Tensor product of quantum logics

Sylvia Pulmannová
Mathematical Institute of the Slovak Academy of Sciences, 814 73 Bratislava, Czechoslovakia

(Received 23 March 1983; accepted for publication 11 May 1984)

A quantum logic is the couple (L,M) where L is an orthomodular σ-lattice and M is a strong set of states on L. The Jauch–Piron property in the σ-form is also supposed for any state of M. A "tensor product" of quantum logics is defined. This definition is compared with the definition of a free orthodistributive product of orthomodular σ-lattices. The existence and uniqueness of the tensor product in special cases of Hilbert space quantum logics and one quantum and one classical logic are studied.

I. INTRODUCTION

By a quantum logic we mean the couple (L,M), where L is an orthomodular σ-lattice (a logic) and M is a set of states on L, which is strong for L, i.e., the statement

$$\{m\in M:m(a)=1\}\subset\{m\in M:m(b)=1\}$$

implies that $a\leqslant b$, $a,b\in L$. We also suppose that the Jauch–Piron property holds in the σ-form, i.e., that $m(a_i)=1$, $a_i\in L$, $i=1,2,\ldots$ if and only if

$$m\Big(\bigwedge_{i=1}^{\infty} a_i\Big)=1.$$

The free orthodistributive product of the orthomodular σ-lattices was defined as follows.[1]

Definition 1: Let $\mathscr{C}$ be a subcategory of the category of orthomodular σ-lattices. Assume L_i $(i\in I)$ and L are objects of $\mathscr{C}$. Then $(L,(u_i)_{i\in I})$ is a tensor product (or free orthodistributive product) of the L_i's if (i) $u_i:L_i\to L$ are injections in $\mathscr{C}$ $(i\in I)$; (ii) $\cup_{i\in I}u_i(L_i)$ generates L; (iii) for every at most countable subset F of I, $\wedge_{i\in F}u_i(a_i)=0$ for $a_i\in L_i$ if and only if at least one a_i is zero; (iv) $u_i(a_i)\leftrightarrow u_j(a_j)$ [i.e., $u_i(a_i)$ is compatible with $u_j(a_j)$] for any $i,j\in I$, $i\neq j$.

In the previous paper,[2] tensor product of quantum logics was defined as follows.

Definition 2: Let (L_1,M_1), (L_2,M_2), and (L,M) be quantum logics. We say that $(L,M)_{\alpha,\beta}$ is a tensor product of (L_i, M_i), $i=1,2$ if there are mappings α,β such that (i) $\alpha:L_1\times L_2\to L$, $\beta:M_1\times M_2\to M$, $\beta(m_1,m_2)(\alpha(a_1,a_2))=m_1(a_1)\times m_2(a_2)$; (ii) $\beta[M_1\times M_2]=\{\beta(m_1,m_2):m_1\in M_1,\ m_2\in M_2\}$ is strong for L; (iii) L is generated by $\alpha[L_1\times L_2]$.

Without any loss of generality we may put $\beta[M_1\times M_2]=M$. To make the difference, we shall call the product of orthomodular σ-lattices by Definition 1 the "free" product and the product of quantum logics the "tensor" product. In this paper, we shall study these two products. We shall show that the tensor product of complete quantum logics in the sense of Definition 2 can exist only if at least one of the logics L_1 and L_2 is a classical one (i.e., a Boolean algebra). Then we shall introduce another definition of the tensor product of quantum logics, in which the property (ii) in Definition 2 is weakened. This latter definition also includes the tensor product of Hilbert space logics.

II. PROPERTIES OF THE FREE PRODUCT

At first, we shall study the properties of the free products in the category of complete atomistic orthomodular lattices.

Lemma 1: Let L_1,L_2,L be complete atomistic orthomodular lattices and let (L,u_1,u_2) be the free product of L_1 and L_2. Then

$$u_1(a_1)\wedge u_2(a_2)\perp u_1(b_1)\wedge u_2(b_2)$$

iff $a_1\perp b_1$ or $a_2\perp b_2$, where $a_1,b_1\in L_1$ and $a_2,b_2\in L_2$ are atoms.

Proof: Let us consider the set $K=\{u_1(a_1),u_1(b_1)^\perp,u_2(a_2),u_2(b_2)^\perp\}$. Among any three of the elements of K there is one which is compatible with the other two, so that K is a distributive set.[3] Then from

$$u_1(a_2)\wedge u_2(a_2)\leqslant[u_1(b_1)\wedge u_2(b_2)]^\perp=u_1(b_1)^\perp\vee u_2(b_2)^\perp$$

we get

$$\begin{aligned}u_1(a_1)\wedge u_2(a_2)&\leqslant(u_1(b_1)^\perp\vee u_2(b_2)^\perp)\wedge u_2(a_2)\\&=u_1(b_1)^\perp\wedge u_2(a_2)\vee u_2(a_2)\wedge u_2(b_2)^\perp\\&=u_1(b_1)^\perp\wedge u_2(a_2)\vee u_2(a_2\wedge b_2^\perp),\end{aligned}$$

but $a_2\wedge b_2^\perp=0$ or a_2, as a_2 is an atom. If $a_2\wedge b_2^\perp=a_2$, we have $a_2\leqslant b_2^\perp$. If $b_2^\perp\wedge a_2=0$, then

$$u_1(a_1)\wedge u_2(a_2)\leqslant u_1(b_1^\perp)\wedge u_2(a_2),$$

i.e., $u_{1,a_2}(a_1)\leqslant u_{1,a_2}(b_1^\perp)$, which implies $a_1\leqslant b_1^\perp$, because the map

$$u_{1,a_2}:L_1\to L$$
$$a\mapsto u_1(a)\wedge u_2(a_2)$$

is injective.[4]

Lemma 2: Let L_1, L_2, L be complete atomistic orthomodular lattices and (L,u_1,u_2) be a free product of L_1 and L_2. Let A_1,A_2,A be the sets of all atoms in L_1, L_2, L, respectively. If (i) $u_1(a)\wedge u_2(b)\in A$ for any $a\in A_1$ and $b\in A_2$; (ii) all atoms of A lying under $u_2(a_2)$, for any $a_2\in A_2$, are of the form $u_1(b)\wedge u_2(a_2)$ for some $b\in A_1$; then the maps

$$u_{1,a_2}:L_1\to L_{[0,u_2(a_2)]}$$
$$a\mapsto u_1(a)\wedge u_2(a_2)$$

are surjective for any $a_2\in A_2$.

Proof: Let $c\leqslant u_2(a_2)$, $a_2\in A_2$. Then there is $B\subset A_1$ such that

 0022-2488/85/010001-05$02.50

$$c=\bigvee_{b\in B}u_1(b)\wedge u_2(a_2)=\Big(\bigvee_{b\in B}u_1(b)\Big)\wedge u_2(a_2)$$
$$=u_1\Big(\bigvee_{b\in B}b\Big)\wedge u_2(a_2)=u_{1,a_2}\Big(\bigvee_{b\in B}b\Big).$$

Q.E.D.

Theorem 1: Let L_1, L_2, L be complete atomistic orthomodular lattices and (L, u_1, u_2) be the free product of L_1 and L_2. If (i) and (ii) of Lemma 2 are satisfied, then L_1 and L_2 are irreducible if and only if L is irreducible.

Proof: By Ref. 5, if L is irreducible then L_1 and L_2 are irreducible. Let c be an element in the center of L different from 0 and 1. We show that there must be an atom b_2 in L_2 such that $c\wedge u_2(b_2)\neq 0$, $u_2(b_2)$. Indeed, let there be for any $b\in A_2$

$$c\wedge u_2(b)=u_2(b)\quad\text{or}\quad c\wedge u_2(b)=0.$$

Then $u_2(b)\leqslant c$ or $u_2(b)\leqslant c^\perp$ for any $b\in A_2$. Put $q_1=\bigvee\{b\in A_2:u_2(b)\leqslant c\}$, $q_2=\bigvee\{b:u_2(b)\leqslant c^\perp\}$. If $b_1\leqslant q_1$ and $b_2\leqslant q_2$, then $u_2(b_1)\leqslant c$ and $u_2(b_2)\leqslant c^\perp$, which implies that $u_2(b_1)\perp u_2(b_2)$ and, as u_2 is an injection, this implies that $b_1\perp b_2$. From this it follows that $q_1\perp q_2$. For any $b\in A_2$ we have $b<q_1$ or $b\perp q_1$, i.e., $q_1\in L_2'$, where L_2' is the center of L_2. This is in contradiction with the irreducibility of L_2 unless $q_1=1$ or $q_1=0$. If $q_1=1$, then $1=u_2(q_1)=\bigvee_{b<q_1}u_2(b)\leqslant c$, which contradicts the supposition. If $q_1=0$, then $q_2=1$ and we get a contradiction in a similar way. Now we can use Corollary 3 (Ref. 6) to show that L is irreducible.

We say that the elements b,c in a complete atomistic orthomodular lattice are separated by a superselection rule[7] if for any atom $a\leqslant b\vee c$ we have $a\leqslant b$ or $a\leqslant c$.

Lemma 3: Let (L, u_1, u_2) be a free orthodistributive product of L_1 and L_2 in the category of complete atomistic lattices. Let A_1, A_2, A be the sets of all atoms in L_1, L_2, and L, respectively. If $A=\{u_1(a_1)\wedge u_2(a_2):a_1\in A_1,\ a_2\in A_2\}$, then $u_1(a_1)\wedge u_2(a_2)$ and $u_1(b_1)\wedge u_2(b_2)$, where $a_1, b_1\in A_1$, $a_2, b_2\in A_2$, $a_1\neq b_1$, $a_2\neq b_2$, are separated by a superselection rule.

