$$C_1(p)=\int\int \frac{dq\,dk\,k\cdot q}{q^2(k-q)^2(k-p)^2\,q\cdot n\,k\cdot n}\,, \tag{A3}$$

$$C_2(p)=\int\int \frac{dq\,dk}{(k-q)^2(k-p)^2q\cdot n\,k\cdot n}\,. \tag{A4}$$

Since the "four-propagator" integral $C_2(p)$ is easier to compute than $C_1(p)$, we shall discuss it first. By following the procedure for B_μ, Eq. (8), $C_2(p)$ can be derived in closed form. Integration over q_μ-space first, yields

$$C_2(p)=\frac{i(-\pi)^\omega\Gamma(2-\omega)}{\omega-1}\left(\frac{2}{n\cdot n^*}\right)^{\omega-1}$$
$$\times\int\frac{dk\,(k\cdot n^*)^{\omega-1}}{(k-p)^2(k\cdot n)^{3-\omega}}$$
$$=-(-\pi)^{2\omega}[\Gamma(\epsilon)/\epsilon]\,T\,, \tag{A5}$$

$$T=\frac{2[\Gamma(\omega-1)]^2}{\Gamma(2\omega-1)}\left(\frac{2p\cdot n^*}{n\cdot n^*}\right)^{2\omega-2}(p\cdot n)^{2\omega-4}\,. \tag{A6}$$

Thus $C_2(p)$ possesses a double pole.

It is worth noting, in this connection, that reversing the order of integration, i.e., integrating (A4) over k_μ *first*, seems to lead only to a simple pole for $C_2(p)$. The line of reasoning in this case would go something like this. Since the momentum integral

$$\int dk\,[(q-k)^2(k-p)^2k\cdot n]^{-1}\equiv F(q,p,n),$$

$q\neq 0, p\neq 0$, is *finite* [see Eq. (14b) of Ref. 4], the total integral

$$\int\frac{dq}{q\cdot n}F(q,p,n) \tag{A7}$$

should at most contain a simple pole. This argument is obviously too simplistic, as can be seen by studying the purely *covariant* integral

$$I=\int\int\frac{dq\,dk}{k^2(q-k)^2(k-p)^2(q-p)^2}\,. \tag{A8}$$

Integrating (A8) over q_μ first [i.e., proceeding as between Eqs. (A4) and (A6)], we get

$$I=i(-\pi)^\omega\Gamma(2-\omega)B(\omega-1,\omega-1)\int\frac{dk}{k^2[(k-p)^2]^{3-\omega}}$$
$$=-(-\pi)^{2\omega}\Gamma(2-\omega)(\Gamma(4-2\omega)$$
$$\times[\Gamma(3-\omega)]^{-1}B(\omega-1,\omega-1)$$
$$\times B(\omega-1,2\omega-3)(p^2)^{2\omega-4}\,. \tag{A9}$$

The double pole arising from $\Gamma(2-\omega)\Gamma(4-2\omega)$ is well documented in the literature.[1]

By contrast, suppose we start the evaluation of (A8) by doing the k-integration first:

$$I'=\int\frac{dq}{(q-p)^2}\int\frac{dk}{k^2(k-p)^2(k-q)^2}\,.$$

Since $\int dk[k^2(k-p)^2(k-q)^2]^{-1}$ is both ultraviolet and infrared finite, the double integral I' would appear to contain only a simple pole, thus contradicting expression (A9).

There remains the integral $C_1(p)$ in (A3), which may be analyzed by using either one of the following two procedures.

Procedure A: Using Eq. (12) in the text with $\sigma=1$, and simplifying we get

$$C_1(p)=\frac{i(-\pi)^\omega\Gamma(1-\omega)}{2}\int_0^1 dy\,y^{\omega-3}(1-y)^{\omega-2}(1-3y)\int\frac{dk\,(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-1}}{(p-k)^2(k\cdot n)^2}$$
$$+\frac{2i(-\pi)^\omega\Gamma(2-\omega)}{n\cdot n^*}\int_0^1 dy\,y^{\omega-2}(1-y)^{\omega-3}\int\frac{dk\,k\cdot n^*(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-2}}{(p-k)^2k\cdot n}$$
$$+\frac{i(-\pi)^\omega\Gamma(2-\omega)B(\omega-1,\omega-2)}{2}\int\frac{dk\,(k^2)^{\omega-1}}{(p-k)^2(k\cdot n)^2}\,, \tag{A10}$$

with $t=2y/((1-y)n\cdot n^*)$. The necessary integrals are given in Appendix C, specifically Eqs. (C1), (C2), and (C4).

Procedure B: Let $k\cdot q=\frac{1}{2}[k^2+q^2-(k-q)^2]$ and take $\int dq(q^2q\cdot n)^{-1}\equiv 0$, in which case (A3) reduces to

$$C_1(p)=\frac{1}{2}C_2(p)+\frac{1}{2}\int\int\frac{dk\,dq\,k^2}{q^2(k-p)^2(q-k)^2k\cdot n\,q\cdot n}\,. \tag{A11}$$

Hence $C(p)$ in (A2) becomes

$$C(p)=(-\pi)^{2\omega}[\Gamma(\epsilon)/\epsilon]\,T+C_3(p)\,, \tag{A12}$$

where

$$C_3(p)=\int\int\frac{dk\,dq\,k^2}{q^2(q-k)^2(k-p)^2k\cdot n\,q\cdot n}\,. \tag{A13}$$

Reduction of the double integral (A13) may be summarized as follows. Express $C_3(p)$ as

$$C_3(p)=\int\frac{dk\,k^2F(k)}{(k-p)^2k\cdot n}\,, \tag{A14}$$

where

$$F(k)\equiv\int dq[q^2(q-k)^2q\cdot n]^{-1}$$
$$=i(-\pi)^{\omega}\Gamma(3-\omega)k\cdot n^*\times\int_0^1 d\alpha\,d\beta\,\beta^{\omega-2}(1-\beta)^{\omega-3}A^{-2}\times(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-3}, \quad \text{(A15)}$$

$t\equiv(1-\alpha)\beta/(A(1-\beta))$, $A=\alpha+(1-\alpha)n_0^2$, $n_0^2=\frac{1}{2}n\cdot n^*$,

so that

$$C_3(p)=i(-\pi)^{\omega}B(J_1-tJ_2), \quad \text{(A16)}$$

with

$$B=\Gamma(3-\omega)\int_0^1 d\alpha\,d\beta\,\beta^{\omega-2}(1-\beta)^{\omega-3}A^{-2}, \quad \text{(A17)}$$

$$J_1(p)=\int dk\,k\cdot n^*[(k-p)^2k\cdot n(k^2+t\,k\cdot n\,k\cdot n^*)^{2-\omega}]^{-1}, \quad \text{(A18)}$$

$$J_2(p)=\int dk\,(k\cdot n^*)^2[(k-p)^2(k^2+t\,k\cdot n\,k\cdot n^*)^{3-\omega}]^{-1}. \quad \text{(A19)}$$

The integrals (A18) and (A19) are given, respectively, by Eqs. (C2) and (C3) in Appendix C.

The final expression for $C(p)$ reads

$$C(p)=(-\pi)^{2\omega}[\Gamma(\epsilon)/\epsilon]T+C_3(p), \quad \text{(A20)}$$

with T and $C_3(p)$ given by Eqs. (A6) and (A16), respectively. The result (A20) is to be compared with the form of $C(p)$ given in the main text, Eqs. (18) and (20).

APPENDIX B: SOME GAUSSIAN INTEGRALS

1. The integrals $\int_{-\infty}^{\infty} dq_4(q_4)^n e^{-M}$, $M\equiv a\,q_4^2-2bq_4p_4$

The subsequent integrals can be derived from the standard expression (11a) in the text by partial differentiation with respect to the parameter a or b:

$$\int_{-\infty}^{\infty} dq_4\,q_4\,e^{-M}=\frac{b\,p_4\sqrt{\pi}}{a^{3/2}}E_0, \quad E_0\equiv\exp\left(\frac{b^2p_4^2}{a}\right), \quad \text{(B1)}$$

$$\int_{-\infty}^{\infty} dq_4\,q_4^2e^{-M}=\sqrt{\pi}\left(\frac{1}{2a^{3/2}}+\frac{b^2p_4^2}{a^{5/2}}\right)E_0, \quad \text{(B2)}$$

$$\int_{-\infty}^{\infty} dq_4\,q_4^3e^{-M}=b\,p_4\sqrt{\pi}\left(\frac{3}{2a^{5/2}}+\frac{b^2p_4^2}{a^{7/2}}\right)E_0. \quad \text{(B3)}$$

2. The integrals $\int d^{2\omega-1}\mathbf{q}\,f(\mathbf{q})e^{-V}$, $V\equiv\gamma\mathbf{q}^2-2\beta\,\mathbf{q}\cdot\mathbf{p}+\alpha(\mathbf{n}\cdot\mathbf{q})^2$

The integrals (B5) to (B9) below can be derived from the basic formula (11b) by differentiating the latter partially, either with respect to α, β, or γ. We define

$$E\equiv\exp\left[\frac{\beta\mathbf{p}^2}{\gamma}-\frac{\alpha\beta^2(\mathbf{p}\cdot\mathbf{n})^2}{\gamma A}\right], \quad A\equiv\gamma+\alpha\mathbf{n}^2. \quad \text{(B4)}$$

$$\int d^{2\omega-1}\mathbf{q}\,\mathbf{q}e^{-V}=\frac{\pi^{\omega-1/2}\gamma^{-\omega}\beta}{A^{1/2}}\left(\mathbf{p}-\frac{\alpha\mathbf{n}\,\mathbf{p}\cdot\mathbf{n}}{A}\right)E, \quad \text{(B5)}$$

$$\int d^{2\omega-1}\mathbf{q}\,\mathbf{q}\cdot\mathbf{n}\,e^{-V}=\frac{\pi^{\omega-1/2}\,\mathbf{p}\cdot\mathbf{n}\,\beta\gamma^{1-\omega}E}{A^{3/2}}, \quad \text{(B6)}$$

$$\int d^{2\omega-1}\mathbf{q}\,\mathbf{q}^2e^{-V}=\frac{\pi^{\omega-1/2}\gamma^{-\omega}}{2A^{1/2}}\left\{2\omega-1-\frac{\alpha\mathbf{n}^2}{A}+\frac{2\beta^2}{\gamma}\left[\mathbf{p}^2-\frac{2\alpha(\mathbf{p}\cdot\mathbf{n})^2}{A}+\frac{\alpha^2\mathbf{n}^2(\mathbf{p}\cdot\mathbf{n})^2}{A^2}\right]\right\}E, \quad \text{(B7)}$$

$$\int d^{2\omega-1}\mathbf{q}\,\mathbf{q}(\mathbf{q}\cdot\mathbf{n})e^{-V}=\frac{\pi^{\omega-1/2}\gamma^{1-\omega}}{2A^{3/2}}\left[\mathbf{n}+\frac{2\beta^2\,\mathbf{p}\cdot\mathbf{n}}{\gamma}\left(\mathbf{p}-\frac{\alpha\mathbf{n}\,\mathbf{p}\cdot\mathbf{n}}{A}\right)\right]E, \quad \text{(B8)}$$

$$\int d^{2\omega-1}\mathbf{q}\,\mathbf{q}\cdot\mathbf{n}\,\mathbf{q}^2e^{-V}=\frac{\pi^{\omega-1/2}\gamma^{-\omega}\beta\,\mathbf{p}\cdot\mathbf{n}}{A^{3/2}}\left\{\omega+\frac{1}{2}-\frac{3\alpha\mathbf{n}^2}{2A}+\frac{\beta}{\gamma}\left[\mathbf{p}^2-\frac{2\alpha(\mathbf{p}\cdot\mathbf{n})^2}{A}+\frac{\alpha^2\mathbf{n}^2(\mathbf{p}\cdot\mathbf{n})^2}{A}\right]\right\}E. \quad \text{(B9)}$$

APPENDIX C: SPECIAL INTEGRALS IN THE LIGHT-CONE GAUGE

The following momentum integrals, obtained by using the procedure discussed in the text between Eqs. (9) and (11), are required in the computation of the coefficient $C(p)$, Eq. (18). We find, in Minkowski space,

$$\int\frac{dk(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-1}}{(p-k)^2(k\cdot n)^2}=\frac{4i(-\pi)^{\omega}\Gamma(4-2\omega)}{\Gamma(1-\omega)(n\cdot n^*)^2}(p\cdot n^*)^2\times\int_0^1 du\,dv(1-u)v^{-\omega}(1-v)^{2\omega-2}(1+uvtn_0^2)^{-3}\times\left\{vp^2+\frac{2(1-v)(1-u+uvtn_0^2)p\cdot n\,p\cdot n^*}{(1+uvtn_0^2)n\cdot n^*}\right\}^{2\omega-4}, \quad \text{(C1)}$$

$$\int\frac{dk\,k\cdot n^*(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-2}}{(p-k)^2k\cdot n}=\frac{2i(-\pi)^{\omega}\Gamma(4-2\omega)(p\cdot n^*)^2}{\Gamma(2-\omega)n\cdot n^*}\times\int_0^1 du\,dv\,uv^{1-\omega}(1-v)^{2\omega-2}(1+uvtn_0^2)^{-3}\times\left\{vp^2+\frac{2(1-v)(1-u+uvtn_0^2)p\cdot n\,p\cdot n^*}{(1+uvtn_0^2)n\cdot n^*}\right\}^{2\omega-4}, \quad \text{(C2)}$$

$$\int\frac{dk(k\cdot n^*)^2(k^2+t\,k\cdot n\,k\cdot n^*)^{\omega-3}}{(p-k)^2}$$

$$= \frac{i(-\pi)^{\omega}\Gamma(4-2\omega)(p\cdot n^*)^2}{\Gamma(3-\omega)} \times \int_0^1 dv\, v^{\omega-2}(1-v)^{2\omega-2}(1+vtn_0^2)^{-3} \times \left\{p^2 + \frac{(1-v)t\, p\cdot n\, p\cdot n^*}{1+vtn_0^2}\right\}^{2\omega-4}, \tag{C3}$$

and

$$\int \frac{dk(k^2)^{\omega-1}}{(p-k)^2(k\cdot n)^2} = \frac{4i(-\pi)^{\omega}\Gamma(4-2\omega)(p\cdot n^*)^2}{\Gamma(1-\omega)(n\cdot n^*)^2} \times \int_0^1 du\, dv(1-u)v^{-\omega}(1-v)^{2\omega-2} \times \left\{vp^2 + \frac{2(1-u)(1-v)p\cdot n\, p\cdot n^*}{n\cdot n^*}\right\}^{2\omega-4}, \tag{C4}$$

where t is an arbitrary parameter. Note that the last integral may be obtained from (C1) by setting $t = 0$.

[1] J. L. Rosner, Ann. Phys. NY **44**, 11 (1967); G. 't Hooft and M. Veltman, *Diagrammar*, CERN preprint 73-9, 1973; M. J. Levine and R. Roskies, Phys. Rev. D **9**, 421 (1974); E. Mendels, Nuovo Cimento A **45**, 87 (1978); V. K. Cung, A. Devoto, T. Fulton, and W. W. Repko, Phys. Rev. D **18**, 3893 (1978); A. A. Vladimirov, Teor. Mat. Fiz. **43**, 210 (1980) [Theor. Math. Phys. USSR **43**, 417 (1980)]; A. E. Terrano, Phys. Lett. B **93**, 424 (1980); K. G. Chetyrkin, A. L. Kataev, and F. V. Tkachov, Nucl. Phys. B **174**, 345 (1980); Phys. Lett. B **99**, 147 (1981); B **101**, 457E (1981); INR (Moscow) preprint P-0200, 1981; F. V. Tkachov, Phys. Lett. B **100**, 65 (1981); D. R. T. Jones and J. P. Leveille, Nucl. Phys. B **206**, 473 (1982); D. I. Kazakov, Phys. Lett. B **133**, 406 (1983); F. R. Graziani, SLAC preprint SLAC-PUB-3369, 1984; N. Marcus and A. Sagnotti, Caltech preprint CALT-68-1128 and 1129, 1984; M. T. Grisaru and D. Zanon, Brandeis preprint BRX-TH-163 and 169, 1984.

[2] R. Delbourgo, A. Salam, and J. Strathdee, Nuovo Cimento A **23**, 237 (1974); W. Kummer, Acta Phys. Austriaca **41**, 315 (1975); W. Konetschny and W. Kummer, Nucl. Phys. B **100**, 106 (1975); B **108**, 397 (1976); B **124**, 145 (1977); W. Konetschny, Phys. Rev. D **28**, 354 (1983); A. Burnel, Nucl. Phys. B **198**, 531 (1982); D. M. Capper and G. Leibbrandt, Phys. Rev. D **25**, 1002 (1982); A. Bassetto, I. Lazzizzera, and R. Soldati, Nucl. Phys. B **236**, 319 (1984); A. Andraši and J. C. Taylor, Nucl. Phys. B **192**, 283 (1981); A. I. Mil'shtein and V. S. Fadin, Yad. Fiz. **34**, 1403 (1981) [Sov. J. Nucl. Phys. **34**, 779 (1981)].

[3] S. Mandelstam, Nucl. Phys. B **213**, 149 (1983); D. M. Capper, J. J. Dulwich, and M. J. Litvak, Nucl. Phys. B **241**, 463 (1984); Mats Ögren, University of Göteborg preprint 84-15, 1984 (unpublished); D. M. Capper and D. R. T. Jones, Nucl. Phys. B **252**, 718 (1985).

[4] G. Leibbrandt, Phys. Rev. D **29**, 1699 (1984).

[5] A. Bassetto, University of Trento preprint UTF-109, 1984; A. Bassetto, M. Dalbosco, I. Lazzizzera, and R. Soldati, Phys. Rev. D **31**,2012 (1985); H. C. Lee and M. S. Milgram, Phys. Lett. B **133**, 320 (1983); Chalk River preprints CRNL-TP-83-XII-13 and 14; CRNL-TP-84-X-24; CRNL-TP-85-II-11; A. Smith, Lawrence Berkeley Laboratory preprint LBL-18668, 1984.