Proof: Let $u_1(c_1)\wedge u_2(c_2)\leqslant u_1(a_1)\wedge u_2(a_2)\vee u_1(b_1)\wedge u_2(b_2)$, where $c_1\in A_1$, $c_2\in A_2$. As $\{u_1(a_1), u_2(b_2), u_2(c_2)\}$ is a distributive set, we have under supposition that $c_2\neq a_2,b_2$,

$$u_1(c_1)\wedge u_2(c_2)\leqslant(u_1(a_1)\vee u_2(b_2))\wedge u_2(c_2)$$
$$=u_2(c_2)\wedge u_1(a_1)\vee u_2(b_2\wedge c_2)=u_1(a_1)\wedge u_2(c_2),$$

which implies that $a_1=c_1$. Similarly we obtain that $c_1=b_1$, a contradiction. Q.E.D.

Lemma 4: If a, b are two different atoms in a complete atomistic orthomodular lattice L which are separated by a superselection rule, then $a\perp b$.

Proof: Let A be the set of all atoms in L. Then

$$(a\vee b)\wedge b^\perp=\bigvee\{f\in A:f\leqslant(a\vee b)\wedge b^\perp\}$$
$$=\bigvee(\{a,b\}\wedge\{f\in A:f\leqslant b^\perp\})=a\wedge b^\perp,$$

but $a\wedge b^\perp=a$ or $a\wedge b^\perp=0$. If $a\wedge b^\perp=a$, then $a\perp b$. If $a\wedge b^\perp=0$, then by the orthomodularity

$$a\vee b=b\vee(a\vee b)\wedge b^\perp=b\vee a\wedge b^\perp=b,$$

which implies $a=b$; a contradiction. Q.E.D.

Theorem 2: Let (L, u_1, u_2) be a free product of L_1 and L_2 in the category of complete atomistic lattices. Let A_1, A_2, A be the sets of all atoms in L_1, L_2, L, respectively. If

$$A=\{u_1(a_1)\wedge u_2(a_2):a_1\in A_1,a_2\in A_2\},$$

then at least one of L_1 and L_2 must be a Boolean algebra.

Proof: Let $a_2, b_2\in A_2$ be such that $a_2\neq b_2$ and $a_2\not\perp b_2$. Then for any a_1, $b_1\in A_1$, $a_1\neq b_1$, $u_1(a_1)\wedge u_2(a_2)$, and $u_1(b_1)\wedge u_2(b_2)$ are separated by a superselection rule (Lemma 3). By Lemma 4 $u_1(a_1)\wedge u_2(a_2)\perp u_1(b_1)\wedge u_2(b_2)$. By Lemma 1 this implies that $a_1\perp b_1$. From this we get that any two different atoms in A_1 are orthogonal, i.e., L_1 is a Boolean algebra. Q.E.D.

Let (L,M) be a quantum logic such that L is a complete atomistic lattice and M is a set of pure states on L such that there is a one-to-one correspondence between the set of all atoms A in L and the set M. To any $a\in A$ there is exactly one state $m\in M$ such that $m(a)=1$, and $a=\wedge\{b\in L:m(b)=1\}$. The atom a is called the support of the state m, $a=\operatorname{supp}m$. In the sequel, a quantum logic of this type will be called "complete." In a complete quantum logic, the Jauch–Piron property is satisfied in the complete form.

Lemma 5: Let (L_1,M_1), (L_2,M_2), (L,M) be complete quantum logics and let $(L,M)_{\alpha,\beta}$ be the tensor product of (L_1,M_1) and (L_2,M_2). Then, for any $p_1\in M_1$ and $p_2\in M_2$,

$$\operatorname{supp}\beta(p_1,p_2)=\alpha(\operatorname{supp}p_1,\operatorname{supp}p_2).$$

Proof: $\alpha(\operatorname{supp}p_1,\operatorname{supp}p_2)$ is an atom in $L_{[0,\alpha(1,\operatorname{supp}p_2)]}$.[8] From this it follows that $\alpha(\operatorname{supp}p_1, \operatorname{supp}p_2)$ is an atom in L. Further,

$$\beta(p_1,p_2)[\alpha(\operatorname{supp}p_1,\operatorname{supp}p_2)]$$
$$=p_1(\operatorname{supp}p_1)p_2(\operatorname{supp}p_2)=1,$$

i.e., $\alpha(\operatorname{supp}p_1, \operatorname{supp}p_2)=\operatorname{supp}\beta(p_1,p_2)$. Q.E.D.

Theorem 3: Let $(L,M)_{\alpha,\beta}$ be a tensor product of complete quantum logics (L_1,M_1) and (L_2,M_2). Then at least one of (L_1,M_2) and (L_2,M_2) must be a classical quantum logic.[9]

Proof: As $\beta[M_1\times M_2]$ is strong for L, we may put $\beta[M_1\times M_2]=M$. From Lemma 5 we see that all atoms in L of the type $\alpha(a_1,a_2)$, $a_1\in L_1$, $a_2\in L_2$ are atoms. If we put

$$u_1:L_1\to L,\quad u_2:L_2\to L,$$
$$a\mapsto\alpha(a,1)\quad a\mapsto\alpha(1,a)$$

then (L,u_1,u_2) is the free product of L_1 and L_2.[10] Theorem 2 then implies that at least one of L_1 and L_2 is a Boolean algebra. Q.E.D.

III. A NEW DEFINITION OF THE TENSOR PRODUCT

Now we shall formulate a more general definition of a tensor product for quantum logics. To this aim we need some preliminary remarks.

Let S be a set of states on a logic L, we say that a state m is the superposition of the states in S if $S(a)=0$ implies $m(a)=0$, $a\in L$, or equivalently, if $S(a)=1$ implies $m(a)=1$.[11] Here $S(a)=j$, $j=0,1$ means that $s(a)=j$ for all $s\in S$. If (L,M) is a quantum logic, we put $\overline{S}=\{m\in M:S(a)=1\Rightarrow m(a)=1\}$ for any $S\subset M$.

Definition 3: Let (L_1,M_1), (L_2,M_2), and (L,M) be quantum logics. We say that $(L,M)_{\alpha,\beta}$ is a tensor product of (L_1,M_1) and (L_2,M_2) if there are mappings α,β such that

(i) $\alpha:L_1\times L_2\to L$, $\beta:M_1\times M_2\to M$,

$$\beta(m_1,m_2)(\alpha(a_1,a_2))=m_1(a_1)m_2(a_2)$$

for $m_i \in M_i$, $a_i \in L_i$, $i = 1,2$;

(ii) $\{m \in M : m(a) = 1\} = \{\beta(m_1,m_2) : \beta(m_1,m_2)(a) = 1\}^-$ for $a \in L$ of the form

$$a = \bigwedge_k \alpha(a_1{}^k, a_2{}^k), \quad a_1{}^k \in L_1, \quad a_2{}^k \in L_2, \quad k = 1,2,...$$

or

$a = \alpha(a_1,1)^\perp$, resp., $a = \alpha(1,a_2)^\perp$, $a_1 \in L_1$, $a_2 \in L_2$;

(iii) $\alpha[L_1 \times L_2]$ generates L;

(iv) $\beta[M_1 \times M_2]^- = M$.

We shall suppose in the following two propositions that $(L,M)_{\alpha,\beta}$ is the tensor product of (L_i,M_i), $i = 1,2$ by Definition 3.

Proposition 1: Let us define

$$u_1 : L_1 \to L, \qquad u_2 : L_2 \to L$$
$$a_1 \mapsto \alpha(a_1,1), \quad a_2 \mapsto \alpha(1,a_2).$$

Then u_1 and u_2 are orthoinjections.

Proof: By (i) of Definition 3 we have $\beta(m_1,m_2) \times (\alpha(1,1)) = m_1(1)m_2(1) = 1$ for any $m_i \in M_i$, $i = 1,2$. From this we obtain that $m(\alpha(1,1)) = 1$ for any $m \in \beta[M_1 \times M_2]^-$, and by (iv) then $m(\alpha(1,1)) = 1$ for any $m \in M$, i.e., $\alpha(1,1) = 1$.

Further, $\beta(m_1,m_2)(\alpha(a^\perp,1)) = m_1(a^\perp)m_2(1) = (1 - m_1(a)) \times m_2(1) = 1 - m_1(a)\, m_2(1) = 1 - \beta(m_1,m_2)\, (\alpha(a,1)) = \beta(m_1,m_2)(\alpha(a,1)^\perp)$ for all $m_i \in M_i$, $i = 1,2$, $a \in L_1$. From this we get

$$\{\beta(m_1,m_2) : \beta(m_1,m_2)(\alpha(a^\perp,1)) = 1\}^- = \{\beta(m_1,m_2) : \beta(m_1,m_2)(\alpha(a,1)^\perp) = 1\}^-,$$

which implies by (ii) that $\alpha(a^\perp,1) = \alpha(a,1)^\perp$, i.e., $u_1(a^\perp) = u_1(a)^\perp$. Similarly we show that $u_2(a^\perp)$, $a \in L_2$.

By the Jauch–Piron property we have

$$\beta(m_1,m_2)\Big(\alpha\Big(\bigwedge_{k=1}^{\infty} a_1{}^k,1\Big)\Big) = 1 \text{ iff } m_1\Big(\bigwedge_{k=1}^{\infty} a_1{}^k\Big) = 1$$

iff $m_1(a_1{}^k) = 1$ for all $k = 1,2,...$ iff

$\beta(m_1,m_2)(\alpha(a_1{}^k,1)) = 1$

for all k iff $\beta(m_1,m_2)\Big(\bigwedge_{k=1}^{\infty} \alpha(a_1{}^k,1)\Big) = 1.$

From this we obtain

$$\Big\{\beta(m_1,m_2) : \beta(m_1,m_2)\Big(\alpha\Big(\bigwedge_{k=1}^{\infty} a_1{}^k,1\Big)\Big) = 1\Big\}^- = \Big\{\beta(m_1,m_2) : \beta(m_1,m_2)\Big(\bigwedge_{k=1}^{\infty} \alpha(a_1{}^k,1\Big) = 1\Big\}^-,$$

which implies by (ii) that

$$\alpha\Big(\bigwedge_{k=1}^{\infty} a_1{}^k,1\Big) = \bigwedge_{k=1}^{\infty} \alpha(a_1{}^k,1),$$

i.e.,

$$u_1\Big(\bigwedge_{k=1}^{\infty} a_1{}^k\Big) = \bigwedge_{k=1}^{\infty} u_1(a_1{}^k).$$

By the duality we obtain

$$\bigvee_{k=1}^{\infty} u_1(a_1{}^k) = u_1\Big(\bigvee_{k=1}^{\infty} a_1{}^k\Big).$$

This shows that u_1 and u_2 are orthohomomorphisms.

Now $u_1(a_1) = u_1(a_1')$, a_1, $a_1' \in L_1$, implies that $\beta(m_1,m_2)(\alpha(a_1,1)) = \beta(m_1,m_2)(\alpha(a_1',1))$ for any $m_i \in M_i$, $i = 1,2$, which implies that $m_1(a_1) = m_1(a_1')$ for any $m_1 \in M_1$, i.e., $a_1 = a_1'$. From this we see that u_1 and u_2 are orthoinjections. Q.E.D.

Proposition 2: For any $a_i \in L_i$, $i = 1,2$, $u_1(a_1) \leftrightarrow u_2(a_2)$.

Proof: For any $m_i \in M_i$, $i = 1,2$, we have

$\beta(m_1,m_2)(\alpha(a_1,1)) \wedge \alpha(1,a_2) = 1$ iff

$\beta(m_1,m_2)(\alpha(a_1,1)) = 1$ and $\beta(m_1,m_2)(\alpha(1,a_2)) = 1$

iff $m_1(a_1) = 1$ and $m_2(a_2) = 1$ iff $\beta(m_1,m_2)(\alpha(a_1,a_2)) = 1$. From this we have

$$\{\beta(m_1,m_2) : \beta(m_1,m_2)(\alpha(a_1,a_2)) = 1\}^- = \{\beta(m_1,m_2) : \beta(m_1,m_2)(\alpha(a_1,1) \wedge \alpha(1,a_2)) = 1\}^-,$$

i.e., by (ii), $\alpha(a_1,1) \wedge \alpha(1,a_2) = \alpha(a_1,a_2)$. Now for any $a_i \in L_i$, $i = 1,2$,

$$\begin{aligned}\beta(m_1,m_2)&(u_1(a_1) \wedge u_2(a_2))\\ &= \beta(m_1,m_2)(\alpha(a_1,1) \wedge \alpha(1,a_2))\\ &= \beta(m_1,m_2)(\alpha(a_1,a)) = m_1(a_1)m_2(a_2)\\ &= \beta(m_1,m_2)(\alpha(a_1,1)).\end{aligned}$$

$\beta(m_1,m_2)(\alpha(1,a_2)) = \beta(m_1,m_2)(u_1(a_1))\beta(m_1,m_2)(u_2(a_2))$ for any $m_i \in M_i$, $i = 1,2$. This implies that $u_1(a_1)$ and $u_2(a_2)$ are independent (in the probablistic sense) in all states from $\beta[M_1 \times M_2]$, i.e., they have the joint distributions in all states in $\beta[M_1 \times M_2]^- = M$, i.e., they are compatible.[12] Q.E.D.

Theorem 4: Let $(L,M)_{\alpha,\beta}$ be a tensor product of (L_1,M_1) and (L_2,M_2) by Definition 3. Let us put

$$u_1 : L_1 \to L, \qquad u_2 : L_2 \to L,$$
$$a_1 \mapsto \alpha(a_1,1), \quad a_2 \mapsto \alpha(1,a_2).$$

Then (L,u_1,u_2) is a free orthodistributive product of L_1 and L_2.

Proof: We have only to show (ii) and (iii) of Definition 1. Let $a_1 \neq 0$, and $u_1(a_1) \wedge u_2(a_2) = 0$. Let $m_1{}^0 \in M_1$ be such that $m_1{}^0(a_1) = 1$. Then $\beta(m_1{}^0,m_2)(u_1(a_1) \wedge u_2(a_2)) = \beta(m_1{}^0,m_2)(\alpha(a_1,a_2)) = m_1{}^0(a_1)m_2(a_2) = 0$ iff $m_2(a_2) = 0$. Thus $u_1(a_1) \wedge u_2(a_2) = 0$ implies that $m_2(a_2) = 0$ for all $m_2 \in M_2$, i.e., $a_2 = 0$. As $\alpha(a_1,a_2) = u_1(a_1) \wedge u_2(a_2)$, and $\alpha[L_1 \times L_2]$ generates L, we obtain that $u_1(L_1) \cup u_2(L_2)$ generates L. Q.E.D.

In a special case when $\beta[M_1 \times M_2]^- = \beta[M_1 \times M_2]$, the tensor product by Definition 3 becomes identical with the tensor product by Definition 1.

IV. TENSOR PRODUCT OF THE HILBERT SPACE LOGICS

The tensor product by Definition 3 involves also the case of Hilbert space logics. Let H_1 and H_2 be Hilbert spaces with the dimensions of at least three. If the Hilbert spaces are complex, then there exist exactly two (nonequivalent) free orthodistributive products of the logics $L(H_1)$ and $L(H_2)$, satisfying the condition of fullness.[13] They are given by

(i) $H = H_1 \otimes H_2$, $u_1(P_1) = P_1 \otimes H_2$, $u_2(P_2) = H_1 \otimes P_2$, $P_1 \in L(H_1)$, $P_2 \in L(H_2)$,

(ii) $H = \overline{H}_1 \otimes H_2$, $u_2(P_1) = \overline{P}_1 \otimes H_2$, $u_1(P_2) = \overline{H}_1 \otimes P_2$, $P_1 \in L(H_1)$, $P_2 \in L(H_2)$,

where $H_1 \otimes H_2$ denotes the usual tensor product of Hilbert spaces H_1 and H_2, and $\overline{H}$ is the dual space of H. If the Hilbert spaces H_1 and H_2 are real, then there is exactly one free product given by (i). The condition of fullness requires that the mappings

$$u_{1,x}: P_1 \mapsto u_1(P_1) \wedge u_2(x), \quad P_1 \in L(H_1)$$

and

$$u_{2,y}: P_2 \mapsto u_1(y) \wedge u_2(P_2), \quad P_2 \in L(H_2)$$

are surjective for all atoms $x \in L(H_2)$ and $y \in L(H_1)$. It was shown[14] that the condition of fullness is equivalent to the requirement that $u_1(y) \wedge u_2(x)$ is an atom in L.

Let H be a Hilbert space (complex or real, dim $H \geqslant 3$). Let us put $M = \{m_\varphi: \varphi \in H, \|\varphi\| = 1\}$, where $m_\varphi: P \mapsto (P\varphi,\varphi)$ $P \in L(H)$, is the vector state corresponding to φ. Then $(L(H),M)$ is a complete quantum logic. Let $(L(H_1),M_1)$, $(L(H_2),M_2)$, $(L(H_1 \otimes H_2),M)$, and $(L(\overline{H}_1 \otimes H_2),\overline{M})$ be the complete quantum logics corresponding to H_1, H_2, $H_1 \otimes H_2$, and $\overline{H}_1 \otimes H_2$, respectively. If we put

$$\begin{aligned}
(1)\quad \alpha:&\quad L(H_1) \times L(H_2) \to L(H_1 \otimes H_2),\\
&\quad (P_1,P_2) \mapsto P_1 \otimes P_2,\\
\beta:&\quad M_1 \times M_2 \to M,\\
&\quad (m_{\varphi_1}, m_{\varphi_2}) \mapsto m_{\varphi_1 \otimes \varphi_2},\\
(2)\quad \bar{\alpha}:&\quad L(H_1) \times L(H_2) \to L(\overline{H}_1 \otimes H_2),\\
&\quad (P_1,P_2) \mapsto \overline{P}_1 \otimes P_2,\\
\overline{\beta}:&\quad M_1 \times M_2 \to \overline{M},\\
&\quad (m_{\varphi_1}, m_{\varphi_2}) \mapsto m_{\bar{\varphi}_1 \otimes \varphi_2},
\end{aligned}$$

then $(L(H_1 \otimes H_2),M))_{\alpha,\beta}$ and $(L(\overline{H}_1 \otimes H_2),\overline{M})_{\bar{\alpha},\overline{\beta}}$ are tensor products by Definition 3. Indeed,

(i) $\beta(m_{\varphi_1}, m_{\varphi_2}) \quad (\alpha(P_1,P_2)) = m_{\varphi_1 \otimes \varphi_2}(P_1 \otimes P_2) = (P_1 \otimes P_2\varphi_1 \otimes \varphi_2, \varphi_1 \otimes \varphi_2) = (P_1\varphi_1,\varphi_1)(P_2\varphi_2,\varphi_2) = m_{\varphi_1}(P_1) m_{\varphi_2}(P_2)$;

(ii) Let $S \subset M$ and $Q = \{\varphi \in H: m_\varphi \in S\}$. Then $\overline{S} = \{m_\varphi: \varphi \in \overline{Q}\}$, where $\overline{Q}$ is the closed linear subspace of H generated by Q. The closed linear subspace of $H_1 \otimes H_2$ corresponding to $P_1 \otimes P_2$ has the base $\{\varphi_i \otimes \psi_j\}_{i,j}$, where $\{\varphi_i\}_i$ and $\{\psi_j\}_j$ are the bases of P_1 and P_2, respectively. Indeed,

$$\begin{aligned}
\bigvee_{i,j} [\varphi_i \otimes \psi_j] &= \bigvee_{i,j} u_1[\varphi_i] \wedge u_2[\psi_j]\\
&= \Big(\bigvee_i u_1[\varphi_i]\Big) \wedge \Big(\bigvee_j u_2[\psi_j]\Big)\\
&= u_1(P_1) \wedge u_2(P_2) = P_1 \otimes P_2,
\end{aligned}$$

where $[\varphi]$ denotes the one-dimensional subspace generated by the vector φ.