[6] Although the x-integrals in Eq. (12) may be done explicitly, we prefer to keep them in the compact form shown.

[7] D. M. Capper and G. Leibbrandt, Phys. Rev. D **25**, 1009 (1982).

[8] G. Leibbrandt and S. -L. Nyeo, Phys. Lett. **140**, 417 (1984); G. Leibbrandt, Phys. Rev. D **30**, 2167 (1984).

Determination of coexistence wave functions from two-neutron-transfer data

M. Carchidi and H. T. Fortune
Physics Department, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 26 March 1985; accepted for publication 20 September 1985)

An analysis of wave functions in a generalized two-state model for a series of even–even nuclei is presented here. With minimal assumptions, necessary and sufficient conditions for consistency in obtaining these from data on (p,t) and (t,p) cross-section ratios are exhibited.

I. INTRODUCTION

If Ψ_j represents the jth physical state of a system and $\{\phi_1, \phi_2, \phi_3, ...\phi_n, ...\}$ a set of basis states, then we may write

$$\Psi_j = \sum_{i=1}^{\infty} c_{ji}\phi_i$$

with the usual orthonormality conditions, i.e., $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ or

$$\sum_{i=1}^{\infty} c_{ji}c_{ik}^{*} = \delta_{jk}.$$

In general, the set of eigenstates is infinite in number. An example of such a system would be the 0^+ levels in an even–even nucleus. Each physical level can be represented as a linear combination of basis levels. In general, all 0^+ levels can interact with each other requiring an infinite number of basis levels needed to describe the physics. If, however, a set n of these states, $\{\phi_1,\phi_2,\phi_3,...,\phi_n\}$ are related strongly to each other and only weakly with any state outside of the set, then this particular set of physical states may be described "almost" completely by a finite number of (suitably chosen) basis states $\{\phi_1,\phi_2,\phi_3,...,\phi_n\}$. This circumstance is referred to as an n-state model configuration of the physical states. An example of such a situation is again found in the 0^+ levels of an even–even nucleus, $0_1^+, 0_2^+, 0_3^+, ..., 0_n^+$.

For the purpose of this paper, we want to focus on a two-state model. We assume in an even–even nucleus that the lowest 0^+ basis state, labeled g, mixes with only one other 0^+ level, labeled e, which usually is, but need not be, the first excited 0^+ above g. If we let Ψ^A (g.s.) and Ψ^A $(0^{+\prime})$ represent the physical ground state and excited 0^+ state in nucleus A and ϕ_g^A and ϕ_e^A be the basis levels, then because of the orthonormality in ϕ_g^A and ϕ_e^A for each A, we can write for Ψ^A (g.s.) and Ψ^A $(0^{+\prime})$ the most general two-state model wave function as

$$\begin{aligned}\Psi^A(\text{g.s.}) &= \alpha_A\phi_g^A + \beta_A\phi_e^A,\\ \Psi^A(0^{+\prime}) &= \beta_A\phi_g^A - \alpha_A\phi_e^A,\end{aligned} \tag{1}$$

with $\alpha_A^2 + \beta_A^2 = 1$.

In a series of even–even nuclei, it may be that the properties of the basis states ϕ_g^A and ϕ_e^A change slowly and smoothly with A whereas the physical states exhibit more complicated behavior. If this hypothesis is to be tested, then the test must involve a process that can connect a 0^+ in A to a 0^+ level in A' in a simple, direct way so as to involve only Ψ^A (g.s.), Ψ^A $(0^{+\prime})$, $\Psi^{A'}$ (g.s.), and $\Psi^{A'}$ $(0^{+\prime})$. The (p,t) and (t,p) reactions are ideal for this purpose. We therefore wish to investigate the degree to which (t,p) and (p,t) cross-section data on a series of nearby nuclei can be described in such a model, with as few assumptions as possible.

Throughout the analysis, we assume that all kinematic effects, etc., have been removed, e.g., by dividing experimental cross sections by distorted-wave Born-approximation (DWBA) cross sections, calculated for the appropriate kinematics and standard form factor. Then, for the (p,t) and (t,p) cross sections, we write

$$\begin{aligned}\sigma(^{A+2}X(p,t)\,^AX(\text{g.s.})) &= \langle\Psi^A(\text{g.s.})|(p,t)|\Psi^{A+2}(\text{g.s.})\rangle^2,\\ \sigma(^{A+2}X(p,t)\,^AX(0^{+\prime})) &= \langle\Psi^A(0^{+\prime})|(p,t)|\Psi^{A+2}(\text{g.s.})\rangle^2,\\ \sigma(^AX(t,p)\,^{A+2}X(\text{g.s.})) &= \langle\Psi^{A+2}(\text{g.s.})|(t,p)|\Psi^A(\text{g.s.})\rangle^2,\\ \sigma(^AX(t,p)\,^{A+2}X(0^{+\prime})) &= \langle\Psi^{A+2}(0^{+\prime})|(t,p)|\Psi^A(\text{g.s.})\rangle^2.\end{aligned}$$

Putting in Eq. (1), we get for each cross section

$$\begin{aligned}\sigma(^{A+2}X(p,t)\,^AX(\text{g.s.})) &= f_A^2(\alpha_A\alpha_{A+2} + \alpha_A\beta_{A+2}s_A\\ &\quad + \alpha_{A+2}\beta_A r_A + \beta_A\beta_{A+2}R_A)^2,\\ \sigma(^{A+2}X(p,t)\,^AX(0^{+\prime})) &= f_A^2(\beta_A\alpha_{A+2} + \beta_A\beta_{A+2}s_A\\ &\quad - \alpha_{A+2}\alpha_A r_A - \alpha_A\beta_{A+2}R_A)^2,\\ \sigma(^AX(t,p)\,^{A+2}X(\text{g.s.})) &= f_A^2(\alpha_A\alpha_{A+2} + \alpha_A\beta_{A+2}s_A\\ &\quad + \alpha_{A+2}\beta_A r_A + \beta_A\beta_{A+2}R_A)^2,\\ \sigma(^AX(t,p)\,^{A+2}X(0^{+\prime})) &= f_A^2(\alpha_A\beta_{A+2} + \beta_A\beta_{A+2}r_A\\ &\quad - \alpha_{A+2}\alpha_A s_A - \beta_A\alpha_{A+2}R_A)^2,\end{aligned} \tag{2}$$

where f_A^2, r_A, s_A, and R_A are measures of the *basis* states $2n$-transfer overlaps and are given schematically in Fig. 1. To eliminate f_A^2 from the analysis, we consider the cross-section ratios

$$\frac{\sigma(^{A+2}X(p,t)\,^AX(0^{+\prime}))}{\sigma(^{A+2}X(p,t)\,^AX(\text{g.s.}))} = \left[\frac{x_{A+2}s_A - x_Ax_{A+2}r_A - x_AR_A}{x_Ax_{A+2} + x_{A+2}r_A + x_As_A + R_A}\right]^2, \tag{3}$$

$$\frac{\sigma(^AX(t,p)\,^{A+2}X(0^{+\prime}))}{\sigma(^AX(t,p)\,^{A+2}X(\text{g.s.}))} = \left[\frac{x_A + r_A - x_Ax_{A+2}s_A - x_{A+2}R_A}{x_Ax_{A+2} + x_{A+2}r_A + x_As_A + R_A}\right]^2, \tag{4}$$

where

 0022-2488/86/020633-12$02.50 © 1986 American Institute of Physics

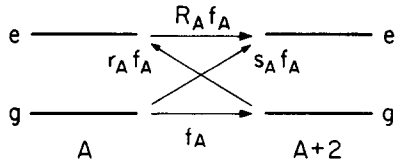


FIG. 1. Basis state definitions of r_A, s_A, and R_A.

$$x_A = \alpha_A/\beta_A, \tag{5}$$

and hence

$$\alpha_A^2 = x_A^2/(1+x_A^2) = 1-\beta_A^2. \tag{6}$$

We shall denote these ratios as P_A^2 and T_A^2, respectively, and call these the *calculated* values of the (p,t) and (t,p) cross-section ratios. The *experimental* ratios and their *uncertainties* are denoted by $P_{A_0}^2$, $T_{A_0}^2$, $\Delta P_{A_0}^2$, and $\Delta T_{A_0}^2$. The goal of this and any empirical model is to minimize the expression

$$\chi^2 = \left(\frac{1}{n}\right)\sum_{A=A'}^{A''}\left\{\left[\frac{P_A^2 - P_{A_0}^2}{\Delta P_{A_0}^2}\right]^2 + \left[\frac{T_A^2 - T_{A_0}^2}{\Delta T_{A_0}^2}\right]^2\right\}$$

with as few parameters as necessary. Here, n is the number of data points fitted minus the number of degrees of freedom.

If we allow r_A, s_A, and R_A to vary arbitrarily with A then we have more parameters than data and everything can be fitted exactly. The problem is overdetermined and information deduced from such a model tells us almost no physics. We therefore make the first assumption about the basis states ϕ_g^A and ϕ_e^A.

Assumption 1: We assume the basis state $2n$-transfer overlap *ratios* are independent of A, i.e., $r_A = r$, $s_A = s$, and $R_A = R$ are all constants with A. This assumption may be too restrictive in some cases, e.g., in nuclei in which a single orbital is dominant. However, we would normally treat those cases in a totally different manner.

With this assumption P_A^2 and T_A^2 become

$$P_A^2 = \left[\frac{x_{A+2} + s - x_A x_{A+2} r - x_A R}{x_A x_{A+2} + x_{A+2} r + x_A s + R}\right]^2, \tag{7a}$$

$$T_A^2 = \left[\frac{x_A + r - x_A x_{A+2} s - x_{A+2} R}{x_A x_{A+2} + x_{A+2} r + x_A s + R}\right]^2. \tag{7b}$$

At this stage, we have two major concerns: (1) What are the calculated values of T_A^2 and P_A^2 that give the best fit to the data $T_{A_0}^2$ and $P_{A_0}^2$ with their uncertainties $\Delta T_{A_0}^2$ and $\Delta P_{A_0}^2$? (2) What wave-function probabilities α_A^2 and β_A^2 would produce this fit via Eq. (7)? We shall answer question (2) first and in the process we shall also develop a procedure for answering question (1).

To deduce α_A^2 and β_A^2 from T_A^2 and P_A^2, we need merely invert Eq. (7) for each value of A. Such a process is simplified by considering T_A and P_A instead of their squares. This would introduce sign ambiguities but these will turn out to be not much of a problem and will be considered later. The equations to solve are then

$$P_A = \frac{x_{A+2} + s - x_A x_{A+2} r - x_A R}{x_A x_{A+2} + x_{A+2} r + x_A s + R}, \tag{8}$$

$$T_A = \frac{x_A + r - x_A x_{A+2} s - x_{A+2} R}{x_A x_{A+2} + x_{A+2} r + x_A s + R}. \tag{9}$$

Solving these equations for x_{A+2} yields

$$x_{A+2} = \frac{(r - T_A R) + (1 - T_A s)x_A}{(T_A r + R) + (s + T_A)x_A}, \tag{10}$$

and

$$x_{A+2} = \frac{(s - P_A R) - (R + P_A s)x_A}{(P_A r - 1) + (r + P_A)x_A}. \tag{11}$$

Equating these yields the following quadratic equation for x_A:

$$\begin{aligned}&[P_A(1+s^2) + T_A(R - rs) + (r + Rs)]x_A^2 \\ &\quad + [(R^2 + r^2) + 2P_A(r + sR) - (1 + s^2)]x_A \\ &\quad + [P_A(R^2 + r^2) + T_A(R - rs) - (r + sR)] = 0.\end{aligned} \tag{12}$$

If we then take Eqs. (8) and (9) for T_{A-2} and P_{A-2} and solve each for x_{A-2}, then we have

$$x_{A-2} = \frac{(r - T_{A-2}R) - (R + rT_{A-2})x_A}{(sT_{A-2} - 1) + (s + T_{A-2})x_A} \tag{13}$$

and

$$x_{A-2} = \frac{(s - P_{A-2}R) + (1 - rP_{A-2})x_A}{(sP_{A-2} + R) + (r + P_{A-2})x_A}. \tag{14}$$

Equating these gives another quadratic equation for x_A,

$$\begin{aligned}&[T_{A-2}(1+r^2) + P_{A-2}(R - rs) + (s + rR)]x_A^2 \\ &\quad + [(R^2 + s^2) + 2T_{A-2}(s + rR) - (1 + r^2)]x_A \\ &\quad + [T_{A-2}(R^2 + s^2) + P_{A-2}(R - rs) - (s + rR)] = 0.\end{aligned} \tag{15}$$

Solving Eqs. (12) and (15) simultaneously for x_A and x_A^2 and setting $(x_A)^2$ equal to x_A^2 leads, after much algebra, to an expression involving only r, s, and R, along with the T_A's and P_A's given as follows:

$$\begin{aligned}&[(R-1)^2 + (s+r)^2]\{(R - rs)(R+1)^2[\beta_{1A}(R - rs) + \beta_{2A}((R+1)^2 + (s-r)^2)] \\ &\quad + (s-r)(R+1)[\beta_{3A}(R - rs)^2 + \beta_{4A}(R - rs)(R^2 + r^2 + s^2 + 1) + \beta_{5A}(R+1)^2((R-1)^2 + (s+r)^2)] \\ &\quad + (s-r)^2[\beta_{6A}(R - rs)^2 + \beta_{7A}(R - rs)(R^2 + r^2 + s^2 + 1) + \beta_{8A}(R+1)^2((R-1)^2 + (s+r)^2)] \\ &\quad + (s-r)^3(R+1)\beta_{9A}[(R-1)^2 + (s+r)^2] + (s-r)^4\beta_{10A}[(R-1)^2 + (s+r)^2]\} = 0,\end{aligned} \tag{16}$$

where

$$\begin{aligned}\beta_{1A} &= (T_{A-2}T_A - P_{A-2}P_A)^2 + (T_A \\ &\quad + T_{A-2} - P_A - P_{A-2})^2,\end{aligned} \tag{17a}$$

$$\beta_{2A} = (T_{A-2} - P_A)(P_{A-2} - T_A), \tag{17b}$$

$$\beta_{3A} = 4(T_{A-2} - P_A)(T_{A-2}P_A + T_A P_{A-2}), \tag{17c}$$

$$\beta_{4A} = 2(T_{A-2} - P_A)(T_{A-2}T_A + P_A P_{A-2}), \tag{17d}$$

$$\beta_{5A} = -(T_{A-2} - P_A), \tag{17e}$$

$$\beta_{6A} = (T_{A-2}T_A + P_A P_{A-2})^2 + (P_{A-2} + T_A)^2 + (T_{A-2} + P_A)^2 + 4T_{A-2}^2 P_A^2, \tag{17f}$$

$$\beta_{7A} = (T_{A-2} + P_A)(P_{A-2} + T_A) + 2T_{A-2}P_A(T_{A-2}T_A + P_{A-2}P_A), \tag{17g}$$

$$\beta_{8A} = (T_{A-2} - P_A)^2 - (1 + T_{A-2}P_A), \tag{17h}$$

$$\beta_{9A} = (T_{A-2} - P_A)(1 + 2T_{A-2}P_A), \tag{17i}$$

$$\beta_{10A} = T_{A-2}P_A(1 + T_{A-2}P_A). \tag{17j}$$

An immediate solution to Eq. (16) is for $(R-1)^2 + (s+r)^2 = 0$, i.e., $R = 1$ and $s = -r$. When these are put back into Eqs. (7) we get $P_A^2 = T_A^2$, the purely symmetric situation. As such a situation is uncommon, we shall assume that $(R-1)^2 + (s+r)^2 \neq 0$ and proceed. Equation (16) divided by this quantity holds for all values of A for which data exist. We shall refer to this equation as the constraint equation and each such equation represents a surface in the r, s, R space and these equations represent the intersection of a set of such surfaces. To simplify matters considerably, we now make our second and last assumption about the basis states.

Assumption 2: We assume symmetry in the basis state $2n$-transfer overlap ratios, i.e., $r = s$. (The general R, r, s problem with $r \neq s$ is left for later development.) Of course, the special case $s = r$ can always be transformed, by a change of basis, into the case $r = s = 0$. However, we keep the general form with $r \neq 0$ because of its similarity to the $r \neq s$ situation and because the "best" choice of basis states will probably have $r \neq 0$.

With assumption 2, the expressions for P_A and T_A become

$$P_A = \frac{x_{A+2} + (1 - x_A x_{A+2})r - x_A R}{x_A x_{A+2} + (x_{A+2} + x_A)r + R}, \tag{18}$$

$$T_A = \frac{x_A + (1 - x_A x_{A+2})r - x_{A+2} R}{x_A x_{A+2} + (x_{A+2} + x_A)r + R}, \tag{19}$$

and the constraint equation [Eq. (16)] becomes

$$(R - r^2)(R+1)[\beta_{1A}(R - r^2) + \beta_{2A}(R+1)^2] = 0. \tag{20}$$

Equations (12) and (15) become

$$[P_A(1 + r^2) + T_A(R - r^2) + r(R+1)]x_A^2 + [R + 2P_A r - 1](R+1)x_A + [P_A(R^2 + r^2) + T_A(R - r^2) - r(R+1)] = 0 \tag{21}$$

and

$$[T_{A-2}(1 + r^2) + P_{A-2}(R - r^2) + r(R+1)]x_A^2 + [R + 2T_{A-2}r - 1](R+1)x_A + [T_{A-2}(R^2 + r^2) + P_{A-2}(R - r^2) - r(R+1)] = 0. \tag{22}$$

If $R + 1 = 0$, these equations give $(P_A - T_A)(1 + r^2)(1 + x_A^2) = 0$, which is again the purely symmetric situation $P_A^2 = T_A^2$. If $R - r^2 = 0$, these equations give $(1 + r^2)(P_A - T_{A-2})(x_A + r)^2 = 0$. Since $x_A + r = 0$ (with $R = r^2$) would yield $\sigma(^A X(t,p)\ ^{A+2}X(\text{g.s.})) = 0$, we are led to $P_A = T_{A-2}$, again a very symmetric situation. Both of these special cases are actually covered by the single equation $\beta_{1A}(R - r^2) + \beta_{2A}(R+1)^2 = 0$. Therefore, we can reduce Eq. (20) to the *constraint equation*,

$$r^2 = R + K_A(R+1)^2, \tag{23}$$

where

$$K_A = (T_{A-2} - P_A)(P_{A-2} - T_A)/[(T_{A-2}T_A - P_A P_{A-2})^2 + (T_{A-2} + T_A - P_A - P_{A-2})^2]. \tag{24}$$

Requiring r and R to be constants automatically requires K_A to be independent of A. We can thus test assumptions 1 and 2 for the basis states $2n$-transfer overlap ratios. When written as

$$K_A = (r^2 - R)/(R+1)^2, \tag{25}$$

the equation relates the basis states $2n$-transfer overlap ratios r and R to the physical state $2n$-transfer overlap ratios, T_A, P_A, T_{A-2}, and P_{A-2} and can serve as a starting point for constructing the basis states ϕ_g^A and ϕ_e^A because it relates their overlaps to experimentally determined quantities in K_A. Since it is wave-function overlaps that are directly measurable, we can continue with the analysis even though we do not have explicit expressions for ϕ_g^A and ϕ_e^A. Equations (23) and (24) represent necessary conditions for the existence of solutions x_A to Eqs. (18) and (19) for all A. The A independence of K_A, however, is not a sufficient condition for solutions x_A to exist. To determine a sufficient condition on the T_A and P_A's, it will become useful to make the following definitions:

$$z_A = (T_A - P_{A-2})/(T_{A-2} - P_A), \tag{26}$$

$$w_A = K_A(1 + z_A), \tag{27}$$

$$v_A = K_A(T_A + P_A z_A). \tag{28}$$

Then,

$$K_A = -z_A/[(T_A + P_A z_A)^2 + (1 + z_A)^2]. \tag{29}$$

These parameters are all related by the identities

$$K_A = v_A^2 + w_A^2 + w_A, \tag{30}$$

$$K_A z_A + v_A^2 + w_A^2 = 0. \tag{31}$$

The mathematical and physical importance of each parameter will be discussed later.

To continue with the derivation of x_A, we take Eqs. (21) and (22) along with Eq. (23) to deduce the simple results

$$x_A = \frac{R + w_A(R+1)}{-r + v_A(R+1)} = \frac{-r - v_A(R+1)}{1 + w_A(R+1)}, \tag{32}$$

leading to

$$\alpha_A^2 = \frac{-(1 + 2w_A) + 2rv_A + (R+1)(1 + w_A + 2K_A)}{(R+1)(1 + 4K_A)}. \tag{33}$$

For a given R value we can calculate r via Eq. (23) and then α_A^2 (and hence $\beta_A^2 = 1 - \alpha_A^2$) via Eq. (33). These equations represent all necessary conditions for consistency in Eqs. (18) and (19). They are all based on the premise that K_A is independent of A. We still must obtain a sufficient condition

for Eq. (32) or (33) to represent solutions to the T_A and P_A equations. To find this condition we go back to Eqs. (18) and (19) and rewrite them as

$$\frac{x_A + r}{rx_A + R} = \frac{x_{A-2} + P_{A-2}}{1 - x_{A-2}P_{A-2}} \tag{34}$$

and

$$\frac{x_A + r}{rx_A + R} = \frac{x_{A+2} + T_A}{1 - x_{A+2}T_A}. \tag{35}$$

Equating these gives

$$x_{A-4} = (\delta_{A-2}x_A + 1)/(\delta_{A-2} - x_A) \tag{36}$$

or

$$x_{A+4} = (\delta_{A+2}x_A - 1)/(\delta_{A+2} + x_A), \tag{37}$$

where

$$\delta_A = (1 + T_A P_{A-2})/(T_A - P_{A-2}). \tag{38}$$

Equations (36)–(38) are really special cases of a more general result. Using Eqs. (30), (34) and the fact that $K_A = K_{A'}$ for any A and A', we arrive at the more general result

$$\epsilon_{AA'} = \frac{x_A - x_{A'}}{1 + x_A x_{A'}} = \frac{w_A - w_{A'}}{v_A + v_{A'}} = \frac{v_A - v_{A'}}{1 + w_A + w_{A'}}, \tag{39}$$

an expression that can be evaluated from the experimental quantities without knowing the value of R. If we incorporate this result with the previous expressions for x_A, then after much algebra, we arrive at the condition

$$(w_A + T_A v_A) + z_A(w_{A+2} + T_A v_{A+2} + 1) = 0. \tag{40}$$

Solving this equation for T_{A+2} and using the fact that Eq. (29) can be written as

$$\frac{1 + 2K_A(1 + P_A T_A)}{z_A} = \frac{-K_A[z_A^2(1 + P_A^2) + (1 + T_A^2)]}{z_A^2},$$

Eq. (40) reduces to

$$L_A = 1, \tag{41}$$

where

$$L_A = z_{A+2}(1 + T_A^2)/z_A(1 + P_A^2). \tag{42}$$

Having K_A independent of A does not imply $L_A = 1$, but $L_A = 1$ does yield K_A independent of A. For proof, we write

$$K_A = \frac{-z_A}{(T_A + P_A z_A)^2 + (1 + z_A)^2} = \frac{-z_A}{(1 + T_A^2) + z_A^2(1 + P_A^2) + 2z_A(1 + P_A T_A)}.$$

Putting in Eqs. (41) and (42), which give

$$z_A = z_{A+2}(1 + T_A^2)/(1 + P_A^2),$$

we get

$$K_A = \frac{-z_{A+2}}{\{(1 + T_A^2) + z_A^2(1 + P_A^2) + 2z_A(1 + P_A T_A)\}[(1 + P_A^2)/(1 + T_A^2)]}$$

$$= \frac{-z_{A+2}}{\{(1 + T_A^2) + z_A^2[(1 + P_A^2)^2/(1 + T_A^2)] + 2z_{A+2}(1 + P_A T_A)\}}$$

$$= \frac{-z_{A+2}}{\{(1 + P_A^2) + z_{A+2}^2(1 + T_A^2) + 2z_{A+2}(1 + P_A T_A)\}} = \frac{-z_{A+2}}{(P_A + T_A z_{A+2})^2 + (1 + z_{A+2})^2}.$$

But by definition of $z_A, z_{A+2} = (T_{A+2} - P_A)/(T_A - P_{A+2})$, so

$$P_A + T_A z_{A+2} = T_{A+2} + P_{A+2}z_{A+2},$$

hence

$$K_A = \frac{-z_{A+2}}{(T_{A+2} + P_{A+2}z_{A+2})^2 + (1 + z_{A+2})^2} = K_{A+2}.$$

Therefore we can replace the necessary condition K_A independent of A by the condition $L_A = 1$. The advantage here is that $L_A = 1$ is also a *sufficient* condition for solutions to Eqs. (18) and (19) to exist. As proof, it is best to consider the summed (t,p) and (p,t) strengths

$$S_A(t,p) = \sigma({}^A X(t,p)\,{}^{A+2}X(\text{g.s.})) + \sigma({}^A X(t,p)\,{}^{A+2}X(0^{+\prime})), \tag{43}$$

$$S_A(p,t) = \sigma({}^{A+2}X(p,t)\,{}^A X(\text{g.s.})) + \sigma({}^{A+2}X(p,t)\,{}^A X(0^{+\prime})),$$

which becomes

$$S_A(t,p) = f_A^2[r^2 + 1 + (R^2 - 1)\beta_A^2 + 2\alpha_A\beta_A r(R + 1)]. \tag{44}$$

Using the fact that $\alpha_A\beta_A = x_A/(1 + x_A^2)$ and Eqs. (32) and (33), we get

$$\alpha_A\beta_A = \frac{v_A(R + 1) - r(1 + 2w_A) - 2v_A}{(R + 1)(1 + 4K_A)}. \tag{45}$$

Finally, if we put this and Eq. (23) into Eq. (44) we get the surprisingly simple result

$$S_A(t,p) = -z_A K_A f_A^2(R + 1)^2. \tag{46}$$

By definition of T_A^2, we can also write

$$S_A(t,p) = [\sigma({}^A X(t,p)\,{}^{A+2}X(\text{g.s.}))](1 + T_A^2) = [\sigma({}^A X(t,p)\,{}^{A+2}X(0^{+\prime}))](1 + T_A^{-2})$$

leading to both

$$\sigma({}^A X(t,p)\,{}^{A+2}X(\text{g.s.})) = -z_A K_A f_A^2(R + 1)^2/(1 + T_A^2) \tag{47}$$

and

$$\sigma({}^A X(t,p)\,{}^{A+2}X(0^{+\prime})) = -z_A K_A T_A^2 f_A^2(R + 1)^2/(1 + T_A^2), \tag{48}$$

showing that the solutions deduced from Eqs. (23), (32), and (33) will give

$$\sigma({}^A X(t,p)\,{}^{A+2}X(0^{+\prime}))/\sigma({}^A X(t,p)\,{}^{A+2}X(\text{g.s.})) = T_A^2,$$

for any value of R, that is, the solutions will work. If the same calculation is done for the (p,t) summed strengths, we get

$$S_A(p,t) = -z_{A+2}K_{A+2}f_A^2(R+1)^2\,, \tag{49}$$

so

$$\sigma(^{A+2}X(p,t)\,^AX(\text{g.s.})) = \frac{-z_{A+2}K_{A+2}f_A^2(R+1)^2}{(1+P_A^2)}\,, \tag{50}$$

and

$$\sigma(^{A+2}X(p,t)\,^AX(0^{+\prime})) = \frac{-z_{A+2}K_{A+2}P_A^2f_A^2(R+1)^2}{(1+P_A^2)}\,, \tag{51}$$

yielding

$$\sigma(^{A+2}X(p,t)\,^AX(0^{+\prime}))/\sigma(^{A+2}X(p,t)\,^AX(\text{g.s.})) = P_A^2\,,$$

again independent of R. The only condition which needs to be satisfied is the result $\sigma(^AX(t,p)^{A+2}X(\text{g.s.})) = \sigma(^{A+2}X(p,t)^AX(\text{g.s.}))$, which is true as long as $L_A = 1$. Remember we are using kinematically reduced cross sections, always. So $L_A = 1$ is both *necessary* and *sufficient* to guarantee that the wave functions of Eq. (33) with Eq. (23) do reproduce the T_A and P_A ratios used to deduce them for any value of R! We have therefore an infinite number of solutions to Eqs. (18) and (19) all described in terms of the one parameter R.

The beauty of the preceding discussion is that it not only answers question (2) from the introduction, but also question (1). To find the best fit to the data $T_{A_0}^2$ and $P_{A_0}^2$ we need only look at T_{A_0} and P_{A_0} and minimize

$$\chi^2 = \frac{1}{n}\sum_{A=A'}^{A''}\left\{\left[\frac{T_A - T_{A_0}}{\Delta T_{A_0}}\right]^2 + \left[\frac{P_A - P_{A_0}}{\Delta P_{A_0}}\right]^2\right\} \tag{52}$$

subject to the restriction that $L_A = 1$ for $A = A' + 2$, $A' + 4, \ldots, A'' - 2$.

This calculation will produce the best fit calculated values T_A and P_A which are then used to find α_A^2 and β_A^2. In addition, the selection of $\chi^2_{\min}$ as a function of sign combination will aid in choosing the best sign for the quantities T_A and P_A in a chi-squared sense. If, after looking at all sign choices, we find there is no satisfactory $\chi^2_{\min}$ (i.e., $\chi^2_{\min} > 1$), then we conclude that a two-state model will not fit the existing data under assumptions 1 and 2.

II. THE MINIMIZATION TECHNIQUE

The problem of minimizing χ^2 subject to the $L_A = 1$ constraint is itself a very interesting mathematical problem with many properties. In discussing these properties, we consider the string of isotopes from $(A-2)$ to $(A+4)$ as shown in Fig. 2. Changing the signs of all the T's and P's (both calculated and experimental), gives $K_A \to K_A$, $K_{A+2} \to K_{A+2}$, $L_A \to L_A$, and hence $\chi^2_{\min} \to \chi^2_{\min}$. So a change in sign of all experimental and calculated numbers leaves K_A, K_{A+2}, L_A, and $\chi^2_{\min}$ invariant. The transformation

$$\begin{aligned} &T_{A-2} \to -T_{A-2}, \quad && P_{A-2} \to P_{A-2}, \\ &T_A \to T_A, && P_A \to -P_A, \\ &T_{A+2} \to -T_{A+2}, && P_{A+2} \to P_{A+2} \end{aligned}$$

(or its negative) in both the experimental and calculated points leaves L_A and $\chi^2_{\min}$ invariant but changes K_A in such a way that $1 + 4K_A \to (1 + 4K_A)^{-1}$, as can be seen by writing

$$\begin{aligned} 1 + 4K_A = {} & (T_{A-2}T_A - P_{A-2}P_A)^2 \\ & + (T_{A-2} - T_A + P_{A-2} - P_A)^2 \\ & \times [(T_{A-2}T_A - P_{A-2}P_A)^2 \\ & + (T_{A-2} + T_A - P_{A-2} - P_A)^2]^{-1}\,. \end{aligned} \tag{53}$$

Because of this result, we can arbitrarily choose that sign combination which sets $1 + 4K_A \leqslant 1$ (or just $K_A \leqslant 0$) and not affect the goodness of fit in a chi-squared sense. In addition, Eq. (53) requires $1 + 4K_A \geqslant 0$ so we need only consider

$$-\tfrac{1}{4} \leqslant K_A \leqslant 0\,. \tag{54}$$

This result is especially appealing from a physical point of view as $K_A \leqslant 0$ gives $r^2 \leqslant R$ [from Eq. (23)]. Since r measures the "nonorthogonality" of the basis state $2n$-transfer overlap ratio between ϕ_g^A and ϕ_e^{A+2}, we would expect it to be small or at least smaller than R which measures the $2n$-transfer overlap between ϕ_e^A and ϕ_e^{A+2}.

The $\chi^2_{\min}$ for various sign combinations should therefore occur in groups of four with each group given by

$$\{|T_{A-2}|, |T_A|, |T_{A+2}|, |P_{A-2}|, |P_A|, |P_{A+2}|\}\,,$$
$$\{-|T_{A-2}|, -|T_A|, -|T_{A+2}|, -|P_{A-2}|, -|P_A|, -|P_{A+2}|\}\,,$$
$$\{-|T_{A-2}|, |T_A|, -|T_{A+2}|, |P_{A-2}|, -|P_A|, |P_{A+2}|\}\,,$$
$$\{|T_{A-2}|, -|T_A|, |T_{A+2}|, -|P_{A-2}|, |P_A|, -|P_{A+2}|\}\,,$$

as choices of signs.

The calculations involved in minimizing Eq. (52) subject to the $L_A = 1$ conditions in general are difficult because of the nonlinear nature of L_A as a function of the P_A's and T_A's. But, since any useful physical result should have $\chi^2_{\min} \leqslant 1$ and hence calculated values not far from experimental values of the (p,t) and (t,p) ratios, we can expand L_A as a six-dimensional Taylor series about the experimental numbers T_{A_0} and P_{A_0}. This expansion linearizes the minimization problem and reduces it to a standard problem in multivariable calculus whose solution is unique.

We have therefore accomplished our goal in that (1) we have a technique for determining the best fit calculated points in a chi-squared sense; (2) we have expressions for the solutions α_A^2 and β_A^2 that are infinite in number and can be described in terms of the one parameter R; and (3) we have a necessary and sufficient condition for testing our two assumptions and the two-state model. We shall apply this to the germanium region in Sec. VII, but first we want to study some of the properties of the preceding model.

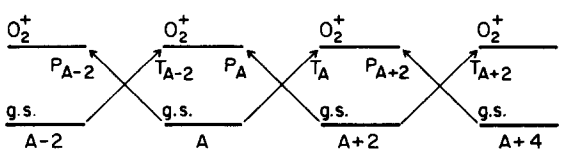


FIG. 2. Schematic representation of the ratios P_{A-2}, P_A, P_{A+2}, T_{A-2}, T_A, and T_{A+2}.

III. THE CONSTRAINT EQUATION: $r^2 = R + K_A(R+1)^2$

By completing the square we obtain

$$\frac{(R+1+1/(2K_A))^2}{[(1+4K_A)/(4K_A^2)]} - \frac{r^2}{[(1+4K_A)/(4K_A)]} = 1 . \quad (55)$$

As $1+4K_A \geqslant 0$, the curve in the R vs r plane is a conic section whose shape depends on the sign of K_A. In particular, for $K_A < 0$ we have an ellipse, for $K_A = 0$, a parabola, and for $K_A > 0$, a hyperbola. Because we can always choose K_A negative, the result will always be an ellipse with the mathematical restriction on R given by

$$|R+1+1/(2K_A)| \leqslant [1+4K_A]^{1/2}/(-2K_A) . \quad (56)$$

This restriction is appealing because physically we would expect R to be bounded. In the same way, r is bounded by

$$|r| \leqslant [(1+4K_A)/(-4K_A)]^{1/2} . \quad (57)$$

Physically we can require $x_A \geqslant 0$ for all A as a choice of basis phase, and so we may restrict the value of R even more. If no value of R can be consistent with both $x_A \geqslant 0$ for all A and Eq. (56), then a two-state model with assumptions 1 and 2 is not adequate to fit the (p,t) and (t,p) cross-section ratio data. As x_A can have singularities [see. Eq. (32)], it becomes more convenient to study, in place of x_A, the mixing potential in a two-state model. This potential is given by

$$U_A = -E_A \alpha_A \beta_A , \quad (58)$$

where E_A is the energy separation between the two physical states (and is a positive number). In terms of x_A, this becomes

$$U_A = -E_A x_A/(1+x_A^2) , \quad (59)$$

and since E_A is positive, the sign of $-U_A$ is the same as that of x_A. In addition, if we use Eq. (45) we can write

$$\frac{-U_A}{E_A} = \frac{v_A(R+1) - r(1+2w_A) - 2v_A}{(R+1)(1+4K_A)} . \quad (60)$$

IV. PROPERTIES OF THE FUNCTIONS α_A^2, β_A^2, x_A, and U_A

We had shown earlier that the best fit chi-squared solutions occur in sets of four. We now show that the functions $\alpha_A^2, \beta_A^2, x_A$, and U_A also occur in sets of four. Table I lists four solutions which all have the same $\chi^2_{\min}$ and K_A value. These four solutions, however, are not all independent in that sets 2, 3, and 4 in Table I are all related to set 1. To see this, we use the full functional notation for $\alpha_A^2, \beta_A^2, x_A$, and U_A. We therefore write Eq. (32) as,

$$x_A(r,R;\mathbf{P}_A,\mathbf{T}_A) = \frac{R + w_A(R+1)}{-r + v_A(R+1)} = \frac{-r - v_A(R+1)}{1 + w_A(R+1)} ,$$

TABLE I. Four sets, all with the same χ^2 min, $L_A = 1$, and $-\frac{1}{4} \leqslant K_A < 0$. Here $(\mathbf{P}_A, \mathbf{T}_A)$ represents a set of P_A, and T_A values as defined in the text.