(iii) is evident.

The same can be proved for $(L(\overline{H}_1 \otimes H_2),\overline{M})_{\bar{\alpha},\overline{\beta}}$. As any tensor product is also a free product, there are exactly two tensor products of $(L(H_1),M_1)$ and $(L(H_2),M_2)$ in the complex case.

We shall say that the tensor product $(L,M)_{\alpha,\beta}$ of complete quantum logics (L_1,M_1) and (L_2,M_2) satisfies the condition of fullness if the mappings

$$\begin{aligned}
u_{1,a_2}:&\quad L_1 \to L_{[0,u_2(a_2)]}\\
&\quad a_1 \mapsto \alpha(a_1,a_2) = u_1(a_1) \wedge u_2(a_2)
\end{aligned}$$

and

$$\begin{aligned}
u_{2,a_1}:&\quad L_2 \to L_{[0,u_1(a_1)]}\\
&\quad a_2 \mapsto \alpha(a_1,a_2) = u_1(a_1) \wedge u_2(a_2)
\end{aligned}$$

are surjective for any atoms $a_1 \in L_1$ and $a_2 \in L_2$.

Proposition 3: Let (L_1,M_1), (L_2,M_2), (L,M) be complete quantum logics and let $(L,M)_{\alpha,\beta}$ be the tensor product of (L_i, M_i), $i = 1,2$, in the sense of Definition 3. The condition of fullness is fulfilled if and only if

$$\begin{aligned}
&\{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(1,b)) = 1\}^-\\
&\quad = \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(1,b)) = 1\}
\end{aligned}$$

for any atom $b \in L_2$, and

$$\begin{aligned}
&\{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(a,1)) = 1\}^-\\
&\quad = \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(a,1)) = 1\}
\end{aligned}$$

for any atom $a \in L_1$.

Proof: First we show that $\alpha(\operatorname{supp} m_1, \operatorname{supp} m_2)$ is an atom in L for any $m_i \in M_i$, $i = 1,2$. Put $b = \operatorname{supp} m_2$. If there is $B \in L$ such that $B \leqslant \alpha(\operatorname{supp} m_1, b)$, then there must be a state $m \in M$ such that $m(B) = 1$. But then also $m(\alpha(1,b)) = 1$, so that $m \in \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(1,b)) = 1\}^- = \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(1,b)) = 1\}$. From this it follows that $m = \beta(m_1', m_2)$ for some $m_1' \in M_1$. Now $\beta(m_1',m_2) \times (\alpha(\operatorname{supp} m_1, b)) = 1$ implies that $m_1' = m_1$. Then

$$\begin{aligned}
&\{m \in M: m(\alpha(\operatorname{supp} m_1, b)) = 1\}\\
&\quad = \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(\operatorname{supp} m_1, b)) = 1\}^-\\
&\quad \subset \{\beta(m_1,m_2): \beta(m_1,m_2)(\alpha(1,b)) = 1\},
\end{aligned}$$

which implies that

$$\begin{aligned}
&\{m \in M: m(\alpha(\operatorname{supp} m_1, b)) = 1\}\\
&\quad = \{\beta(m_1,m_2)\} \subset \{m \in M: m(B) = 1\},
\end{aligned}$$

i.e., $\alpha(\operatorname{supp} m_1, b) \leqslant B$.

From $\beta(m_1,m_2)(\alpha(\operatorname{supp} m_1, \operatorname{supp} m_2)) = m_1(\operatorname{supp} m_1) \cdot m_2(\operatorname{supp} m_2) = 1$ it follows that

$$\operatorname{supp} \beta(m_1,m_2) = \alpha(\operatorname{supp} m_1, \operatorname{supp} m_2).$$

Let A_1, A_2, A be the sets of all atoms in L_1, L_2, L, respectively. The last statement implies that any atom under $\alpha(1,b)$, $b \in A_2$ is of the form $\alpha(a,b)$, $a \in A_1$. If $C \leqslant \alpha(1,b)$, we get

$$\begin{aligned}
C &= \bigvee \{a \in A: a \leqslant C\} = \bigvee \{\alpha(a,b): a \in A_1, \alpha(a,b) \leqslant C\}\\
&= \alpha(\bigvee \{a \in A_1: \alpha(a,b) \leqslant C\} b.
\end{aligned}$$

The inverse statement is clear. Q.E.D.

It can be easily seen that the condition of Proposition 3 is fulfilled for the Hilbert space logics.

V. TENSOR PRODUCT OF ONE QUANTUM AND ONE CLASSICAL LOGICS

Theorem 5: Let (L_1,M_1), (L_2,M_2), and (L,M) be complete quantum logics and let $(L,M)_{\alpha,\beta}$ be a tensor product of (L_1,M_1) and (L_2,M_2) by Definition 3. Let (L_1,M_1) be a classical quantum logic and let the mappings u_{1,a_2} be surjective for any atom $a_2 \in L_2$. Then $\beta[M_1 \times M_2]^- = \beta[M_1 \times M_2]$.

Proof: Let A_1, A_2, A be the sets of all atoms in L_1, L_2, L, respectively. As L_2 is a Boolean algebra, $\{\alpha(1,a_2): a_2 \in A_2\}$ are mutually orthogonal elements in the center L' of L. Let $m \in \beta[M_1 \times M_2]^-$. As m is a pure state on L, we have $m(c) = 1$ or $m(c) = 0$ for any $c \in L'$. But then there is an element $a_2 \in A_2$

such that supp $m \leqslant \alpha(1,a_2)$. Since the map u_{1,a_2} is surjective, there is $a_1 \in A_1$ such that $\text{supp}\, m = \alpha(a_1,a_2)$, i.e., $m = \beta(m_1,m_2)$, where supp $m_1 = a_1$, supp $m_2 = a_2$. Q.E.D.

From Theorem 5, Theorem 3, and Theorem 5 in Ref. 15 we obtain the following statement.

Theorem 6: Let (L_1,M_1), (L_2,M_2), and (L,M) be complete quantum logics. Then $(L,M)_{\alpha,\beta}$ is a tensor product of (L_1,M_1) and (L_2,M_2) by Definition 1 if and only it is their tensor product by Definition 3 satisfying the condition of fullness and at least one of (L_1,M_1) and (L_2,M_2) is classical.

Theorem 7: Let (L,M) be a complete quantum logic and $(S,\mathscr{M})$ be a classical logic such that S is the set of all subsets of X, where X is a nonreal-measurable set, and $\mathscr{M} = \{m_x : x \in X\}$, where m_x is the probability measure on S concentrated in the point $x \in X$. Then $(\widetilde{L},\widetilde{M})_{\alpha,\beta}$ is the tensor product of (L,M) and $(S,\mathscr{M})$ by Definition 3, satisfying the condition of fullness if and only if $L = \oplus_{x\in X} L_x$, $L_x = L$ for any $x \in X$, and $M = \{m \circ \delta_x : m \in M, x \in X\}$, where

$$\delta_y : \oplus_{x\in X} L_x \to L_y,$$
$$(a_x)_{x\in X} \mapsto a_y .$$

Proof: Let $(\widetilde{L},\widetilde{M})_{\alpha,\beta}$ be a tensor product of (L,M) and $(S,\mathscr{M})$, satisfying the condition of fullness. Then $\{\alpha(1,x) : x \in X\}$ are pairwise orthogonal elements in the center L' of L, so that we can write $L = \oplus_{x\in X} L_{[0,\alpha(1,x)]}$. As the condition of fullness is fulfilled, the logics $L_{[0,\alpha(1,x)]}$ are isomorphic with L for all $x \in X$. By Ref. 16, any state on L is of the form

$$\sum_{n\in N} p_{x_n} \cdot m_{x_n},$$

where $\{x_n : n \in \mathrm{N}\}$ is a countable subset of X, $\{p_{x_n} : n \in N\}$ is the partition of the unity, and m_{x_n} is a state on L_{x_n}. It can be easily checked that any pure state on L is of the form $m \circ \delta_x$ for some $x \in X$ and $m \in M$. As the set $\widetilde{M}$ is strong for $\widetilde{L}$, it must contain all $m \circ \delta_x$, $m \in M$, $x \in X$.

On the other hand, let $(\widetilde{L},\widetilde{M})$ be such that

$$\widetilde{L} = \oplus_{x\in X} L_x, \quad L_x = L, \ x \in X,$$

and

$$\widetilde{M} = \{m \circ \delta_x : m \in M, x \in X\}.$$

Clearly, $\widetilde{M}$ is strong for $\widetilde{L}$. Let us define

$$\alpha: \quad L \times S \to \widetilde{L},$$
$$(\alpha,E) \mapsto (a_x)_{x\in X}, \quad a_x = \begin{cases} a & \text{if } x \in E, \\ 0 & \text{if } x \notin E, \end{cases}$$
$$\beta: \quad M \times \mathscr{M} \to \widetilde{M},$$
$$(m, m_x) \mapsto m \cdot \delta_x .$$

Then

$$\beta(m,m_{x_0})(\alpha(a,E)) = m \circ \delta_{x_0}((a_x)_{x\in X}) = \begin{cases} m(a) & \text{if } x_0 \in E, \\ 0 & \text{if } x_0 \notin E, \end{cases}$$

i.e., $\beta(m,m_{x_0})(\alpha(a,E)) = m(a) \cdot m_{x_0}(E)$. As every pure state on $\oplus_{x\in X} L_x$ is of the form $m \circ \delta_x$, we get $\beta[M \times \mathscr{M}] = \beta[M \times \mathscr{M}]^- = \widetilde{M} \cdot \widetilde{L}$ is generated by the elements $\alpha(a,x)$, $a \in L$, $x \in X$. The condition of fullness follows by Theorem 5 in Ref. 17. Q.E.D.