Set #	r	$(\mathbf{P}_A,\mathbf{T}_A)$
1	+	$(\mathbf{P}_A,\mathbf{T}_A)$
2	−	$(\mathbf{P}_A,\mathbf{T}_A)$
3	+	$(-\mathbf{P}_A,-\mathbf{T}_A)$
4	−	$(-\mathbf{P}_A,-\mathbf{T}_A)$

where $\mathbf{P}_A$ and $\mathbf{T}_A$ represent the set of P_A and T_A numbers. It is clear from this expression and Eq. (23) that

$$x_A(-r,R;-\mathbf{P}_A,-\mathbf{T}_A) = -x_A(r,R;\mathbf{P}_A,\mathbf{T}_A) , \quad (61)$$

and

$$x_A(-r,1/R;\mathbf{P}_A,\mathbf{T}_A) = -1/x_A(r,R;\mathbf{P}_A,\mathbf{T}_A) , \quad (62)$$

where $-\mathbf{P}_A$ and $-\mathbf{T}_A$ mean change the sign of all the P_A and T_A numbers. So sets 2 and 1 are related by

$$x_A(-r,R;\mathbf{P}_A,\mathbf{T}_A) = -1/x_A(r,1/R;\mathbf{P}_A,\mathbf{T}_A)$$

yielding

$$\begin{array}{ll} \text{Set 2} & \text{Set 1} \\ \alpha_A^2(-r,R;\mathbf{P}_A,\mathbf{T}_A) = \beta_A^2(r,1/R;\mathbf{P}_A,\mathbf{T}_A), \\ \beta_A^2(-r,R;\mathbf{P}_A,\mathbf{T}_A) = \alpha_A^2(r,1/R;\mathbf{P}_A,\mathbf{T}_A) , \\ U_A(-r,R;\mathbf{P}_A,\mathbf{T}_A) = -U_A(r,1/R;\mathbf{P}_A,\mathbf{T}_A) . \end{array}$$

Sets 3 and 1 are related by

$$x_A(r,R;-\mathbf{P}_A,-\mathbf{T}_A) = 1/x_A(r,1/R;\mathbf{P}_A,\mathbf{T}_A)$$

yielding

$$\begin{array}{ll} \text{Set 3} & \text{Set 1} \\ \alpha_A^2(r,R;-\mathbf{P}_A,-\mathbf{T}_A) = \beta_A^2(r,1/R;\mathbf{P}_A,\mathbf{T}_A), \\ \beta_A^2(r,R;-\mathbf{P}_A,-\mathbf{T}_A) = \alpha_A^2(r,1/R;\mathbf{P}_A,\mathbf{T}_A) , \\ U_A(r,R;-\mathbf{P}_A,-\mathbf{T}_A) = U_A(r,1/R;\mathbf{P}_A,\mathbf{T}_A) . \end{array}$$

Sets 4 and 1 are related by

$$x_A(-r,R;-\mathbf{P}_A,-\mathbf{T}_A) = -x_A(r,R;\mathbf{P}_A,\mathbf{T}_A)$$

yielding

$$\begin{array}{ll} \text{Set 4} & \text{Set 1} \\ \alpha_A^2(-r,R;-\mathbf{P}_A,-\mathbf{T}_A) = \alpha_A^2(r,R;\mathbf{P}_A,\mathbf{T}_A) , \\ \beta_A^2(-r,R;-\mathbf{P}_A,-\mathbf{T}_A) = \beta_A^2(r,R;\mathbf{P}_A,\mathbf{T}_A) , \\ U_A(-r,R;-\mathbf{P}_A,-\mathbf{T}_A) = -U_A(r,R;\mathbf{P}_A,\mathbf{T}_A) . \end{array}$$

In essence, physically, sets 1 and 3 (and sets 2 and 4) are related by just interchanging the nature of ϕ_g^A and ϕ_e^A. Thus, only one set of these four is independent and could be used to describe the wave functions. Of course we should choose that set yielding $-U_A \geqslant 0$, as we can physically require $x_A \geqslant 0$ for all A.

From the constraint equation (23), if we take d/dR of both sides, we get

$$\frac{dr}{dR} = \frac{1 + 2K_A(R+1)}{2r} . \quad (63)$$

Equations (33) and (60) then produce

$$\frac{d\alpha_A^2}{dR} = \frac{U_A}{E_A r(R+1)} , \quad (64)$$

showing that the critical points for α_A^2 correspond to the zeros of U_A. Solving for the critical points, $R_C(A)$, yields $U_A(R_c(A)) = 0$ or

$$R_c^{\pm}(A) = -(1 + w_A^{-1})^{\pm 1} . \quad (65)$$

The corresponding $r_c(A)$ values are then obtained from

$$r_c^{\pm}(A) = \pm v_A(R_c^{\pm}(A) + 1) , \quad (66)$$

and the values $\alpha_A^2(R_c^{\pm}(A))$ are

$$\alpha_A^2(-(1 + w_A^{-1})^{\pm 1}) = (1 \pm 1)/2 . \quad (67)$$

V. SIGNIFICANCE OF THE PARAMETERS: w_A, v_A, L_A, Z_A, AND K_A

A. The parameter w_A

Mathematically, w_A gives the location of critical points for α_A^2. Physically, since $U_A(R_c^{\pm}(A)) = 0$, w_A also gives a measure of the physical region for R in the following way. Since we can always choose $-\frac{1}{4} \leqslant K_A \leqslant 0$, then from Eq. (30)

$$(w_A + \tfrac{1}{2})^2 + v_A^2 = K_A + \tfrac{1}{4} \tag{68}$$

would give

$$(w_A + \tfrac{1}{2})^2 + v_A^2 \leqslant \tfrac{1}{4},$$

showing that the allowed values of w_A and v_A in the $w_A v_A$ plane lies inside a circle of radius $\frac{1}{2}$ and center $(-\frac{1}{2},0)$ and that $-\frac{1}{2} \leqslant v_A \leqslant \frac{1}{2}$ and $-1 \leqslant w_A \leqslant 0$. If we choose the $r > 0$ set, then we can define

$$M = \min_{\{A \mid v_A < 0\}} \{ -(1 + w_A^{-1})^{-1} \}$$

and

$$m = \max_{\{A \mid v_A > 0\}} \{ -(1 + w_A^{-1})^{+1} \},$$

then $-U_A \geqslant 0$ for all A in the region $m \leqslant R \leqslant M$. For any other set (e.g., $r < 0$) we can use the results of Eqs. (61) and (62) to get the physical bound on R. If we couple this with the mathematical limits of the ellipses

$$\frac{-(1+2K_A) + \sqrt{1+4K_A}}{2K_A} \leqslant R \leqslant \frac{-(1+2K_A) - \sqrt{1+4K_A}}{2K_A},$$

we get the physical bounds on R given by

$$\max\left[\frac{-(1+2K_A) + \sqrt{1+4K_A}}{2K_A}, m\right] \leqslant R \leqslant \min\left[\frac{-(1+2K_A) - \sqrt{1+4K_A}}{2K_A}, M\right]. \tag{69}$$

If these restrictions cannot be met, then the two-state model with assumptions 1 and 2 will not fit the (p,t) and (t,p) cross-section ratio data for any value of R.

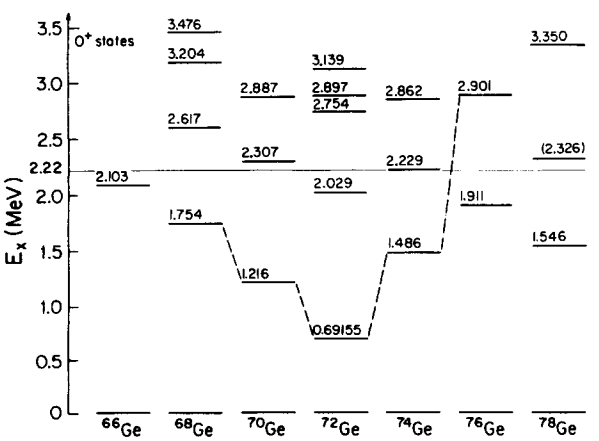


FIG. 3. Low-lying 0^+ states in the even germanium isotopes.

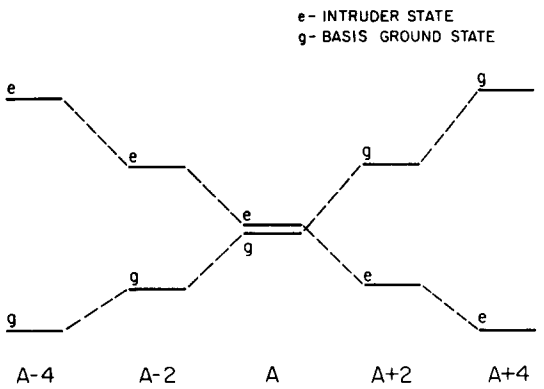


FIG. 4. Schematic representation of the basis ground state and basis intruder state in a string of isotopes.

B. The parameter v_A

We have seen that the parameter w_A gives a measure of the physical region of R. The parameter v_A gives the choice of which critical point occurs depending on whether the $r \geqslant 0$ or $r \leqslant 0$ solution is chosen. In particular, from Eqs. (65), (66), and (67), if the $r \geqslant 0$ solution set is chosen, then for $v_A > 0$ we have $R_c^+(A) = -(1 + w_A^{-1})^{+1}$ and $\alpha_A^2(R_c^+(A)) = 1$ while for $v_A < 0$, we have $R_c^-(A) = -(1 + w_A^{-1})^{-1}$ and $\alpha_A^2(R_c^-(A)) = 0$. The sign of v_A gives a measure of whether α_A^2 is 0 or 1 at its critical points. In particular, using $r \geqslant 0$ shows that for $\alpha_A^2(R_c(A))$ to give 1, we need $v_A > 0$ and $\alpha_A^2(R_c(A))$ gives 0 when $v_A < 0$. So v_A can be a measure of whether the physical state will favor ϕ_g^A or ϕ_e^A. We shall elaborate later when we look at the method applied to the germanium isotopes.

C. The parameter L_A

The significance of the parameter L_A is that it serves as a test of the entire model mathematically in that $L_A = 1$ serves as the necessary and sufficient condition. Physically, $L_A = 1$ is just the statement that

$$\sigma(^A X(t,p)^{A+2} X(\text{g.s.})) = \sigma(^{A+2}X(p,t)\,^A X(\text{g.s.})),$$

assuming, of course, that all kinematic corrections have been divided out.

D. The parameter z_A

Equation (46) gives $S_A(t,p) = -z_A K_A f_A^2 (R+1)^2$ showing that the ratio z_{A+2}/z_A is the ratio of $S_{A+2}(t,p)$ to $S_A(t,p)$,

TABLE II. Experimental 0^+ cross-section ratios in the germanium isotopes. Here E_A is the excitation energy of the excited $0^{+\prime}$ state.

A	E_A (MeV)	$\dfrac{\sigma(^{A+2}\text{Ge}(p,t)\,^A\text{Ge}(0^+))}{\sigma(^{A+2}\text{Ge}(p,t)^A\text{Ge}(\text{g.s.}))}$	$\dfrac{\sigma(^A\text{Ge}(t,p)\,^{A+2}\text{Ge}(0^+))}{\sigma(^A\text{Ge}(t,p)\,^{A+2}\text{Ge}(\text{g.s.}))}$
68	1.753	$0.0058 \pm 8\%$	
70	1.216	$0.068 \pm 4\%$	$0.0020 \pm 20\%$
72	0.6915	$0.280 \pm 3.5\%$	$0.200 \pm 5\%$
74	1.486	$0.010 \pm 80\%$	$0.025 \pm 10\%$
76	2.901		$0.0167 \pm 8\%$
78	2.326		

TABLE III. DWBA 0^+ cross-section ratios in the germanium isotopes. Here E_A is the excitation energy of the excited $0^{+\prime}$ state.

A	E_A (MeV)	$\frac{\sigma(^{A+2}\mathrm{Ge}(p,t)\,^{A}\mathrm{Ge}(0^+))}{\sigma(^{A+2}\mathrm{Ge}(p,t)\,^{A}\mathrm{Ge}(\mathrm{g.s.}))}$	$\frac{\sigma(^{A}\mathrm{Ge}(t,p)\,^{A+2}\mathrm{Ge}(0^+))}{\sigma(^{A}\mathrm{Ge}(t,p)\,^{A+2}\mathrm{Ge}(\mathrm{g.s.}))}$
68	1.753	$0.8198 \pm 2\%$	
70	1.216	$1.0040 \pm 2\%$	$1.0230 \pm 5\%$
72	0.6915	$1.0450 \pm 3\%$	$1.0110 \pm 3\%$
74	1.486	$1.1370 \pm 2\%$	$0.9660 \pm 3\%$
76	2.901		$0.8780 \pm 2\%$
78	2.326		

i.e., physically, z_A is a measure of summed (t,p) and (p,t) strengths.

E. The parameter K_A

Finally, the parameter K_A provides the connection between the physical state $2n$-transfer overlap ratios represented by P_A and T_A and the basis states of $2n$-transfer overlap ratios r and R. In addition the parameter K_A along with the critical points $R_c^{\pm}(A)$ given by Eq. (65) can give a measure of the physical (but unmeasurable) cross-section sum $Q_A(t,p)$ given by

$$Q_A(t,p) = \sigma(^{A}X(0^{+\prime})(t,p)^{A+2}X(\mathrm{g.s.})) + \sigma(^{A}X(0^{+\prime})(t,p)^{A+2}X(0^{+\prime}))\,.$$

We could now easily show that

$$S_A(t,p) + Q_A(t,p) = f_A^2(1 + 2r^2 + R^2)\,, \tag{70}$$

which is just the condition that total flux is conserved. Putting in Eqs. (23) and (46) gives

$$\frac{Q_A(t,p)}{S_A(t,p)} = \frac{(1+w_A)+K_A}{K_A - w_A} = \frac{K_A + (R_c^-(A)+1)^{-1}}{K_A + (R_c^+(A)+1)^{-1}}\,,$$

giving the A dependence of Q_A/S_A.

VI. A SIMPLE LIMITING CASE

If $K_A \to -\frac{1}{4}$, Eq. (23) yields $R \to 1$ and $r \to 0$. Equation (53), rewritten as

$$1 + 4K_A = \frac{(T_A + P_A z_A)^2 + (1 - z_A)^2}{(T_A + P_A z_A)^2 + (1 + z_A)^2}\,,$$

shows that $z_A \to 1$ and $T_A + P_A \to 0$, for all A. In addition, we get $w_A \to -\frac{1}{2}$ and $v_A \to 0$. Such a result is an important limiting case as previous models of this type[1] have assumed $R = 1$ and $r = 0$, which is equivalent, via Eq. (23), to $K_A = -\frac{1}{4}$. As $K_A = -\frac{1}{4}$ if and only if $P_A + T_A = 0$, we have an immediate test to whether such a naive model ($r = 0$, $R = 1$) will work. The quantity $1 + 4K_A$ is then a measure of the deviation from the simple model.

VII. EXAMPLE: APPLICATION TO THE GERMANIUM DATA

In many regions of the periodic table, there exist chains of nuclei in which an intruder state (or more than one) is obviously present. A signature of such a state can be found in the low-lying energy-level spectra. For example, in the germanium isotopes (Fig. 3), we notice a parabolic dependence in E_A (the energy of the 0_2^+ state) with A. Such a phenomenon is interpreted by saying that the ground state and the physical $0^{+\prime}$ state (usually the 0_2^+ state, but need not be) result from mixing between the basis ground state and a basis 0_e^+ excited intruder state. These basis states mix or interfere with each other and can therefore "switch" positions as we move from lighter mass to heavier. This situation is shown schematically in Fig. 4. These nuclei, for which an intruder state exists, are somewhat collective, but not strongly deformed. From Fig. 3, we see that the place where the basis 0_e^+ and basis ground state of germanium "switch" position occurs at the minimum in E_A—i.e., around $N = 40$ ($A = 72$). There exists in the literature[1] a staggering amount of data for the germanium region, all of which also indicates a change in structure occurring at $N = 40$. Many models have been proposed[1] to try and explain the origin of this transition, all with limited success. We shall show that the generalized two-state model developed above will not only account for the Ge(p,t) and Ge$(t,p)\sigma(0^{+\prime})/\sigma(\mathrm{g.s.})$ ratio data, but also present a possible explanation for the existence of this transition between the lighter and heavier mass germanium isotopes.

We shall use the Ge(p,t) and Ge$(t,p)\sigma(0^{+\prime})/\sigma(\mathrm{g.s.})$ data measured[2–10] in the literature. These are summarized in Table II. In the above model, all (p,t) and (t,p) $0^{+\prime}$/g.s. cross-section ratios must be corrected for Q-value effects. To compensate for Q-value effects, DWBA calculations were performed with the code DWUCK,[11] using optical-model parameters from Ref. 9, and a two-neutron bound-state (B.S.) wave function of the form

TABLE IV. Q-corrected 0^+ cross-section ratios and calculated ratios using the sign combinations given in Table VI in the Ge isotopes. Here E_A is the excitation energy of the excited $0^{+\prime}$ state.

		Q-corrected ratios		Calculated ratios	
A	E_A (MeV)	$P_{A_0}^2$	$T_{A_0}^2$	P_A^2	T_A^2
68	1.753	0.0071 ± 0.0007		0.0071	
70	1.216	0.0680 ± 0.0040	0.0020 ± 0.0005	0.0680	0.0020
72	0.6915	0.2700 ± 0.0160	0.2000 ± 0.0160	0.2740	0.1950
74	1.486	0.0090 ± 0.0072	0.0250 ± 0.0034	0.0060	0.0250
76	2.901		0.0190 ± 0.0020		0.0190
78	2.326				

TABLE V. The sign combinations in the P_A's and T_A's that give the best fit chi-squared. Also given are the corresponding K_A values.

set	P_{70}	P_{72}	P_{74}	T_{70}	T_{72}	T_{74}	K_A
1	−	−	−	+	+	+	−0.2452
2	−	+	−	−	+	−	12.7519
3	+	−	+	+	−	+	12.7519
4	+	+	+	−	−	−	−0.2452

$$\Psi_{\text{B.S.}}(2n) = \frac{-N}{\sqrt{2(9/2)+1}}(1g_{9/2})_0^2 + \frac{N}{\sqrt{2(5/2)+1}}(1f_{5/2})_0^2 + \frac{N}{\sqrt{2(3/2)+1}}(2p_{3/2})_0^2 + \frac{N}{\sqrt{2(1/2)+1}}(2p_{1/2})_0^2 ,$$

where $N = 0.992$ for normalization.

To reduce Coulomb effects, calculations were performed for all beam energies at which experimental numbers were measured. Ratios of the DWBA cross-sections are summarized in Table III. To investigate configuration dependences in the DWBA calculations, we also ran the code DWUCK for two-neutron bound-state wave functions of pure $(1g_{9/2})_0^2$ and pure $(2p_{1/2})_0^2$ and used these results to estimate uncertainties in the DWBA ratios. The final Q-corrected ratios, along with their uncertainties to be used in this analysis, are summarized in Table IV. Note that in the above data and DWBA calculations, we are using the 0_2^+ state as the physical excited state in $^{68-74}$Ge, and the 0_3^+ state in ^{76,78}Ge. This is because from Fig. 3 we see that from ^{68}Ge to ^{78}Ge, each isotope has a 0^+ state at about 2.22 MeV as indicated by the solid horizontal line. These states at about 2.22 MeV could be inert for all the germanium isotopes and not mix with any other 0^+ states. If that is the case, then for ^{76}Ge, the next candidate to consider with the ground state in a two-state model would be the 0_3^+ state at 2.901 MeV. For ^{78}Ge, the 0_4^+ state at 3.350 MeV would seem too far away from the ground state and the 0_2^+ state at 1.546 MeV seems to contradict the parabolic trend in E_A with A. Although the 0_3^+ state at 2.326 MeV would want to fall into the possible "inert" state category mentioned above, we shall choose the worst of three evils and assume that the 0_3^+ at 2.326 MeV in ^{78}Ge is the excited $0^{+\prime}$ mixed state. Since ^{78}Ge represents an "endpoint" in the germanium isotope chain, the particular choice of state here will not affect any of the calculations for $^{68-76}$Ge. The reason is because data exist for enough nuclei (four) to calculate one L_A, viz., L_{72} and so the cross-section ratio in

TABLE VI. The sign combinations in the P_A's and T_A's that give the best fit chi-squared. Each set has $K_A = -0.2452$.

set	r	P_{70}	P_{72}	P_{74}	T_{70}	T_{72}	T_{74}
a	+	−	−	−	+	+	+
b	−	−	−	−	+	+	+
c	+	+	+	+	−	−	−
d	−	+	+	+	−	−	−

TABLE VII. Calculated values of the parameters K_A, v_A, w_A, and z_A for Ge. These are calculated using set a of Table VI, $P_{68} < 0$, and $T_{76} > 0$. (Note that v_A changes sign as A goes from 70 to 76.)

A	K_A	v_A	w_A	z_A
70	−0.2452	+0.0629	−0.5292	1.1583
72	−0.2452	+0.0502	−0.5438	1.2342
74	−0.2452	−0.0141	−0.5679	1.3159
76	−0.2452	−0.0314	−0.5618	1.2912

^{76}Ge(t,p) ^{78}Ge will not be involved in the best fit minimization procedure. In fact, in that procedure, we need only minimize

$$\chi^2 = \left(\frac{1}{6}\right) \sum_{A=70}^{74} \left\{ \left[\frac{(P_A - P_{A_0})}{\Delta P_{A_0}} \right]^2 + \left[\frac{(T_A - T_{A_0})}{\Delta T_{A_0}} \right]^2 \right\},$$

subject to the one condition $L_{72} = 1$. The results of this calculation show that the "best" fit in a chi-squared sense occurs for four sign combinations in the P_A's and T_A's. These are given in Table V along with the K_A values. Each calculation gives $\chi^2_{\min} \approx 0.068$ and K_A values which are related via $1 + 4K_A \rightarrow (1 + 4K_A)^{-1}$ as promised earlier in the model-developing sections above. Also as promised, we can always choose the negative K_A value leaving us with only two sign choices. As for the constraint equation [Eq. (23)], we also have two choices of sign for r, leaving a total of the four possibilities given in Table VI. These all have been shown to be related in Table I and so only one of these four needs to be considered. Using set a in Table VI and assuming that $P_{68} < 0$ and $T_{76} > 0$ allows one to calculate the unmeasurable ratios T_{68} and P_{76} so that L_{70} and L_{74} are equal to 1. Knowledge of P_A and T_A for $A = 68$ to 76 allows us to calculate the parameters K_A, z_A, w_A, and v_A (for $A = 70$, 72, 74, and 76),

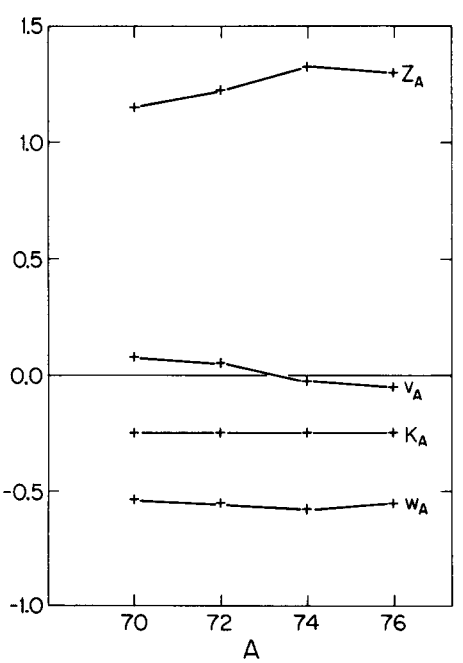


FIG. 5. Plots of z_A, w_A, v_A, and K_A vs A for the germanium data.

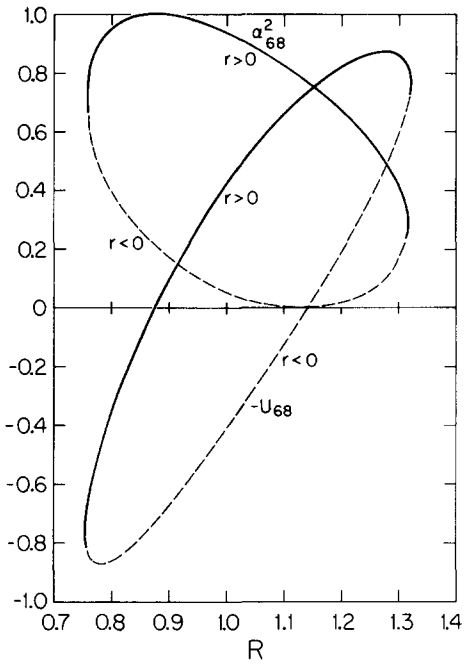


FIG. 6. Plots of α^2_{68} and $-U_{68}$ (in MeV) versus R for ^{68}Ge. (The solid curve is for $r\geqslant 0$ while $r<0$ is given by the dashed curve.)

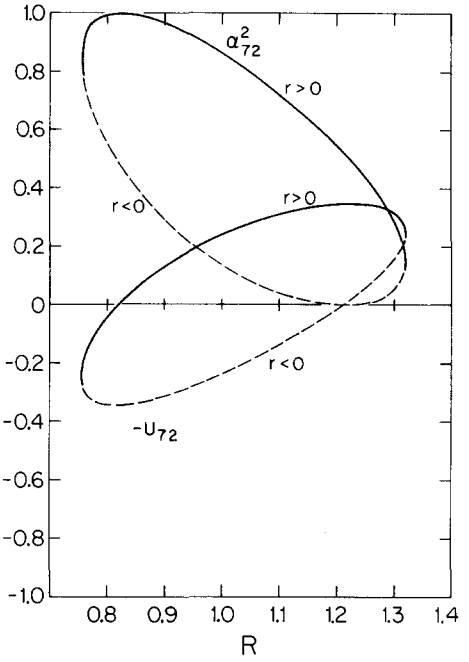


FIG. 8. Plots of α^2_{72} and $-U_{72}$ (in MeV) vs R for ^{72}Ge. (The solid curve is for $r\geqslant 0$ while $r<0$ is given by the dashed curve.)

which are summarized in Table VII and plotted versus A in Fig. 5. Note that v_A changes sign as we move from lighter mass to heavier suggesting a change in structure occurring in the sense that the value of $\alpha^2_A(R^{\pm}_c(A))$ flips from 0 to 1 or vice versa. The quantities α^2_A and $-U_A$ (in MeV) For $A = 68$ to 78 can be calculated using Eqs. (33) and (60) and are plotted (for both $r\geqslant 0$ and $r\leqslant 0$) as functions of R in Figs. 6–11. Notice, as predicted by the v_A parameter, the flip in structure from lighter A to heavier. As indicated earlier, we need only consider, say, $r\geqslant 0$ (i.e., set a of Table VI). The α^2_A and $-U_A$ (in MeV) resulting from these are all plotted on one graph as shown in Figs. 12 and 13. (Note that the plots for α^2_A and $-U_A$ for $A = 68$ and 78 are dotted indicating that the signs of P_{68} and T_{76} are chosen as negative and positive, respectively, and are not determined by the chi-squared minimization process.) Each of the α^2_A will produce fits to the data which are summarized in Table IV and plotted in Fig. 14 for any value of R. The only limits on R are given by Eq. (69), which for the germanium data given is $0.889\leqslant R\leqslant 1.282$.

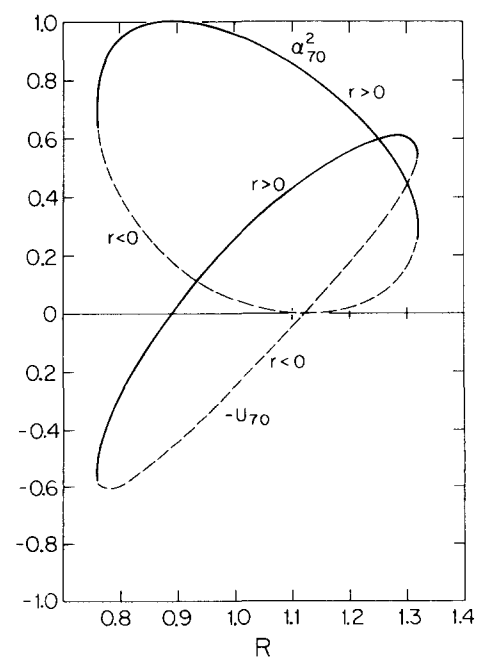


FIG. 7. Plots of α^2_{70} and $-U_{70}$ (in MeV) vs R for ^{70}Ge. (The solid curve is for $r\geqslant 0$ while $r<0$ is given by the dashed curve.)

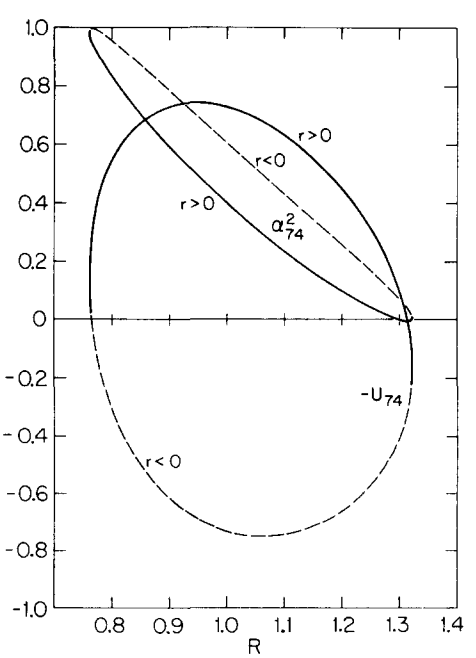


FIG. 9. Plots of α^2_{74} and $-U_{74}$ (in MeV) vs R for ^{74}Ge. (The solid curve is for $r\geqslant 0$ while $r<0$ is given by the dashed curve.)

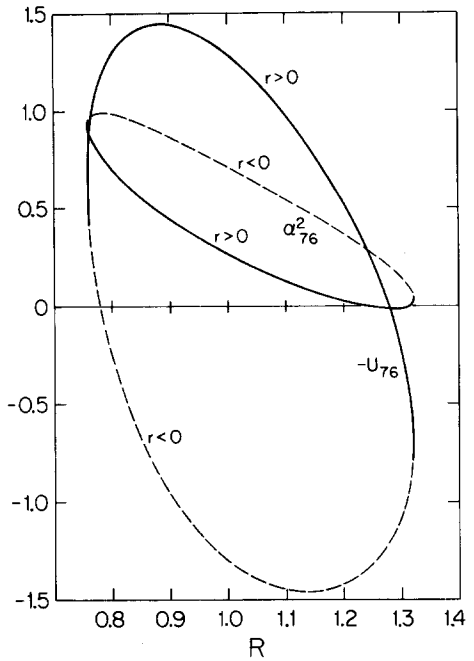


FIG. 10. Plots of α^2_{76} and $-U_{76}$ (in MeV) vs R for ^{76}Ge. (The solid curve is for $r\geq 0$ while $r<0$ is given by the dashed curve.)

Especially evident in Fig. 12 again, is the change in structure between the lighter and heavier Ge isotopes with the transition occurring between ^{72}Ge and ^{74}Ge. Therefore the origin of the transition in the Ge isotopes is due to the drastic change in mixing (α^2_A) from the lighter mass to the heavier. Notice also in Fig. 13 the congestion occurring between $R = 1.06$ and 1.26, suggesting that in that region of R, the perturbed two-state Hamiltonian causing the mixing is slowly changing with A.

The value of $K_A = -0.2452 \pm 0.0009$ shows that K_A is

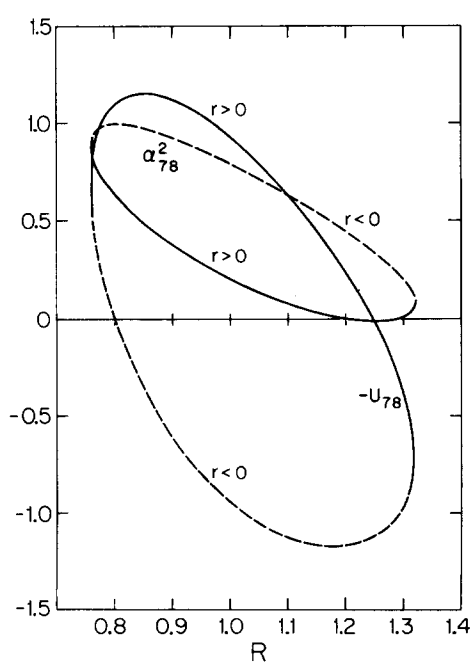


FIG. 11. Plots of α^2_{78} and $-U_{78}$ (in MeV) vs R for ^{78}Ge. (The solid curve is for $r\geq 0$ while $r<0$ is given by the dashed curve.)

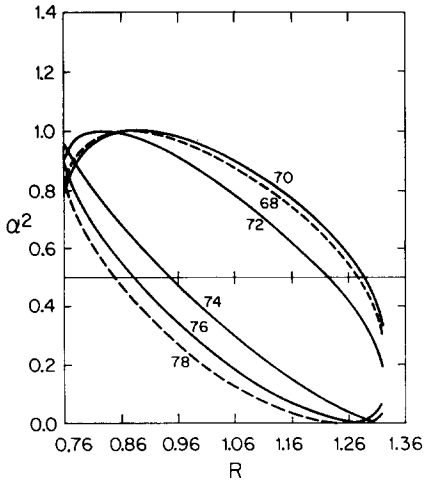


FIG. 12. Plots of α^2_A vs R for the germanium data with $r\geq 0$.

very close to its minimum value of $-\frac{1}{4}$ but is still *five* probable errors away from this limiting value of $-\frac{1}{4}$. This result implies that previous choices[1] of ϕ^A_g and ϕ^A_e assuming $R = 1$ and $r = 0$ and hence $K_A = -\frac{1}{4}$ are incorrect. The nonzero value of r in the physical range shows some "nonorthogonality" between ϕ^A_g and ϕ^{A+2}_e. This nonorthogonality probably arises in the process of constructing particle-hole states of good isospin, and must be considered in the construction of ϕ^A_g and ϕ^A_e.