ACKNOWLEDGMENTS

This work was done at the Vrije Universiteit in Brussels. I am very grateful to Professor Jean Reignier for his kind hospitality and to Dirk Aerts for many helpful discussions.

[1] T. Matolcsi, Acta Sci. Math. **37**, 263 (1975).
[2] S. Pulmannová, Int. J. Theor. Phys. **22**, 837 (1983).
[3] R. Greechie, Proc. Am. Math. Soc. **67**, 17 (1977).
[4] See proof of Theorem 2 in Ref. 2.
[5] See Ref. 2, Corollary 1.
[6] See Ref. 2.
[7] C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, MA, 1976).
[8] See Ref. 2.
[9] A similar result, in another formulation, was obtained by D. Aerts, Found. Phys. **12**, 1131 (1982).
[10] See Ref. 2.
[11] V. S. Varadarajan, *Geometry of Quantum Theory I* (Van Nostrand, Princeton, NJ, 1968).
[12] A. Dvurečenskij and S. Pulamannová, Math. Slovaca **32**, 155 (1982).
[13] D. Aerts and I. Daubechies, Helv. Phys. Acta **51**, 661 (1978); see also Ref. 1.
[14] D. Aerts and I. Daubechies, Helv. Phys. Acta **51**, 661 (1978).
[15] See Ref. 2.
[16] V. Maňasová and P. Pták, Int. J. Theor. Phys. **20**, 451 (1981).
[17] See Ref. 2.

Triality principle and G_2 group in spinor language

Z. Hasiewicz and A. K. Kwasniewski
Institute of Theoretical Physics, University of Wrocław, ul. Cybulskiego 36, 50-205 Wrocław, Poland

(Received 27 December 1983; accepted for publication 15 June 1984)

The presented paper is aimed at providing a systematic study of a relation between octonions and spinors corresponding to S^7 7-sphere, starting from a natural point of view, enabling us to endow spinor space $S_+(8,0)$ with octonion algebra structures. As a result we arrive at formulations of triality principle in its finite form in terms of vector and fundamental representations of Spin (8) group—both for spinors and vectors. The group of automorphisms of octonion algebra, as well as its Lie algebra, gains clear interpretation in the context. The method proposed is purely algebraic and could be applied as well to $\mathscr{C}(4,4)$ Clifford algebra corresponding to ${}_4S^7$ indefinite 7-sphere geometry.

I. INTRODUCTION

Consider $SO(8)/SO(7) = S^7$ an ordinary 7-sphere and $SO(4,4)/SO(3,4) = {}_4S^7$ an indefinite 7-sphere. Both are distinguished manifolds among all irreducible simply connected globally symmetric pseudo-Riemannian manifolds with consistent absolute parallelisms.[1]

The geometries of points, lines etc. from S^7 or ${}_4S^7$ are known to be governed by the geometrical triality principle which could be formulated in terms of octonion algebra[2–4] or split octonion algebra,[4] correspondingly. The principle of triality in the geometry of S^7 was discovered by Cartan[5] while he was studying simple and semisimple groups and corresponding spinors. The principle of triality in non-Euclidean geometries of both S^7 and ${}_4S^7$ spheres was described by Rozenfeld in Ref. 4.

Spinors of SO(8) and SO(4,4) groups are naturally relevant to these exceptional cases of geometries admitting triality principle as the corresponding metrices are invariant under SO(8) or SO(4,4) group. One would expect it then to be possible to formulate the principle of triality in terms of corresponding spinors apart from the possibility of its formulations via octonion or split octonion algebras. Even more, we expect in these very cases octonions and split octonions to be identified with corresponding spinors in some algebraic sense.[6]

The triality principle for spinors of eight-dimensional space (real) with a quadratic form of rank 8 and index 4 was derived by Chevalley.[7] This corresponds to a SO(4,4) group and hence to a ${}_4S^7$ indefinite 7-sphere. His method of arriving at the principle as well as a geometrical interpretation of the construction essentially relies on the existence of maximal totally singular subspaces. Therefore, there is no straightforward application of his construction to the case of a SO(8) group acting on S^7. This paper is aimed at providing a systematic study of a relation between octonions and spinors corresponding to S^7, starting from a natural point of view enabling us to endow spinor spaces and generating space $E(8,0)$ of Clifford algebra $\mathscr{C}(8,0)$ with octonion algebra structures. As a result, we arrive at formulations of the triality principle in its finite form in terms of vector and fundamental representations of Spin (8) group—both for spinors and vectors. One also achieves a clear interpretation of the group of automorphism of the octonions.

Our point of view, via canonical methods of Clifford algebra representations, leads to octonion algebra isomorphisms between generating vector space and spinor spaces (now algebras). Various forms of triality are then natural consequences of these isomorphisms which we call vector-spinor reciprocities.

The point of view proposed here might be treated as a complementary (alternative?) way of understanding the peculiarity of a Spin (8) group and consequently of the geometry of a S^7 sphere, its parallelizability included.

As for that last property, Clifford modules are known to be intrinsically related also to the vector fields problem on spheres.[8] In particular, $\mathscr{C}(0,7) \sim \mathscr{C}^{(+)}(8,0)$ Clifford algebra distinguished properties could be used to prove parallelizability of a S^7 sphere, which could be achieved as well via octonion algebra.

It is no surprise then that octonion algebra can be obtained from $\mathscr{C}(0,7) \sim \mathscr{C}^{(+)}(8,0)$ Clifford algebra. This observation is due to Atiyah and the corresponding construction is presented in Ref. 9. While formulating our point of view in Ref. 6 we were ignorant of this observation and apparently the basic trick is the same, though now, as we start right from $\mathscr{C}(8,0)$ Clifford algebra with generating space $E(8,0)$, the construction should be more refined, especially when one's aim is to also derive all possible global forms of triality principle. The general idea proposed here is the following.[6] Consider any associative algebra $\mathscr{C}$ and S its, say, left ideal. Let E be a vector subspace of $\mathscr{C}$ and $\dim E = \dim S$. Denote by h any vector space isomorphism h: $S \to E$ and then define the bilinear mapping

$$S \times S \to (\psi,\psi') \to \phi(\psi,\psi') = h(\psi)\psi' S. \tag{1.1}$$

The subspace now becomes a nonassociative in general-algebra (S,Φ). If in addition, for example, appropriate quadratic forms are defined on S and E, we shall also require the isomorphism h to be an isometry and if $\mathscr{C}$ is an algebra with unity we shall require (if this is possible) the (S,Φ) algebra to be the one with unity, etc. (S,Φ) algebra is said to be retransmitted via (1.1) from $\mathscr{C}$. Therefore, the idea is to retransmit as many properties of $\mathscr{C}$ as is possible. One easily recognizes the importance of algebra $\mathscr{C}$ representations for construction of the type.[10] In our case of semisimple algebra we then use of course the Wedderburn's theorem. It is rather trivial

 0022-2488/85/010006-06$02.50

to adjust a definition of ϕ-multiplication, now in vector space E, so that the algebra (S,Φ) is isomorphic to (E,ϕ) via the same h. In the relevant case of $\mathscr{C} = \mathscr{C}(8,0)$ Clifford algebra and $E = E(8,0)$ being its generating space, we must slightly generalize the above scheme as now $S = S^{(+)} \oplus S^{(-)}$ and $\dim E \neq \dim S$. However, $\dim E = \dim S^{(+)} = \dim S^{(-)}$ [S-pinor module, $S^{(+)}$—spinor module of $\mathscr{C}(8,0)$ Clifford algebra]. Therefore, we consider a sequency of isometries i, j, h

$$E \xrightarrow{j} S^{(-)} \xrightarrow{i} S^{(+)} \xrightarrow{h} E, \tag{1.2}$$

such that this diagram is commutative ($h = (ij)^{-1}$). An algebra structure $(S^{(+)}, \Phi^{(+)})$ is now defined according to

$$S^{+} \times S^{+} \ni (\psi,\psi') \rightarrow \Phi^{(+)}(\psi,\psi') = \mathrm{i}[\mathrm{h}(\psi)\psi']. \tag{1.3}$$

Similarly to the former (1.1) case one may introduce appropriate algebra structures $(S^{(-)}, \Phi^{(-)})$ and (E,ϕ) so that these two are isomorphic to $(S^{(+)}, \Phi^{(+)})$ algebra. The algebraic structures obtained that way are said to be retransmitted from $\mathscr{C}$ via the commutative diagram (1.2).

The main result of the presented paper is the construction of retransmitted structures with unities, all of them forming the algebras of octonions.

A 14-parameter family of all corresponding isometries h is constructed [hence, sequences (1.2)]. As a result we arrive at a natural form of the global principle of triality in terms of obtained octonionic structures on spinor spaces and also on vector space E. Their groups of automorphisms gain an interesting and clear interpretation.

Our study was stimulated by the recently increased interest of physicists in problems strictly related to the subject of the presented paper (see for example, Refs. 11–17 and also references therein).

II. RETRANSMITTED ALGEBRAS OF OCTONIONS

In this section we shall construct retransmitted via (1.2) and (1.3) structures, i.e., $(S^{(+)}, \Phi^{(+)})$, $(S^{(-)}, \Phi^{(-)})$, and (E,ϕ) algebras which we require to be algebras with unities. It appears that they are isomorphic to Cayley algebra of octonions. First, we establish some preliminary knowledge and notation, which is fairly standard.[8–10] Then we introduce a 14-parameter family of isometries $h: S^{(+)} \rightarrow E$ [consequently, corresponding family of diagrams (1.2) and retransmitted algebra with unity structures].

(1) $E(8,0)$ denotes the Euclidean real eight-dimensional vector space and $\mathscr{C}(8,0)$ is its universal Clifford algebra. The map π

$$\mathscr{C}\,\mathrm{Pin}(8,0) \ni s \rightarrow \pi_s \in \mathrm{O}(8,0) \subset \mathrm{Aut}(E(8,0)),$$

where

$$E(8,0) \ni v \rightarrow \pi_s(v) = sv\alpha(s^{-1}) \in E(8,0)$$

is the vector representation of the Pin(8,0) group (α denotes the main automorphism of Clifford algebra). The pinor module of $\mathscr{C}(8,0)$ algebra is a 16-dimensional real vector space

$$S(8,0) = \mathscr{C}(8,0)e(8,0) \sim R^{16}$$

and the field

$$F(8,0) = e(8,0)\mathscr{C}(8,0)e(8,0)$$

is the isomorphic to R. Here $e(8,0)$ denotes a primitive idempotent, hence $S(8,0)$ is a left minimal ideal of $\mathscr{C}(8,0)$. $e(8,0)$ can be choosen so that it remains primitive also in the even subalgebra $\mathscr{C}^{(+)}(8,0)$ of the algebra $\mathscr{C}(8,0)$.

For example, one may choose

$$e(8,0) = 1/16(1 + E_{1238})(1 + E_{1458})(1 + E_{7816})(1 + E_{2486}),$$

where $E_{ijkl} = E_iE_jE_kE_l$ and $\{E_i\}_1^8$ is an arbitray orthonormal basis of $E(8,0)$.

The pinor module $S(8,0)$ can be given naturally a direct sum structure $S(8,0) = C^{(+)}(8,0)e(8,0) \oplus C^{(-)}(8,0)e(8,0) = S^{(+)}(8,0) \oplus S^{(-)}(8,0)$, where $S^{(+)}(8,0)$ is now a spinor module, i.e., the space of irreducible (though not faithful) representation of $C^{+}(8,0)$ algebra. Also, quadratic forms are naturally defined on $S(8,0)$ pinor space. The only Spin (8,0) invariant quadratic forms $\Theta_{\pm}$ on $S(8,0)$ are those generated by $\beta_{\pm}$ antiauthomorphisms ($\beta_{\pm}\, v = \pm v \in E(8,0)$) according to

$$S(8,0) \times S(8,0) \ni (\psi,\psi') \rightarrow \Theta_{\pm}(\psi,\psi') = \beta_{\pm}(\psi)\psi' \in R$$

and we use the same notation for associated bilinear forms so that

$$\Theta_{\pm}(\psi) = \Theta_{\pm}(\psi,\psi').$$

The forms Θ_+ and Θ_- equip the pinor space with Euclidean and correspondingly maximally isotropic pseudo-Euclidean structure. From now on, we shall use the abbreviations $S^{(\pm)} = S^{(\pm)}(8,0)$, $E = E(8,0)$, Spin (8) = Spn(8,0), and Q for quadratic form on E. Finally, the fundamental representation of Spin (8) is denoted by τ. It is not faithful neither on S_+ nor S_- and $\ker\tau = \{1,J\} \sim Z_2$, where $J = E_1E_2 \cdots E_8$ is a unit pseudoscalar.

(2) Now we proceed to construct a 14-parameter family of isometries $h: S^{(+)} \rightarrow E$ and consequently corresponding sequences (1.2) which in turn generate retransmitted algebra structures $(S^{(+)}, \Phi^{(+)})$, $(S^{(-)}, \Phi^{(-)})$, and (E,ϕ). Let $E^* = E - \{0\}$ and $S^* = S^{(+)} - \{0\}$, then for any $(v,\eta) \in E^* \times S^*$ we have a sequence (commutative diagram)

$$E \xrightarrow{j_\eta} S^{(-)} \xrightarrow{i_v} S^{(+)} \xrightarrow{h_{v,\eta}} E, \tag{2.1}$$

where

$$\begin{aligned} &E \ni w \rightarrow j_\eta(w) := w\eta \in S^{(-)},\\ &S^{(-)} \ni \psi \rightarrow \mathrm{i}_v(\psi) := v\psi \in S^{(+)},\\ &S^{(+)} \ni \psi \rightarrow h_{v,\eta}(\psi) := (i_v \circ j_\eta)^{-1}(\psi) \in E. \end{aligned} \tag{2.2}$$

It is easy to see that following lemmas hold:

Lemma 2.1: For any $(v,\eta) \in E^* \times S^*$: i_v and j_η are isomorphisms of vector spaces.

Proof: The maps are obviously linear. Assume $i_v(\psi) = 0$. This imples that from

$$0 = \Theta(i_v(\psi)) = \beta_+(i_v(\psi))i_v(\psi) = \beta_+(\psi)v^2\psi = Q(v)\Theta(\psi)$$

due to definiteness of Q and Θ, and we obtain $\psi = 0$. Analogously we obtain $w = 0$ from $j_\eta(w) = 0$. ■

Lemma 2.2:

$$S^{(+)} \ni \psi \rightarrow \mathrm{h}_{v,\eta}(\psi) = (\Theta(\eta)Q(v))^{-1}\beta(i_v(\psi))E^a\eta E_a \in E,$$

where $\{E_a\}_1^8$ is any orthonormal basis of E.

Proof: For $S^{(+)} \ni \psi = (i_v \circ j_\eta)(w) = vw\eta$ we can write

$$h_{v,\eta}(\psi) = h_{v,\eta}(vw\eta) = (\Theta(\eta)Q(v))^{-1}\beta(i_v(vw\eta))E^a\eta E_a$$

$$= (\Theta(\eta)Q(v))^{-1}\beta(\eta)wv^2E^a\eta E_a$$
$$= \Theta(\eta)^{-1}\beta(\eta)wE^a\eta E_a$$
$$= \Theta(\eta)^{-1}\beta(\eta)\tfrac{1}{2}(wE^a + E^a w)\eta E_a$$
$$= \Theta(\eta)^{-1}\beta(\eta)w\circ E^a\eta E_a$$
$$= w^aE_a = w. \qquad \blacksquare$$

Lemma 2.3: For any $(v,\eta)\in E^*\times S^*$:

(1) $\Theta(i_v(\psi)) = Q(v)\Theta(\psi), \quad \psi\in S^{(-)}$,

(2) $\Theta(j_\eta(w)) = \Theta(\eta)Q(w), \quad w\in E$,

(3) $\Theta(h^{-1}_{v,\eta}(w)) = Q(v)\Theta(\eta)Q(w), \quad w\in E$,

(4) $Q(h_{v,\eta}(\psi)) = (Q(v)\Theta(\eta))^{-1}\Theta(\psi), \quad \psi\in S^{(+)}$.

Proof: Note that (3) and (4) are equivalent and

(1) $\Theta(i_v(\psi)) = \Theta(v\psi) = \beta(v\psi)v\psi = \beta(\psi)v^2\psi = Q(v)\Theta(\psi)$,

(2) $\Theta(j_\eta(w)) = \Theta(w\eta) = \beta(w\eta)w\eta = \beta(\eta)w^2\eta$
$= Q(w)\Theta(\eta)$,

(3) $\Theta(h^{-1}_{v,\eta}(w)) = \Theta((i_v\circ j_\eta)(w)) = \Theta(vw\eta) = \beta(vw\eta)vw\eta$
$= \beta(\eta)wv^2w\eta = (Q(v)\Theta(\eta))Q(w). \qquad \blacksquare$

One could define retransmitted structures already via the commutative diagram (2.1) with $i_v, j_\eta, h_{v,\eta}$ being the only vector space isomorphisms. We are going to do that, however, subsequently we restrict the (v,η) pair to be an element of $S^7\times S^7\subset E^*\times S^*$ so that $i_v, j_\eta, h_{v,\eta}$ become isometries according to (1.1). (This is obvious in view of Lemma 2.3.) The retransmitted algebraic structures are the following:

Definition 1:

$S^{(+)}\times S^{(+)}\ni(\psi,\psi')\mapsto\Phi^{(+)}_{v,\eta}(\psi,\psi') := i_v[h_{v,\eta}(\psi)\psi']\in S^{(+)}$,

$S^{(-)}\times S^{(-)}\ni(\psi,\psi')\mapsto\Phi^{(-)}_{v,\eta}(\psi,\psi') := h_{v,\eta}[i_v(\psi)]i_v(\psi')\in S^{(-)}$,

$E\times E\ni(w,w')\mapsto\phi_{v,\eta}(w,w') := (h_{v,\eta}\circ i_v)[wh^{-1}_{v,\eta}(w')]\in E. \qquad \blacksquare$

One readily sees the following.

Lemma 2.4: (1) $h_{v,\eta}$ is an algebra isomorphism of $(S^{(+)},\Phi^{(+)}_{v,\eta})$ onto $(E,\phi_{v,\eta})$; (2) i_v is an algebra isomorphism of $(S^{(-)},\Phi^{(-)}_{v,\eta})$ onto $(S^{(+)},\Phi^{(+)}_{v,\eta})$; (3) j_η is an algebra isomorphism of $(E,\phi_{v,\eta})$ onto $(S^{(-)},\Phi^{(-)}_{v,\eta})$; for any $(v,\eta)E^*\times S^*$, i.e., $(E,\phi_{v,\eta})$, $(S^{(+)},\Phi^{(+)}_{v,\eta})$, and $(S^{(-)},\Phi^{(-)}_{v,\eta})$ algebras are isomorphic.

Proof:

(1) $(h_{v,\eta}\circ\Phi^{(+)}_{+v,\eta})(\psi,\psi') = (h_{v,\eta}\circ i_v,h_{v,\eta}(\psi)\psi')$
$= (h_{v,\eta}\circ i_v)(h_{v,\eta}(\psi)h^{-1}_{v,\eta},h_{v,\eta}(\psi'))$
$= \phi_{v,\eta}(h_{v,\eta}(\psi),h_{v,\eta}(\psi'))$,

(2) $i_v\Phi^{(-)}_{v,\eta}(\psi,\psi') = i_v(h_{v,\eta}(i_v(\psi))i_v(\psi')) = \Phi^{(+)}(i_v(\psi),i_v(\psi'))$,

(3) obvious.