VIII. CONCLUSION

We have shown that a two-state model can be used to describe the $\sigma(0^+{}')/\sigma(\text{g.s.})$ cross-section ratio in (t,p) and (p,t). We have determined necessary and sufficient conditions for such a model to work based on two simple assumptions. Although we do not describe ϕ^A_g or ϕ^A_e in terms of any shell-model basis or similar type descriptions, we do give starting points by describing how the basis state $2n$-transfer overlaps must behave via the constraint equation, $r^2 = R + K_A(R+1)^2$ and Eqs. (46) and (49). The quantity K_A is an extremely sensitive measurement of the $2n$-transfer overlap ratios and in the case of the germanium data, is suffi-

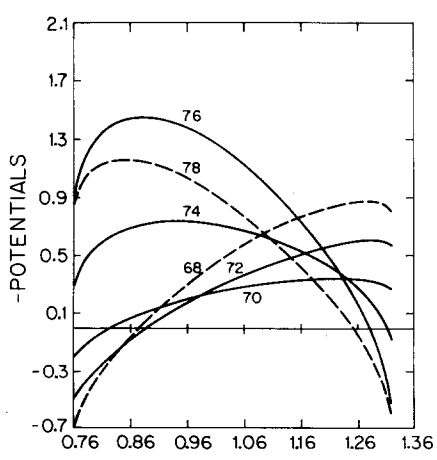


FIG. 13. Plots of $-U_A$ (in MeV) vs R for the germanium data with $r\geq 0$.

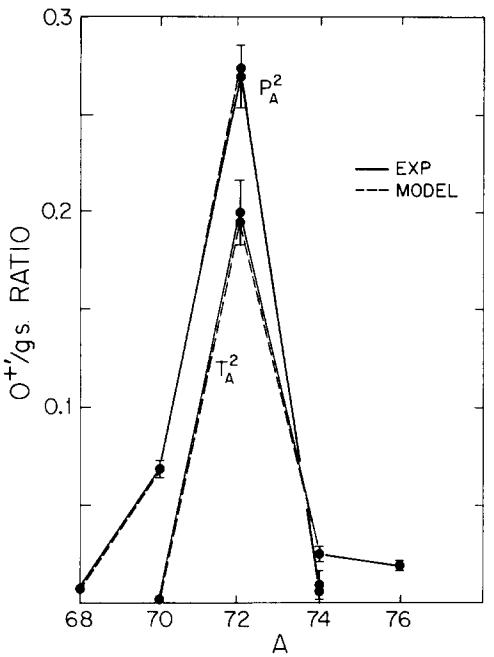


FIG. 14. Experimental and calculated $\sigma(0^{+\prime})/\sigma$(g.s.) ratios for the Ge data.

ciently far from its minimum value of $-\frac{1}{4}$ that the simple models previously applied cannot work. We have been able to fit the $0^{+\prime}$/g.s. state (p,t) and (t,p) cross-section ratio data almost exactly and the results of the fits exhibit the change in structure between the lighter and heavier masses.

The solution of the present germanium data problem allows one to use the resulting wave functions to predict other observables. We have shown these wave functions to be consistent with $1f_{5/2}$ proton occupation numbers in the ^{A}Ge ground states,[12] with $(d,^6\text{Li})$ and $(^6\text{Li},d)$ cross-section ratios leading to ^{A}Ge (see Ref. 13), and with BE2 ratios from 2_2^+, 2_1^+, to 0_2^+, 0_1^+, states in ^{70}Ge and ^{72}Ge (see Ref. 14). With very few assumptions, we have been able to fit much data and we still do not need to specify any value of R, so that we continue to have flexibility in fitting even more data.

We have done a survey of this analysis to (t,p) and (p,t) data for all nuclei (total of 15) from calcium to uranium and have found the present model successful in every case. The detailed results of this work will appear in a later publication.

Other possible special cases of Eq. (16), i.e., $R+1=0$ and $R-rs=0$, have already been considered by us and these results will appear in a later publication. We are also working on the general solution to Eq. (16) for any values of r, s and R.

ACKNOWLEDGMENT

We acknowledge financial support from the National Science Foundation.

[1]M. Vergnes, Inst. Phys. Conf. Ser. **49**, 25 (1980), and references therein.
[2]F. Guilbault, D. Ardouin, J. Uziereau, P. Avignon, R. Tamisier, G. Rotbard, M. Vergnes, Y. Deschamps, G. Berrier, and R. Seltz, Phys. Rev. C **16**, 1840 (1977).
[3]G. C. Ball, R. Fournier, J. Kroan, T. H. Hsu, and B. Hird, Nucl. Phys. A **231**, 334 (1974).
[4]A. C. Rester, J. B. Ball, and R. L. Auble, Nucl. Phys. A **346**, 371 (1980).
[5]C. Lebrun, F. Guilbault, D. Ardouin, E. R. Flynn, D. C. Hanson, S. D. Orbesen, R. Rotbard, and M. N. Vergnes, Phys. Rev. C **19**, 1224 (1979).
[6]S. Mordechai, H. T. Fortune, R. Middleton, and G. Stephans, Phys. Rev. C **19**, 1733 (1979).
[7]S. LaFrance, S. Mordechai, H. T. Fortune, and R. Middleton, Nucl. Phys. A **307**, 52 (1978).
[8]S. Mordechai, H. T. Fortune, R. Middleton, and G. Stephans, Phys. Rev. C **18**, 2498 (1979).
[9]J. F. Mateja, L. R. Medsker, H. T. Fortune, R. Middleton, G. E. Moore, M. E. Cobern, S. Mordechai, J. D. Zumbro, and C. P. Browne, Phys. Rev. C **17**, 2047 (1978).
[10]F. Guilbault, D. Ardouin, R. Tamisier, P. Avignon, M. Vergnes, G. Rotbard, G. Berrier, and R. Seltz, Phys Rev. C **16**, 1840 (1977); **15**, 894 (1977).
[11]P. D. Kunz (private communication).
[12]H. T. Fortune, M. Carchidi, and S. Mordechai, Phys. Lett. B **145**, 4 (1984).
[13]M. Carchidi and H. T. Fortune, Phys. Rev. C **31**, 853 (1985).
[14]M. Carchidi, thesis, University of Pennsylvania, 1985.

Existence and uniqueness properties for the one-dimensional magnetotellurics inversion problem

John A. MacBain
Sohio Petroleum Company, One Lincoln Center—Suite 1200, 5400 LBJ Freeway, Dallas, Texas 75240
J. Bee Bednar
5867 South Joplin, Tulsa, Oklahoma 74135

(Received 7 March 1985; accepted for publication 27 September 1985)

The one-dimensional magnetotelluric (MT) inversion problem is well known to be ill posed and nonlinear. This paper seeks an understanding of the mappings underlying the nonlinear relationships. These properties are used to study the validity of some exploration aspects of the problem that are essential for the practical use of MT as an exploration tool. This study has two major segments—results in response space and results in geologic space. In response space, the existence of optimal admittance curves has been proven. In the case of continuous data, uniqueness has been established. For discrete data, a relationship between alternative optimal solutions has been derived. In geologic space, the results are extremely significant to MT's applicability to exploration. A class of conductivity functions Y, which contains all possible natural geological conditions, has been used as a framework for the study. Over this class Y, the one-dimensional magnetotellurics problem has been proven to be uniquely invertible from the admittance curves. This precludes the possibility of two different geologic models yielding the same complete data. As a result, the path is now clear to generate a description of a unimodal statistical distribution of feasible inversions to the real-world exploration problem with finitely sampled noisy data.

I. INTRODUCTION

Many investigators have contributed to the theory of the one-dimensional magnetotelluric (MT) inversion problem. This survey will be intentionally brief and touch only the more significant recent results. For a related problem, Bailey[1] proved that the inverse was unique, given complete data on the surface of a stratified sphere provided the model space is restricted to nonzero, bounded, infinitely differentiable functions. (Bailey studied the ratio of the externally generated magnetic field to the internally generated magnetic field.) The next major advance was in a paper by Weidelt[2] in which an actual inversion algorithm was presented. This procedure, an application of the Gel'fand–Levitan theory, generates continuous conductivity profiles when the data curve corresponds to a continuous conductivity profile. However, it is not clear that the procedure will recover the same σ as that which generated the data.

The practicability of the inversion was greatly expanded by Sabatier.[3] Sabatier demonstrated the application of linear programming to finite-dimensional approximations of the global problem. More recently, Parker[4,5] has modified and enhanced these techniques to yield conductivity profiles that will fit noisy finite data (arbitrarily close to the optimal possible fit).

II. PRELIMINARIES

This paper will follow the notation and terminology of Parker[4] where possible. The basic equation describing a horizontally polarized electric field of radial frequency ω diffusing into a conductive body is

$$\frac{\partial^2 E(z,\omega)}{\partial^2 z} = i\omega\mu_0\sigma(z)E(z,\omega), \tag{1}$$

where $\sigma(z)$ is the conductivity profile for $z < h$ (h very large) and is assumed to be a constant $\sigma_0 > 0$ when $z \leqslant 0$. The lower boundary conditions will be $E(0,\omega) = 1$ and $(\partial E/\partial z)(0,\omega) = (1+i)\sqrt{\omega\mu_0\sigma_0}/2$. The ratio

$$c(\omega) = \frac{E(h,\omega)}{(\partial E/\partial z)(h,\omega)} \tag{2}$$

is defined to be the admittance curve. The more familiar apparent resistivity is $\rho_a(\omega) = \sqrt{\omega\mu_0}|c(\omega)|^2$. We will consider the inverse problem both where $c(\omega)$ is known at a finite number of frequencies and also on an interval.

The problem will be formulated in terms of geologic models

$$(\sigma,\sigma_0,h), \tag{3}$$

where $\sigma(z) \geqslant 0$ for $z \in [0,h]$, $\sigma(z) = \sigma_0$ for $z \leqslant 0$, and $h > 0$.

Parker[4] demonstrated that the admittance curve $c(\omega)$ corresponding to a geological model can be represented as

$$c(\omega) = b_0 + \int_0^\infty \frac{1 - i\omega\lambda}{\lambda + i\omega}\, db(\lambda), \tag{4}$$

where $b_0 \geqslant 0$ and $b(\lambda)$ is monotone increasing. It can be easily observed that the following formulas hold:

$$\begin{aligned} &b_0 = \mathrm{Re}(c(1)), \\ &0 \leqslant TV(b) = -\mathrm{Im}(c(1)) < \infty, \\ &b_0 \geqslant \int_0^\infty \lambda\, db(\lambda). \end{aligned} \tag{5}$$

If we further require $0 < r \leqslant \rho_A(\omega) \leqslant R$ for all ω, we obtain, respectively,

$$\begin{aligned} &b_0 = \int_0^\infty \lambda\, db(\lambda) \quad (R), \\ &\int_0^\epsilon \frac{db(\lambda)}{\lambda} = \infty, \quad \epsilon > 0 \quad (r). \end{aligned} \tag{6}$$

 0022-2488/86/020645-05$02.50 © 1986 American Institute of Physics

In particular, (6) implies that if $db(\lambda)$ is a fine delta comb corresponding to a model (σ,σ_0,h), then $\lim_{\omega\to 0}\rho_A(\omega)=0$ and $\sigma_0=\infty$.

To establish notation, define

$$c(z,\omega)=\frac{E(z,\omega)}{(\partial E/\partial z)(z,\omega)}.$$

Then

$$\frac{\partial c(z,\omega)}{\partial z}=1-i\mu_0\omega\sigma(z)c^2(z,\omega). \tag{7}$$

Now that some notation has been established, let us give a schematic of the mappings to be studied (see Fig. 1). Under very general settings, f_1 and f_2 will be well defined and nicely behaved. The difficult questions concern the behavior and existence of mappings f_1^{-1} and f_2^{-1}.

III. RESULTS IN RESPONSE SPACE

Here the relationships between $c(\omega)$ and $b(\lambda)$ are analyzed in greater detail. In the case where a finite number of data values are available (as in real experiments), let the data be given by $\mathbf{d}=(d_1,d_2,\ldots,d_n)$, where d_j is the complex data for frequency $\omega_j\in W=\omega_1<\omega_2<\cdots<\omega_n$. If data is available on $[a,u]$, it is given by $d(\omega)\in\mathscr{C}[a,u]$. Define $g(c(\omega))=(c(\omega_1),\ldots,C(\omega_n))\in C^n$. Throughout, the norm in C^n can be modified to reflect statistical knowledge or weightings as in Ref. 4 without affecting results. Let T represent the set of all admittance curves which can be represented as in (4) and (5).

Theorem 1: Let $\mathbf{d}=(d(\omega_1),d(\omega_2),\ldots,d(\omega_n))$ be given and S be a convex subset of T. Then there exists a closed convex subset Q of T such that $c_0\in Q$ implies $\|g(c_0)-\mathbf{d}\|_p=\min_{c\in S}\|g(c)-\mathbf{d}\|_p$ and each member of Q is a pointwise limit of a sequence from S. Moreover, if c_1 and $c_2\in Q$, then $c_1(\omega_j)=c_2(\omega_j)$, whenever $1<p<\infty$ $(j=1,\ldots,n)$.

Proof: Let $\alpha=\min_{c\in S}\|g(c)-\mathbf{d}\|_p$. Choose $c_n\in S$ such that $\|g(c_n)-\mathbf{d}\|_p\to\alpha$.

Claim 1: $\{b_n\}$ is a bounded sequence in NBV$[0,\infty)$. To see this, note that when $c_b\in S$,

$$c_b(\omega)=b_0+\int_0^\infty\frac{\lambda(1-\omega^2)}{\lambda^2+\omega^2}db(\lambda)-i\int_0^\infty\frac{\omega(\lambda^2+1)}{\lambda^2+\omega^2}db(\lambda).$$

If $\lambda\leqslant\omega$, then

$$\omega(\lambda^2+1)/(\lambda^2+\omega^2)\geqslant 1/2\omega.$$

If $\lambda>\omega$, then

$$\omega(\lambda^2+1)/(\lambda^2+\omega^2)>\omega/2.$$

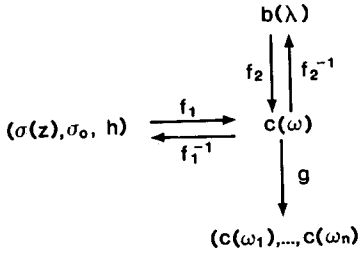


FIG. 1. Schematic of mappings.

Thus, letting $k_0=\min_{j=1,n}\{1/2\omega_j,\omega_j/2\}>0$, one has for all b:

$$\int_0^\infty db(\lambda)<\frac{1}{k_0}\int_0^\infty\frac{\omega(\lambda^2+1)}{\lambda^2+\omega^2}db(\lambda),\quad\text{for }\omega\in W.$$

Thus,

$$\|b\|\leqslant(1/k_0)\min_j[-\mathrm{Im}(c_b(\omega_j))]$$

$$\leqslant(1/k_0)\min_j[|c_b(\omega_j)|].$$

Now for some N_0 and all $n\geqslant N_0$,

$$\|g(c_n)-\mathbf{d}\|_p<\alpha+1.$$

Thus,

$$\|g(c_n)\|_p<\alpha+1+\|\mathbf{d}\|_p.$$

Now

$$\|c_n(\omega_j)\|\leqslant\|g(c_n)\|_p<\alpha+1+\|\mathbf{d}\|_p.$$

Thus,

$$\|b_n\|=\int_0^\infty db_n(\lambda)\leqslant\frac{1}{k_0}(\alpha+1+\|\mathbf{d}\|_p).$$

Now, $\{b_n\}$ being a bounded subset of NBV$[0,\infty)$ means that $\{b_n\}$ has a weak-* convergent subsequence $\{b_{n_k}\}$. Let b^* denote the weak-* limit of this sequence so that

$$\int_0^\infty f\,db(\lambda)=\lim_{k\to\infty}\int_0^\infty f\,db_{n_k}(\lambda),$$

$$f\in C[0,\infty)\text{ and }f\text{ bounded.}$$

We can further require that $\{b_{0n_k}\}$ is convergent. For any c,

$$|\mathrm{Re}(c(\omega_1))-\mathrm{Re}(c(1))|$$
$$=|\mathrm{Re}(c(\omega_1))-b_0|\leqslant|\max f(\omega_1)|\,\|b\|,$$

where f is the real part of the integrand in (4). Thus, for $n\geqslant N_0$ in the original sequence, we know that $\{b_{0n}\}$ is bounded.

Let c^* correspond to the limits $(b_0^*,b^*(\lambda))$.

Claim 2: $\|g(c^*)-\mathbf{d}\|_p=\alpha$.

This follows immediately from the observation that $b_{n_k}\overset{\omega^*}{\to}b^*$, $b_{0n_k}\overset{\omega^*}{\to}b_0^*$ imply

$$c^*(\omega)=\lim_{k\to\infty}c_{b_{n_k}}(\omega)$$

for each w.

Now put $Q=\{c(\omega)\,|\,\|g(c)-\mathbf{d}\|_p=\alpha\}$. Then Q is convex. The final statement of the theorem follows immediately from the strict convexity of the norms in l_p, $1<p<\infty$, for C^n.

Note that S may equal T. Thus the existence of globally optimal fits has been established. Moreover, once S has been selected, the values of admittances for the optimal solutions agree on the observation frequencies if an appropriate norm is chosen. In Ref. 4, p. 4426, Parker claims without proof that in this global setting Q will consist of a unique $b(\lambda)$, where $db(\lambda)$ is a positive finite delta comb. We find this result unlikely.

Theorem 2: Let $1 \leqslant p < \infty$, d: $[a,u] \to c$, $d \in L_p[a,u]$ (the data curve) be given, S be a convex subset of T, then we have the following.

(a) There exists a convex subset Q of T such that $c_0 \in Q$ implies

$$\|h(c) - d\|_p = \min_{c \in S} \|h(c) - d\|_p = \alpha,$$

where

$$h(c) = \chi_{[a,u]}(\omega) c(\omega)$$

and $\chi_{[a,u]}$ = characteristic function of $[a,u]$. We require members of Q to be the pointwise limit of a sequence from S.

(b) The optimal solution is unique if $1 < p < \infty$.