It is important to note that these algebras are algebras with unity as we have

Lemma 2.5: (1) η is unity of $(S^{(+)},\Phi^{(+)}_{v,\eta})$; (2) v^{-1} is unity of $(E,\phi_{v,\eta})$; (3) $v^{-1}\eta$ is unity of $(S^{(-)},\Phi^{(-)}_{v,\eta})$; for any $(v,\eta)\in E^*\times S^*$.

Proof: Note that

$h_{v,\eta}(\eta) = (Q(\eta)Q(v))^{-1}\beta(i_v(\eta))E^a\eta E_a$
$= (\Theta(\eta)Q(v))^{-1}\beta(\eta)v\circ E^a\eta E_a = v/Q(v) = v^{-1}$,

hence,

(1) $\Phi^{(+)}(\eta,\psi) = i_v(h_{v,\eta}(\eta)\psi) = i_v(v^{-1}\psi) = \psi$,
$\Phi^{(+)}(\psi,\eta) = (i_v\circ h_{v,\eta})(\psi)\eta = (i_v\circ j_\eta\circ h_{v,\eta})(\psi)$
$= id(\psi) = \psi$.

(2) Follows from: $h_{v,\eta}(\eta) = v^{-1}$ and Lemma 2.4.

(3) Follows from: $i_v^{-1}(\eta) = v^{-1}\eta$ and Lemma 2.4. $\blacksquare$

Let us now denote by Φ the Cayley algebra of octonions. Using then Lemma 2.3 and the Hurwitz theorem one can prove

Proposition 2.1:

(1) $(E,\phi_{v,\eta})\sim\Phi$ iff $v\in S^7\subset E$,

(2) $(S^{(+)},\Phi^{(+)}_{v,\eta})\sim\Phi$ iff $\eta\in S^7\subset S^{(+)}$,

(3) $(S^{(-)},\Phi^{(-)}_{v,\eta})\sim\Phi$ iff $v^{-1}\eta\in S^7\subset S^{(-)}$.

Proof: Due to the Hurwitz theorem we need to prove only that the above algebras (with unities) have a composition property iff the following conditions are satisfied:

$$(1)\ Q(\phi_{v,\eta}(w,w')) = Q((h_{v,\eta}\circ i_v)(wh^{-1}_{v,\eta}(w')))$$
$$\overset{\text{Lem. 2.3}}{=} \Theta(i_v(wh^{-1}_{v,\eta}(w)))(Q(v)\Theta(\eta))^{-1}$$
$$= \beta(vwh^{-1}_{v,\eta}(w'))vwh^{-1}_{v,\eta}(w')$$
$$\times(Q(v)\Theta(\eta))^{-1} = Q(v)Q(w)$$
$$\times\Theta(h^{-1}_{v,\eta}(w'))Q^{-1}(v)\Theta^{-1}(\eta)$$
$$\overset{\text{Lem. 2.3}}{=} Q(v)Q(w)Q(w').$$

Hence, $(E,\phi_{v,\eta})$ is a composition algebra iff $Q(v)=1$, i.e., $v\in S^7$.

$$(2)\ \Theta(\Phi^{(+)}_{v,\eta}(\psi,\psi'))$$
$$= \Theta(i_v(h_{v,\eta}(\psi)\psi')) = Q(v)Q(h_{v,\eta}(\psi))\Theta(\psi')$$
$$\overset{\text{Lem. 2.3}}{=} Q(v)(Q(v)\Theta(\eta))^{-1}\Theta(\psi)\Theta(\psi')$$
$$= \Theta^{-1}(\eta)\Theta(\psi)\Theta(\psi').$$

Hence, $(S^{(+)},\Phi^{(+)}_{v,\eta})$ is a composition algebra for $\eta\in S^7$.

(3) Analogously as in (1) and (2). $\blacksquare$

We arrive thus at the conclusion that the sequence (2.1) with $(v,\eta)\in S^7\times S^7$ retransmits via Definition 1 octonion algebra structures only or equivalently, the sequence (2.1) becomes a sequence of isomorphic octonion algebras. As an illustration of standard properties let us derive the structure constants of, say $(E,\phi_{v,\eta})$ algebra and let us define also the appropriate conjugations. We then have

Lemma 2.6: For $\{E_i\}_1^7$ orthonormal set of perpendicular to v vectors of E

$$\phi_{v,\eta}(E_i,E_j) = -\delta_{ij}v + \epsilon_{ijk}(v,\eta)E_k, \quad 1\leqslant i,j,k\leqslant 7. \qquad \blacksquare$$

For example, it is enough to note that $\epsilon_{ijk}(v,\eta) := \beta(\eta)E_i\times E_jE_kv\eta$ is totally antisymmetric. Analogous formulas are derived for $(S^{(+)},\Phi^{(+)})$ and $(S^{(-)},\Phi^{(-)})$ algebras via respective isomorphism.

As for conjugations these are defined correspondingly by

Definition 2:

$$E \ni w \to \bar{w} = vwv^{-1} \in E,$$

$$S^{(+)} \ni \psi \to \bar{\psi} = h_{v,\eta}^{-1}(vh_{v,\eta}(\psi)v^{-1}) \in S^{(+)},$$

$$S^{(-)} \ni \psi \to \bar{\psi} = i_v \circ h_{v,\eta}^{-1}(vh_{v,\eta}(i_v\psi)v^{-1}) \in S^{(-)}.$$

The conjugations thus defined have the desired properties.

Lemma 2.7:

(1) $\phi_{v,\eta}(w,\bar{w}) = Q(w)v,$

(2) $\Phi_{v,\eta}^{(+)}(\psi,\bar{\psi}) = \Theta(\psi)\eta,$

(3) $\Phi_{v,\eta}^{(-)}(\psi,\bar{\psi}) = \Theta(\psi)v\eta, \quad (v,\eta) \in S^7 \times S^7.$

Apparently the above conjugations are algebra antiautomorphisms (see Lemma 2.6 and note: $E \perp v$).

The isometries $h_{v,\eta}$; $(v,\eta) \in S^7 \times S^7$, are not the only isometries of $S^{(+)}$ and E vector spaces, as it is enough to note that all isometries of E onto $S^{(+)}$ are given by

$${}_s h_{v,\eta}^{-1} := \tau_s h_{v,\eta}^{-1},$$

where $s \in$ Spin (8). One may define a corresponding sequence for ${}_s h_{v,\eta}^{-1}$ and its associated isometries. Then Lemmas 2.3 and 2.4 are also valid for these general isometries, however, the retransmitted algebra on $S^{(+)}$ (for example) seems to have unity only if $s \in \ker \tau$.

III. TRIALITY PRINCIPLE AND G_2 GROUP

In this section we derive a triality principle for spinors and vectors in terms of appropriate representations of a Spin (8) group. We also supply groups of automorphisms of octonionic structures with a simple geometric interpretation originating from the context created by the point of view proposed in this paper.

(1) Now we concentrate our considerations on $(S^{(+)},\Phi^{(+)})$ and (E,ϕ) algebras. Corresponding results are easily transferred to $(S^{(-)},\Phi^{(-)})$ algebra via a commutative diagram. Let us proceed then to construct various representations of Spin (8) group in terms of End $(S^{(+)})$ and End (E). For that purpose consider

$$\text{Spin}(8) \ni S \quad \begin{array}{l} \rho_s := i_v \circ \tau_s \circ i_v \\ \kappa_s := h_{v,\eta}^{-1} \circ \pi_s \circ h_{v,\eta} \in \text{Aut}(S^{(+)}) \\ \tau_s \end{array}$$

and

$$\text{Spin}(8) \ni S \quad \begin{array}{l} L_s := h_{v,\eta} \circ \tau_{vsv} \circ h_{v,\eta}^{-1} \\ \pi_s := s(\ )s^{-1} \in \text{Aut}(E) \\ R_s := h_{v,\eta} \circ \tau_s \circ h_{v,\eta}^{-1} \end{array}$$

two triples of different representations (having different kernels). The fundamental relation between them is given by

Proposition 3.1: (Triality Principle)

(1) $\rho_s(\Phi_{v,\eta}^{(+)}(\psi,\psi')) = \Phi_{v,\eta}^{(+)}(\kappa_s(\psi),\tau_s(\psi')),$

(2) $L_s(\phi_{v,\eta}(w,w')) = \phi_{v,\eta}(\pi_s(w),R_s(w')),$

for any $s \in$ Spin (8) and arbitrary $\psi,\psi' \in \mathscr{S}^{(+)}$, $w,w' \in E$; $(v,\eta) \in S^7 \times \mathscr{S}^7$.

Proof:

$$(1)\ \rho_s(\Phi^{(+)}(\psi,\psi')) = vsv\, vh_{v,\eta}(\psi)\psi' = i_v(sh_{v,\eta}(\psi)\psi')$$

$$= i_v(h_{v,\eta} \circ h_{v,\eta}^{-1} sh_{v,\eta}(\psi)s^{-1}\psi')$$

$$= i_v(h_{v,\eta} \circ h_{v,\eta}^{-1} \circ \pi_s \circ h_{v,\eta}(\psi)\tau_s\psi')$$

$$= i_v(h_{v,\eta}(\kappa_s(\psi))\tau_s(\psi')) = \Phi_{v,\eta}^{(+)}(\kappa_s(\psi),\tau_s(\psi')).$$

(2) is obtained from (1) by straightforward application of $h_{v,\eta}$ isomorphism. ■

The second part of this section is devoted to groups of automorphisms which we define to be

Definition 3:

$$\text{Aut}(S^{(+)}) = \{s \in \text{Spin}(8);\ \tau_s \Phi_{v,\eta}(\psi,\psi') = \Phi_{v,\eta}^{(1)}(\tau_s\psi,\tau_s\psi');\ \psi,\psi' \in S_+\},$$

$$\text{Aut}(E) = \{s \in \text{Spin}(8);\ \pi_s \phi_{v,\eta}(w,w') = \phi_{v,\eta}(\pi_s(w),\pi_s(w'));\ w,w' \in E\},$$

$$(v,\eta) \in S^7 \times S^7.$$

For considerations to follow, two lemmas are of primary importance:

Lemma 3.1:

(1) If $s \in \text{Aut}(S^{(+)})$, then $\Phi_{v,\eta}^{(+)} = \Phi_{\pi_s^{-1}(v),\eta}^{(+)}$,

(2) If $s \in \text{Aut}(E)$, then $\phi_{v,\eta} = \phi_{v,\tau_s^{-1}\eta}$.