Proof: (a) Choose $c_n \in S \ni \|h(c_n) - d\|_p \to \alpha$. As before, $\{b_n\}$ is bounded and, hence, has a weak-* convergent subsequence $\{b_{n_k}\}$ with limit b^*. Because of (4), the uniform bound on $\|b_n(\lambda)\|$, and

$$c^*(\omega) = \lim_{k \to \infty} c_{b_{n_k}}(\omega),$$

we have

$$\|h(c^*) - \mathbf{d}\|_p = \alpha,$$

and the existence of optimal solutions has been established. Put $Q = \{c\colon \|h(c) - d\|_p = \alpha\}$. Then Q is clearly convex. Uniqueness of $c^*(\omega)$ in the case $1 < p < \infty$ follows again from strict convexity coupled with the analyticity of $c(\omega)$. The uniqueness of $b^*(\lambda)$ follows from Theorem 3.

Corollary: If $1 < p < \infty$, there exists a unique global optimal solution.

The previous theorem relies on the relationship of $c(\omega)$ to $b(\lambda)$ expressed in the next theorem.

Theorem 3: The admittance curves $c(\omega)$ and the associated $(b_0, b(\lambda))$ are related by the equations $m(\gamma) = \mathcal{S}[(\lambda^2 + 1)b(\lambda)]$ and $c(\omega) = m(i\omega) + (b_0 - \int_0^\infty \lambda\, db(\lambda))$, where $\mathcal{S}$ is the Stieljes transform. By assuming $b \in \mathrm{NBV}[0,\infty)$, $\mathcal{S}$ is uniquely invertible. A similar result is available with the Hilbert transform. [Remember that for our purposes in (4), $b_0 \geqslant \int_0^\infty \lambda\, b(\lambda)$, although that is not required in this theorem.]

Proof: See Sneddon,[6] p. 233. The uniqueness at discontinuities comes from the assumption that $b \in \mathrm{NBV}[0,\infty)$.

This section has been based on Parker's most general model setting (Ref. 4, pp. 4421 and 22) and contains the finite-dimensional subspaces of $b(\lambda)$ [where $db(\lambda)$ consists of finite positive delta combs] in which he performs his optimization calculations.[4,5] These correspond to delta comb conductivity models terminating in a perfect conductor basement. Several times Parker seems to imply that the optimal solutions are always in this limited class [for instance, "In addition to optimal models, which always consist of delta functions, two other types of model are examined." (Ref. 5, p. 9574)]. Consider data from a continuous conductivity model (σ_c, σ_0, h) and the corresponding b_c (db_c is not a finite delta comb.) Then (σ_c, σ_0, h) and b_c are clearly optimal in the finite data case for any finite set of observation frequencies W_N. We are not aware of any proof showing that the optimal set (b's or conductivity models) must contain any of Parker's delta comb σ's and corresponding b's. We suggest it is unreasonable to expect that result in either the pure data or noisy data cases, and see no proof to substantiate the claims in either Ref. 4 or Ref. 5. However, one can readily prove a weaker result that is sufficient to support all of Parker's numerical work in both the noise-free and noisy data cases: if α is the optimal 1-D fit to the data in a given metric and $\epsilon > 0$, then there exists one of Parker's delta comb conductivity models with a fit better than $\alpha + \epsilon$.

These theorems have yielded results relating to the existence and uniqueness of optimal admittance curves. The real geophysical interest, though, must lie in the implications of these results to physical models.

IV. RESULTS IN GEOLOGIC SPACE

The following theorem treats the existence of optimal data fits with typical restrictions on the geological model. Note that the common assumption that $1/r \geqslant \sigma(z) \geqslant 1/R$ is covered.

Theorem 4: Assume that discrete data $\mathbf{d}$ is available for $\omega \in W_N$. Select $f, F \in L_2(0,h)$ with $0 \leqslant f(z) \leqslant F(z)$. Let $V = \{\sigma(z) \in L_2[0,h]\colon f(z) \leqslant \sigma(z) \leqslant F(z)\}$. Using the l_p norm in C^N, $1 \leqslant p \leqslant \infty$, there exists an optimal solution $(\sigma(z), \sigma_0) \in V \times [m,M]$, where $m > 0$.

Proof: It should be noted that this theorem only establishes the existence of optimal solutions to the constrained optimization problem. We do not establish uniqueness.

Consider the set of all $c(\omega)$ which correspond to all $(\sigma, \sigma_0) \in V \times [m,M]$. Select a minimizing sequence $\{c_n(\omega)\}$ such that

$$\lim_{n \to \infty} \|g(c_n(\omega)) - \mathbf{d}\| = \min = \alpha.$$

Consider the sequence $\{(\sigma_n, \sigma_{0n})\}$. The sequence $\{\sigma_{0n}\}$ is bounded. Select a subsequence which converges to $\sigma_0 \neq 0$. From this subsequence of $\{(\sigma_n, \sigma_{0n})\}$, one may select another subsequence such that $\sigma_n \overset{\omega^*}{\to} \sigma$ in $L_2[0,h]$. The existence of such a σ is guaranteed by the theorem of Alaoglu, which proves the weak compactness of the unit ball in dual spaces.

Now $\sigma(x) \leqslant F(x)$ a.e. If not, assume $\sigma(x) \geqslant F(x) + \epsilon$, for $x \in A$, $\epsilon > 0$, $m(A) > 0$. Let χ_A be the characteristic function of A. Then

$$\int_0^h \chi_A(x)\sigma(x)dx \geqslant \epsilon m(A) + \int_0^h \chi_A(x)F(x)dx$$

$$\geqslant \epsilon m(A) + \int_0^h \chi_A(x)\sigma_m(x)dx,$$

for all n. Clearly, this precludes $\{\sigma_n\}$ from having a subsequence which converges weak-* to σ. Similarly, $\sigma(x) \geqslant f(x)$ a.e.

Let us write the base differential equation (1) as a pair of first-order equations:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ i\mu_0\omega\sigma(z) & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \tag{8}$$

where $0 \leqslant z \leqslant h$, $X_1(0,\omega) = 1$, $X_2(0,\omega) = \sqrt{i\mu_0\omega\sigma_0}$, where σ_0 is the basement conductivity.

Define

$$A(z) = \begin{pmatrix} 0 & 1 \\ i\mu_0\omega\sigma(z) & 0 \end{pmatrix}.$$

Then the transition matrix for (8) is

$$\Phi[0,h] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \int_0^h A(\lambda)d\lambda + \int_0^h A(\lambda)\int_0^\lambda A(\mu)d\mu\, d\lambda + \cdots .$$

Select $\omega > 0$. Since $\sigma_n \overset{\omega^*}{\to} \sigma$, it is straighforward to see that $\Phi_{n,\omega}[0,h] \to \Phi_\omega[0,h]$, where Φ_ω corresponds to (σ,σ_0), $\Phi_{n,\omega}$ to (σ_n,σ_{0n}). Thus $c_n(\omega) \to c(\omega)$, the admittance curve for (σ,σ_0). In particular, $\|\mathbf{d} - g(c(\omega))\| \leqslant \|\mathbf{d} - g(c_n(\omega))\|$ for all n. Thus the pair (σ,σ_0) is optimal

Corollary: The results remain valid if $V = \{\sigma(z) \in L_2[0,h]: \|\sigma\| \leqslant N\}$.

Proof: The proof is identical to that of Theorem 4, except that we must show $\|\sigma\| \leqslant N$. Assume $\|\sigma\| = N + \epsilon$, $\epsilon > 0$:

$$(N+\epsilon)^2 = \int_0^h \sigma_n(x)dx = \lim_{n\to\infty} \int_0^h \sigma(x)\sigma_n(x)dx \leqslant \lim_{n\to\infty} \|\sigma\|\,\|\sigma_n\| \leqslant (N+\epsilon)N.$$

Theorem 5: Assume that continuous data $d(\omega)$ is available for $\omega \in [a,u]$. Select f, F, and V as above. Then there exists an optimal solution $(\sigma(z),\sigma_0) \in V \times [m,M]$ with $m > 0$ in the sense that if $c(\omega)$ corresponds to $(\sigma(z),\sigma_0)$, then $\|c(\omega) - d(\omega)\|_p$ is a minimum.

Proof: Formulate the proof as in Theorem 4. The resulting (σ,σ_0) corresponds to $c(\omega)$, which is the pointwise limit of $\{c_n(\omega)\}$. By combining (4) and (5), one can prove that $c_n(\omega) \to c(\omega)$ uniformly on $[a,u]$,

$$\|c(\omega) - d(\omega)\|_p \leqslant \|c(\omega) - c_n(\omega)\|_p + \|c_n(\omega) - d(\omega)\|_p.$$

As $n \to \infty$,

$$\|c(\omega) - d(\omega)\|_p \leqslant 0 + \text{minimum} = \text{minimum}.$$

Theorems 4 and 5 yield nice results in that the constrained geological problems do possess optimal solutions. However, it is unfortunate that uniqueness need not follow.

In attempts to prove the invertibility of the 1-D MT problem (Theorem 7), a related result of interest was proved.

Theorem 6: Assume $(\sigma_1,\sigma_{01}, h_1)$ and $(\sigma_2,\sigma_{02},h_2)$ yield the same admittance curve. Then $\sigma_{01} = \sigma_{02}$ and the zeroth, first, and second moments of $\sigma_1(z)$ and $\sigma_2(z)$ are equal on $[0,h]$, where $h = \max\{h_1,h_2\}$ and the functions are aligned to put $z = h$ at the Earth's surface.

Proof: This result can be proved by writing formulas for $c_1(\omega)$ and $c_2(\omega)$ using the transition matrices for the two systems. Using expansions of both in terms of ω yields the result.

The next theorem discusses f_1^{-1} as in Fig. 1.

Theorem 7: Let $c(\omega)$ be generated by $(\sigma(z),\sigma_0,h)$, where $\sigma(z) \in C^\infty[0,h]$, $\sigma(z) \geqslant 0$, $\sigma(h) > 0$, and σ is the restriction of an analytic function to $[0,h]$, $\sigma_0 > 0$, $h > 0$. Then, within this class, $(\sigma(z),\sigma_0, h)$ can be recovered from $c(\omega)$ when h is minimal.

Proof: σ_0 can readily be obtained since $\sigma_0^{-1} = \lim_{\omega\to 0} i\mu_0\omega c^2(\omega)$. The remainder of the proof is based on an argument suggested by Bailey for a related problem on a stratified sphere. This concept has been adapted to the MT problem for a flat earth, and the final arguments have been strengthened to permit weaker hypotheses (both here and in Bailey): we do not assume $\sigma(z) \geqslant m > 0$, but only $\sigma(h) > 0$, $\sigma_0 > 0$, and $c(z,\omega)$ is analytic in the lower half of the complex ω plane. When $|\omega|$ is large and $\sigma(z) > 0$, one can argue that

$$\lim_{|\omega|\to\infty} c'(z,\omega) = \lim_{|\omega|\to\infty} (1 - i\mu_0\omega\sigma(z)c^2(z,\omega)) = 0$$

with the error behaving as $\gamma/\sqrt{\omega}$. This results from the aysmptotic behavior of $c(z,\omega)$ for large ω, which is

$$c(z,\omega) = \frac{1}{k} + \frac{1}{4}\frac{\sigma'(z)}{\sigma(z)k^2} + O(k^{-3}),$$

where

$$k = \sqrt{i\omega\mu_0\sigma(z)}. \tag{9}$$

Formula (9) can be derived using the WKB approximation for E and $\partial E/\partial z$. Select the contour C. By the Cauchy integral formula (for ω in the lower half-plane, and using the contour in Fig. 2)

$$\frac{\partial c}{\partial z}(z,\omega) = \frac{1}{2\pi i}\int_c \frac{1}{\omega' - \omega}\frac{\partial c(z,\omega)}{\partial z}\,d\omega'.$$

The portion of the integral over the curved bottom of C goes to zero as $R \to \infty$. Thus,

$$\frac{\partial c}{\partial z}(z,\omega) = \frac{-1}{2\pi i}\int_{-\infty}^{\infty} \frac{1}{\omega' - \omega}\frac{\partial c(z,\omega')}{\partial z}\,d\omega'. \tag{10}$$

Let us analyze what happens as $\omega \to 0$. Using Parker[4],

$$\frac{\partial c}{\partial z}(z,0) = 1 - \frac{\sigma(z)}{\sigma_0},$$

in the limit. Allowing $\omega \to 0$ in the right side of (10) yields

$$1 - \frac{\sigma(z)}{\sigma_0} = \frac{-1}{\pi i}\oint_{-\infty}^{\infty} \frac{1}{\omega'}\frac{\partial c(z,\omega')}{\partial z}\,d\omega',$$

where $\oint_{-\infty}^{\infty}$ indicates $\lim_{\epsilon\to\infty} \int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}$, the Cauchy principal value. Now let us integrate the modified differential equation

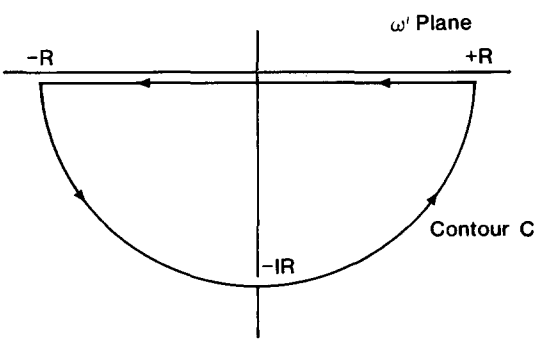


FIG. 2. Contour of integration.

$$\frac{1}{i\mu_0\omega'}\frac{\partial c(z,\omega')}{\partial z} = \frac{1}{i\mu_0\omega'} - \sigma(z)c^2(z,\omega'),$$

$$\oint_{-\infty}^{\infty} \frac{1}{i\mu_0\omega'}\frac{\partial c(z,\omega')}{\partial z}\,d\omega' = \oint_{-\infty}^{\infty} \frac{d\omega'}{i\mu_0\omega'} - \int_{-\infty}^{\infty} \sigma(z)c^2(z,\omega')d\omega'.$$

The second integral is zero. Substituting for the first yields

$$\frac{\pi}{\mu_0}\left(1 - \frac{\sigma(z)}{\sigma_0}\right) = \oint_{-\infty}^{\infty} \sigma(z)c^2(z,\omega)d\omega \tag{11}$$

or

$$\sigma(z) = \frac{\pi/\mu_0}{\pi/\mu_0\sigma_0 + \oint_{-\infty}^{\infty} c^2(z,\omega)dw},$$

$$\rho(z) - \rho_0 = \frac{\mu_0}{2\pi}\oint_{-\infty}^{\infty} c_2(z,\omega)d\omega \quad \left(\rho(z) = \frac{1}{\sigma(z)}\right).$$

Since $c(z,-\omega^*) = c^*(z,\omega)$, (11) may be rewritten

$$\sigma(z) = \frac{\pi/\mu_0}{\pi/\mu_0\sigma_0 + 2\int_0^{\infty}\mathrm{Re}[c^2(z,\omega)]d\omega}. \tag{12}$$

[Equation (12) yields $\lim_{\epsilon\to\infty}\sigma(z-\epsilon)$ when σ is not continuous.]

Using (12), one may obtain values for $\sigma^{(n)}$ at the surface, $n = 0,1,2,\ldots$. Moreover, due to (7), $\sigma^{(n)}$ can be uniquely determined from $c(\cdot,\omega)$ at the surface ($\omega > 0$). The C^∞ portion of $\sigma(z)$ may now be reconstructed uniquely by power series, $S(z)$. Using $S(z)$, $c(z,\omega)$ can be obtained from (7) starting at the surface. This $c(z,\omega)$ can then be plugged into (12) to obtain $\lim_{\epsilon\to\infty}\sigma(z-\epsilon)$. Then h is recovered as the depth where $S(z)$ separates from $\lim_{\epsilon\to\infty}\sigma(z-\epsilon)$. At this depth, $c(\cdot,\omega) = 1/\sqrt{i\mu_0\omega\sigma_0}$.

Consider a partition $\{a_j\}$ of $[0,h]$ such that $0 = a_0 < a_1 < \cdots < a_n = h$. Assume that when $a_{j-1} < x \leqslant a_j$, $\sigma(x) = f_j(x) \geqslant 0$, where $f_j \in C^\infty[a_{j-1},a_j]$, $f_j(a_j) > 0$, and f_j is the restriction of an analytic function to $[a_{j-1},a_j]$. Let Y_h be the class of all such functions σ. Then Y_h includes all the layered earth models, all piecewise polynomials, and all piecewise finite trigonometric-series functions.

Theorem 8: In both the flat earth and spherical earth cases, $c(\omega)$ can be inverted uniquely over the class Y_h, $h > 0$.

Proof: This is almost a trivial but significant extension of Theorem 7 and of Bailey,[1] respectively. The $n+1$ layers can uniquely be reconstructed by n applications of the procedure in Theorem 7.

Y_h is a very broad class of expected geological models. It would be nice to extend the result to the positive cone in $L^2[0,h]$, but all attempts have failed. Nonetheless, Y_h is a sufficiently broad class that Theorem 8 makes it reasonable to expect a unimodal distribution of solutions to the finitely sampled noisy data inversion problem.

V. APPLICATIONS

The theorems stated earlier can be applied to a variety of the iterative inversion techniques in use today. As an example, the techniques of Parker[4,5] fit the general framework of Theorem 1. He selects $b(\lambda)$ from ever-expanding subspaces in NBV$[0,\infty)$ which consist of finite delta combs. In the limit when the functionals are restricted to $f(\lambda,\omega) = (1 - i\omega\lambda)/(\lambda + i\omega)$, where $\omega \in W_N$, these subspaces are weak*-dense in the positive cone of NBV$[0,\infty)$. If one studies the set $\{f(\lambda,\omega): \omega \in W_N\}$, one observes that most of the "action" takes place over a rather small λ range. Thus one would expect the rapid convergence to a small χ^2 value experienced by Parker[4,5] and Sabatier.[3]

[1] R. C. Bailey, "Inversion of the geomagnetic induction problem," Proc. R. Soc. London Ser. A **315**, 185 (1970).

[2] P. Weidelt, "The inverse problem of geomagnetic induction," Zeitschr. Geophys. **38**, 257 (1972).

[3] P. C. Sabatier, "Positivity constraints in linear inverse problems I and II," Geophys. J. R. Astron. Soc. **48**, 415 (1977).

[4] R. Parker, "The inverse problem of electromagnetic induction: Existence and construction of solutions based on incomplete data," J. Geophys. Res. **85**, 4421 (1980).

[5] R. Parker and K. A. Whaler, "Numerical methods for establishing solutions to the inverse problem of electromagnetic induction," J. Geophys. Res. **86**, 9574 (1981).

[6] I. H. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).

Perfect fluid models of Bianchi type-VI_0 in modified Brans–Dicke cosmology

Shri Ram
Applied Mathematics Section, Institute of Technology, Banaras Hindu University, Varanasi 221005, India

(Received 25 June 1985; accepted for publication 16 October 1985)

The spatially homogeneous and anisotropic Bianchi type-VI_0 cosmological solution for barotropic fluid in the context of modified Brans–Dicke theory is obtained. The general behavior of such an anisotropic homogeneous model has been discussed.

I. INTRODUCTION

The Brans–Dicke (BD) theory[1] was proposed more than 20 years ago with the aim of incorporating Mach's principle into general relativity. An enormous effort has been made since then in exploring the gravitational as well as cosmological consequences of this scalar–tensor theory. The role of the cosmological constant Λ has been discussed by various authors in agreement with modern ideas of fundamental particle interactions. Bergmann[2] and Wagoner[3] have suggested that the cosmological term should be a function of a scalar function. Following the work of Endo and Fukui,[4] Banerjee and Santos[5] have obtained cosmological dust solutions for a Bianchi type-I homogeneous space in the modified Brans–Dicke cosmology. Recently spatially homogeneous and anisotropic Bianchi type-I cosmological solutions of the modified BD theory containing barotropic fluid have been obtained by Singh and Singh[6] by the condition of the cosmological term $\Lambda(\phi)$. Since Bianchi type-I models are a very special set of spatially homogeneous models, the present author has presented anisotropic homogeneous Bianchi type-VI_0 cosmological dust solutions in the context of this theory.[7] However, Lorentz[8] has presented Kasner-like perfect fluid solutions of the BD theory of gravitation for the Bianchi type-VI_0 geometry.

In this paper we obtain spatially homogeneous and anisotropic Bianchi type-VI_0 cosmological solutions of modified BD theory containing barotropic fluid. The possibilities of dust-filled universes, radiation-dominated universes, and superdense universes are explored. The general behavior of such anisotropic homogeneous models is discussed.

II. FIELD EQUATIONS

The field equations for the modified BD theory with the introduction of $\Lambda(\phi)$ are[4]

$$G_{\mu\nu}+g_{\mu\nu}\Lambda=-(k/\phi)T_{\mu\nu}-(\omega/\phi^2)(\phi_{,\mu}\phi_{,\nu}-\tfrac{1}{2}g_{\mu\nu}\phi_{,a}\phi^{,a})-(1/\phi)(\phi_{;\mu\nu}-g_{\mu\nu}\Box\phi), \tag{1}$$

$$\phi\frac{\partial\Lambda}{\partial\phi}-\Lambda=\frac{k}{2\phi}T-\frac{2\omega+3}{2\phi}\Box\phi. \tag{2}$$

Here $T_{\mu\nu}$ is the energy-momentum tensor for a perfect fluid given by

$$T_{\mu\nu}=(\rho+p)V_\mu V_\nu-pg_{\mu\nu}, \tag{3}$$

where p and ρ are proper pressure and energy density, respectively, and V^μ is the four-velocity of the fluid. We assume the coordinates to be comoving so that

$$V^1=V^2=V^3=0, \quad V^4=1. \tag{4}$$

We number the coordinates x, y, z, and t as 1, 2, 3, and 4, respectively. Here a comma and a semicolon denote ordinary differentiation and covariant differentiation, respectively.

We further assume that the matter and the scalar fields are related through[5]

$$\Box\phi=k\mu T/(2\omega+3), \tag{5}$$

where the constant μ shows the deviation of this theory from that of BD theory and ω is the coupling constant. Substitution of (5) in (2) yields

$$\Lambda-\phi\frac{\partial\Lambda}{\partial\phi}=a\frac{\Box\phi}{\phi}, \tag{6}$$

a being a constant defined by

$$a=[(2\omega+3)/2]((1/\mu)-1). \tag{7}$$

If Λ is a function of ϕ only, Banerjee and Santos[5] have assumed that $\Box\phi=m\phi^n$, where m and n are arbitrary constants. Substituting for $\Box\phi$ in (6) and integrating the resulting equation, we obtain

$$\Lambda=[am/(2-n)]\,\phi^{n-1}+D_1\phi, \quad n\neq 2, \tag{8}$$

and

$$\Lambda=-am\log\phi+D_2\phi, \quad n=2, \tag{9}$$

D_1 and D_2 being integration constants.

III. SOLUTIONS OF FIELD EQUATIONS

The line element for the spatially homogeneous Bianchi-type VI_0 can be written as

$$ds^2=dt^2+A^2(t)\,dx^2+B^2(t)e^{-2qx}\,dy^2+C^2(t)e^{2qx}\,dz^2, \tag{10}$$

where A, B, and C are cosmic scale functions, and q is a nonzero constant. The nonzero components of the field equations for (10) are

$$\frac{B_{44}}{B}+\frac{C_{44}}{C}+\frac{B_4C_4}{BC}+\frac{q^2}{A^2}=-\Lambda-\frac{kp}{\phi}-\frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right)^2+\frac{A_4\phi_4}{A\phi}+\frac{\Box\phi}{\phi}, \tag{11}$$

$$\frac{A_{44}}{A}+\frac{C_{44}}{C}+\frac{A_4C_4}{AC}-\frac{q^2}{A^2}=-\Lambda-\frac{kp}{\phi}-\frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right)^2+\frac{B_4\phi_4}{B\phi}+\frac{\Box\phi}{\phi}, \tag{12}$$

 0022-2488/86/020650-03$02.50 © 1986 American Institute of Physics

$$\frac{A_{44}}{A}+\frac{B_{44}}{B}+\frac{A_4B_4}{AB}-\frac{q^2}{A^2}$$
$$=-\Lambda-\frac{kp}{\phi}-\frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right)^2+\frac{C_4\phi_4}{C\phi}+\frac{\Box\phi}{\phi}, \tag{13}$$

$$\frac{A_4B_4}{AB}+\frac{B_4C_4}{BC}+\frac{A_4C_4}{AC}-\frac{q^2}{A^2}$$
$$=-\Lambda+\frac{kp}{\phi}+\frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right)^2+\frac{\phi_{44}}{\phi}+\frac{\Box\phi}{\phi}, \tag{14}$$

$$\frac{B_4}{B}-\frac{C_4}{C}=0, \tag{15}$$

where the subscript 4 denotes ordinary differentiation with respect to t.

From the conservation equation $T^{\mu}_{\nu;\mu}=0$ we obtain

$$\rho_4+(\rho+p)\left(\frac{A_4}{A}+\frac{B_4}{B}+\frac{C_4}{C}\right)=0. \tag{16}$$

Equation (15) readily gives $B=\gamma C$, γ being an integration constant. Without loss of any generality we take $\gamma=1$.

We now consider the equation of state

$$p=(\lambda-1)\rho, \quad 1<\lambda<2. \tag{17}$$

Then from (16) we obtain

$$\rho=c/(AB^2)^{\lambda}, \tag{18}$$

where c is a constant of integration. Also we have

$$T=3p-\rho=(3\lambda-4)\rho. \tag{19}$$

Equations (5), (6), (18), and (19) give

$$1/(AB^2)^{\lambda}=-(m/d)\phi^n, \tag{20}$$

where

$$d=k\mu c(4-3\lambda)/(2\omega+3).$$

We observe that then we can write the density (18) in terms of the scalar field

$$\rho=-(mc/d)\phi^n. \tag{21}$$

In order to treat Eqs. (11)–(14) we introduce new variables α, β, and τ by

$$A=e^{\lambda\alpha}, \quad \beta=e^{\lambda\beta}, \quad dt=AB^2\,d\tau, \tag{22}$$

and differentiation with respect to τ is denoted by a dash. Then Eq. (20) gives

$$e^{\lambda(\alpha+2\beta)}=-(d/m)\phi^{-n}. \tag{23}$$

Making use of (22) and (23) we can express $\Box\phi=m\phi^n$ by

$$\phi''=-\frac{d^2}{m}\phi^{-n}. \tag{24}$$

Substituting (8), and (22)–(24) into Eqs. (11)–(14) we obtain the field equations as

$$2\beta''+3\lambda\beta'^2+(2n+2)\beta'(\phi'/\phi)+(q^2/\lambda)e^{4\lambda\beta}$$
$$=\left[-\frac{ad^2}{\lambda m(2-n)}+\frac{k(\lambda-1)cd}{\lambda m}+\frac{d^2}{\lambda m}\right]\bar{\phi}^{(n+1)}$$
$$-\frac{D_1d^2}{\lambda m^2}\bar{\phi}^{(2n-1)}-\frac{1}{\lambda}\left(\frac{\omega}{2}+n\right)\left(\frac{\phi'}{\phi}\right)^2, \tag{25}$$

$$-\beta''+3\lambda\beta'^2+(2n-1)\beta'\frac{\phi'}{\phi}-\frac{q^2}{\lambda}e^{4\lambda\beta}$$
$$=\left[-\frac{ad^2}{\lambda m(2-n)}+\frac{k(\lambda-1)cd}{\lambda m}\right.$$
$$\left.+\frac{d^2}{\lambda m}-\frac{nd^2}{\lambda m}\right]\bar{\phi}^{(n+1)}$$
$$-\frac{D_1d^2}{\lambda m^2}\bar{\phi}^{(2n-1)}-\frac{1}{\lambda}\left(\frac{\omega}{2}+n\right)\left(\frac{\phi'}{\phi}\right)^2, \tag{26}$$

$$3\lambda\beta'^2+2n\beta'\frac{\phi'}{\phi}+\frac{q^2}{\lambda}e^{4\lambda\beta}$$
$$=\left[\frac{ad^2}{\lambda m(2-n)}+\frac{kcd}{\lambda m}\right]\bar{\phi}^{(n+1)}$$
$$+\frac{D_1d^2}{\lambda m^2}\bar{\phi}^{(2n-1)}-\frac{1}{\lambda}\left(\frac{\omega}{2}+n\right)\left(\frac{\phi'}{\phi}\right)^2. \tag{27}$$

Subtracting (25) from (24), we obtain

$$3\beta''+3\beta'\frac{\phi'}{\phi}+\frac{2q^2}{\lambda}e^{4\lambda\beta}=\frac{nd^2}{\lambda m}\bar{\phi}^{(n+1)}. \tag{28}$$

Elimination of β'' from (24) and (25) yields

$$3\lambda\beta'^2+2n\beta'\frac{\phi'}{\phi}-\frac{q^2}{3\lambda}e^{4\lambda\beta}$$
$$=\left[-\frac{ad^2}{\lambda m(2-n)}+\frac{k(\lambda-1)cd}{\lambda m}\right.$$
$$\left.+\frac{d^2}{\lambda m}-\frac{2}{3}\frac{nd^2}{\lambda m}\right]\bar{\phi}^{(n+1)}$$
$$-\frac{D_1d^2}{\lambda m^2}\bar{\phi}^{(2n-1)}-\frac{1}{\lambda}\left(\frac{\omega}{2}+n\right)\left(\frac{\phi'}{\phi}\right)^2. \tag{29}$$

From (27) and (29) we find that

$$q^2e^{4\lambda\beta}=P\bar{\phi}^{(n+1)}+\frac{3}{2}\frac{D_1d^2}{m^2}\bar{\phi}^{(2n-1)}, \tag{30}$$

where

$$P=\frac{3}{4}\left[\frac{2ad^2}{m(2-n)}+\frac{kcd}{m}-\frac{d^2}{m}\right.$$
$$\left.+\frac{2}{3}\frac{nd^2}{m}-\frac{k(\lambda-1)cd}{m}\right].$$

Due to the nonlinearity of the field equations, it is very difficult to obtain a solution in its generality and, therefore, we have to make some simplifying assumptions to derive useful results. The assumptions are motivated either by physical considerations or by mathematical convenience. We consider the case $D_1=0$. Then (30) reduces to

$$q^2e^{4\lambda\beta}=P\bar{\phi}^{(n+1)}. \tag{31}$$

Differentiating (31) we get

$$\beta'=-\frac{n+1}{4\lambda}\left(\frac{\phi'}{\phi}\right), \tag{32}$$

$$\beta''=\frac{n+1}{4\lambda}\left(\frac{\phi'}{\phi}\right)^2+\frac{(n+1)d^2}{4\lambda m}\phi^{-(n+1)}. \tag{33}$$

Equation (28) is satisfied if

$$8\lambda mP=(n-3)d^2. \tag{34}$$

Substituting (31)–(33) into (29), we find that

$$\left(\frac{\phi'}{\phi}\right)^2 = \frac{\lambda^2}{Q^2}\phi^{-(n+1)}, \tag{35}$$

where

$$\frac{1}{Q^2} = 16\left[\frac{-ad^2}{2m(2-n)} + \frac{3k(\lambda-1)cd}{4m} + \frac{3}{4}\frac{d^2}{m} + \frac{kcd}{4m} - \frac{nd^2}{2m}\right] \times [\lambda^2(3+8\omega+14n-5n^2)]^{-1}. \tag{36}$$

Integrating (35), we obtain

$$\phi = \left[\frac{(n+1)(\lambda\tau+E)}{2Q}\right]^{2/(n+1)}, \tag{37}$$

where E is an integration constant and may be chosen such that at $\tau = 0$ one has $\phi = 0$. From (23) and (31) we finally have the solutions for α and β as

$$e^{4\lambda\beta} = \frac{P}{q^2}\left[\frac{(n+1)(\lambda\tau+E)}{2Q}\right]^{-2}, \tag{38}$$

$$e^{2\lambda\alpha} = \frac{1}{P}\left(\frac{qd}{m}\right)^2\left[\frac{(n+1)(\lambda\tau+E)}{2Q}\right]^{-2(n-1)/n+1}. \tag{39}$$

Also, α and β can be expressed in terms of the scalar function ϕ as

$$e^{2\lambda\alpha} = \frac{1}{P}\left(\frac{qd}{m}\right)^2\phi^{-(n-1)}, \quad e^{4\lambda\beta} = \frac{P}{q^2}\phi^{-(n+1)}. \tag{40}$$

The cosmological factor given (8) with $D_1 = 0$ is

$$\Lambda = [am/(2-n)]\,\phi^{n-1}. \tag{41}$$

The density and pressure can be written as a function of the cosmological factor

$$\rho = -(mc/d)[[(2-n)/am]\Lambda]^{n/(n-1)} \tag{42}$$

and

$$p = -(mc/d)(\lambda-1)[[(2-n)/am]\Lambda]^{n/(n-1)}. \tag{43}$$

We now consider the cases of the dust-filled, radiation-dominated, and superdense stages of the universe.

Case (I) Dust-filled universe: This case corresponds to the distribution of incoherent matter for which $p = 0$. Putting $\lambda = 1$ in (37)–(43) we get the model for incoherent matter. This case has already been discussed by the present author.[7]

Case (II) Radiation-dominated universe: Putting $\lambda = \frac{4}{3}$ in (17) we get $\rho = 3p$. In this case of radiation dominated universe we obtain, from (19), that $T = 0$. Then, from (5), we have $\Box\phi = 0$, which implies that $m = 0$. Therefore (42) and (43) give $\rho = 0$ and $p = 0$. Thus for $\rho = 3p$ no solution will exist and the model reduces to the vacuum one.

Case (III) Superdense universe: Choosing $\lambda = 2$ the equation of state (17) becomes $\rho = p$. This equation of state for stiff matter has been widely used in general relativity to obtain stellar and cosmological models for ultradense matter.[9] Therefore, inserting $\lambda = 2$ in (37)–(43) we arrive at the model for ultradense matter in modified BD theory.

Finally observe that for $q = 0$ the line element (10) reduces to a Bianchi type-I metric and consequently we get plane symmetric Bianchi type-I models for a perfect fluid in modified BD theory.

IV. CONCLUSIONS

In Sec. III we obtained a perfect fluid solution of Bianchi type-VI_0 in modified BD theory. The general behavior of such an anisotropic model is analogous to that of the zero-curvature Robertson–Walker model of Endo and Fukui,[4] the Bianchi type-I perfect fluid model of Singh and Singh,[6] the Bianchi type-VI_0 dust model of Ram,[7] and the Bianchi type-VI_0 perfect fluid model of Lorentz[8] in BD theory. From (37), we observe that ϕ is an increasing function of τ, if $n < 0$, and a decreasing function for $n > 0$. The later case is of no physical interest as it contradicts the choice of initial condition. For an expanding universe the spatial volume increases with time for negative values of n; in which case $V \to 0$ and $\rho \to \infty$ at the epoch $\phi \to 0$. In course of time the model expands and attains infinite volume $V \to \infty$ and $\rho \to 0$ as $\phi \to \infty$. Solutions obtained in this paper are of considerable interest and may be useful to study the large-scale dynamics of the physical universe.

ACKNOWLEDGMENT

The author is extremely thankful to Professor K. P. Singh for encouragement and useful discussions.

[1] C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
[2] P. G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968).
[3] R. V. Wagoner, Phys. Rev. D **1**, 3209 (1970).
[4] M. Endo and T. Fukui, Gen. Relativ. Gravit. **8**, 833 (1977).
[5] A. Banerjee and N. O. Santos, Gen. Relativ. Gravit. **14**, 559 (1982).
[6] T. Singh and T. Singh, J. Math. Phys. **25**, 2800 (1984).
[7] S. Ram, J. Math. Phys. **26**, 2915 (1985).
[8] D. Lorentz-Petzold, Acta Phys. Austriaca **55**, 209 (1984).
[9] Ya. B. Zeldovich, Sov. Phys. JETP **14**, 1143 (1962).