Proof: (1) Note that

$$\tau_s \Phi_{v,\eta}^{(+)}(\psi,\eta) = \tau_s\psi = \Phi_{v,\eta}^{(+)}(\tau_s\psi,\tau_s\eta)$$

implies that $\tau_s\eta = \eta$.

The defining identity of $\text{Aut}(\mathscr{S}^{(+)})$ rewritten explicitly

$$svh_{v,\eta}(\psi)\psi'$$
$$= \tau_s\Phi_{v,\eta}^{(+)}(\psi,\psi') = \Phi_{v,\eta}^{(+)}(\tau_s\psi,\tau_s\psi') = vh_{v,\eta}(s\psi)s\psi'$$
$$= v\beta(\psi)s^{-1}vE_a\eta E_a s\psi' = v\beta(\psi)s^{-1}vss^{-1}E_a s\eta E_a s\psi'$$
$$= vs\beta(\psi)\pi_s^{-1}(v)s^{-1}E_a s\eta s^{-1}E_a s\psi'$$
$$= vs\beta(\psi)\pi_{s^{-1}}(v)E_a\eta E_a\psi'$$

and multiplied by s^{-1} leads to

$$\Phi_{v,\eta}^{(+)}(\psi,\psi') = \pi_{s^{-1}}(v)\chi_{\pi_s^{-1}(v),\eta}(\psi)\psi' = \Phi_{\pi_s^{-1}(v),\eta}^{(+)}(\psi,\psi').$$

(2) As above, $\pi_s(v) = v$ is a necessary condition for $s \in \text{Aut}(E)$. Using this fact and the definition of $\text{Aut}(E)$ we get

$$\beta(\eta)w'vwE_a\eta\pi_s(E_a) = \pi_s\phi_{v,\eta}(w,w') = \phi_{v,\eta}(\pi_s w,\pi_s w')$$
$$= (h_{v,\eta} \circ i_v)(sws^{-1}vww's^{-1}\eta)$$
$$= \beta(\eta)sw's^{-1}vsws^{-1}E_a\eta E_a$$
$$= \beta(s^{-1}\eta)w'vw\pi_{s^{-1}}(E_a)(s^{-1}\eta)E_a$$
$$= \beta(s^{-1}\eta)w'vwE_a s^{-1}\eta\pi_s(E_a),$$

which leads immediately to the thesis. ■

Lemma 3.2:

(1) If $\Phi_{v,\eta}^{(+)} = \Phi_{v',\eta}^{(+)}$, then $v = \pm v$;

(2) If $\phi_{v,\eta} = \phi_{v,\eta'}$, then $\eta = \pm\eta$.

Proof: (1) From

$$0 = \Phi_{v,\eta}^{(+)}(\psi,\psi') - \Phi_{v,\eta}^{(+)}(\psi,\psi') = (vh_{v,\eta}(\psi) - v'h_{v\psi,\eta}(\psi))\psi'$$

for arbitrary ψ', we deduce that the endomorphism acting on ψ' belongs to the kernel of representation $\tilde{\tau}$ (of $e^{(+)}(8,0)$ Clifford algebra), i.e., it has the form

$$vh_{v,\eta}(\psi) - v'h_{v',\eta}(\psi') = \tfrac{1}{2}(1-J)c$$

for some $c \in C^+(8,0)$ [see (1.1)]. Note that the right-hand side of the above equation is self-dual, whereas dual to the left-hand side element belongs to the subspace of $\mathscr{C}(8,0)$, which does not contain preimage of the dual. Thus

$$h_{v,\eta}(\psi) - vv'h_{v',\eta}(\psi) = 0$$

for arbitrary $\psi \in \mathscr{S}^{(+)}$, from which, after subtraction of the three-vector part

$$v \wedge v' \wedge h_{v',\eta}(\psi) = 0, \quad \psi \in \mathscr{S}^{(+)},$$

we obtain $v \wedge v' = 0$, i.e., $v = \pm v'$ due to arbitrariness of $\psi \in \mathscr{S}^{(+)}$.

(2) The transformation $h_{v,\eta}^{-1} \circ h_{v,\eta'}$ is orthogonal and we can always find such $s \in \mathrm{Spin}(8)$, that $\tau_s = h_{v,\eta}^{-1} \circ h_{v,\eta'}$, and moreover $\tau_s \eta' = \eta$.

Acting with $h_{v,\eta}^{-1}$ on the equality of octonionic structures we obtain

$$\begin{aligned}\Phi^{(+)}_{v,\eta}&(h_{v,\eta}^{-1}(w),h_{v,\eta}^{-1}(w'))\\ &= h_{v,\eta}^{-1} \circ h_{v,\eta'}(\Phi^{(+)}_{v,\eta'}(h_{v,\eta'}^{-1}(w),h_{v,\eta'}^{-1}(w')))\\ &= \tau_s \Phi^{(+)}_{v,\eta'}(\tau_{s^{-1}} h_{v,\eta}^{-1}(w)\tau_s^{-1} h_{v,\eta}^{-1}(w')).\end{aligned}$$

The last term rewritten explicitly with $\psi,\psi' = h_{v,\eta}^{-1}(w)$, $h_{v,\eta}^{-1}(w')$,

$$\begin{aligned}\tau_s \Phi^{(+)}_{v,\eta'}(\tau_{s^{-1}}\psi,\tau_{s^{-1}}\psi') &= sv\beta(\psi)svs^{-1}sE_a\eta'E_as^{-1}\psi'\\ &= sv\beta(\psi)\pi_s(v)\pi_s(E_a)\eta E_a s^{-1}\psi'\\ &= sv\beta(\psi)\pi_s(v)E_a\eta\pi_s^{-1}(E_a)s^{-1}\psi'\\ &= \pi_s(v)\beta(\psi)\pi_s(v)E_a\eta E_a\psi'\\ &= \Phi^{(+)}_{\pi_s(v),\eta}(\psi,\psi')\end{aligned}$$

leads to the equation

$$\Phi^{(+)}_{v,\eta}(\psi,\psi') = \Phi^{(+)}_{\pi_s(v),\eta}(\psi,\psi'),$$

from which we deduce $\pi_s(v) = \pm v$ (Sec. I).

Composing again $h_{v,\eta}^{-1}$ with an initial identity we get

$$vw\psi' = \pi_s(vw)\psi', \quad \psi' = h_{v,\eta}(w'), \quad w,w' \in E,$$

and after the same discussion on $\ker \tilde{\tau}$ as in (1) and using

$\pi_s(v) = \pm v$ we have

$$\pi_s(w) = \pm w, \quad w \in E.$$

Thus, $s \in \{1, -1, J, -J\}$, i.e., $\eta' = \tau_{s^{-1}}\eta = \pm \eta$. ∎

From Lemmas 3.2 and 3.3 follows an important proposition.

Proposition 3.2:

$$\mathrm{Aut}(S^{(+)}) = \{s \in \mathrm{Spin}(8);\ \tau_s(\eta) = \eta,\ \pi_s(v) = \pm v\},$$

$$\mathrm{Aut}(E) = \{s \in \mathrm{Spin}(8);\ \pi_s(v) = v,\ \tau_s\eta = \pm \eta\}. \quad ∎$$

This proposition enables a clear interpretation of the automorphism groups of $(S^{(+)}, \Phi^{(+)})$ and (E,ϕ) algebras. It is also natural to define

Definition 4:

$$\begin{aligned}\mathrm{Aut}(S^{(+)},E) &= \mathrm{Aut}(S^{(+)}) \cap \mathrm{Aut}(E)\\ &= \{s \in \mathrm{Spin}(8);\ \tau_s\eta = \eta,\ \pi_s(v) = v\}.\end{aligned}$$

This very group $\mathrm{Aut}(S^{(+)},E)$ of simultaneous automorphisms (via respective representations) of $(S^{(+)},\Phi^{(+)})$ and (E,ϕ) algebras has a characteristic property expressed by

Corollary 3.1: The following diagram:

$$\begin{array}{llll} & \tau_s: & {}_hS^{(+)} & \to S^{(+)}_h \\ \mathrm{Aut}(S^{(+)},E) \ni s & & & \\ & \pi_s: & E & \to E \end{array}$$

($h = h_{v,\eta}$, $(v,\eta) \in S^7 \times S^7$) is commutative. ∎

This is the distinguished property; as for $\mathrm{Aut}(E)$ or $\mathrm{Aut}(S^{(+)})$, the corresponding diagrams are no more commutative though, they are "Z_2-commutative" instead, that is to say $h \circ \tau_s = \pm \pi_s \circ h$. This is to be related with the fact that τ and π representations of $\mathrm{Aut}(S^{(+)})$ and $\mathrm{Aut}(E)$ groups are not faithful [kernels being those of Spin (8) representations]. However, we have

Corollary 3.2: The restricted and representations

$$\tau|_{\mathrm{Aut}(S^{(+)})}, \quad \pi|_{\mathrm{Aut}(S^{(+)},E)}$$

are faithful. The $\mathrm{Aut}(S^{(+)},E)$ group of automorphisms we identify with a standard G_2 group.

This is not the largest group for which Corollary 3.1 holds, however it is so, due to the triality principle if in addition one requires representations τ and π to be faithful.

Finally, we describe briefly the Lie algebra G_2 of derivations of octonionic structures. From Proposition 3.1 it follows that

$$G_2 = \{b \in \Lambda^2E; [b,v] = 0,\ b\eta = 0\} \subset \mathrm{so}(8).$$

The first condition, $[b,v] = 0$, subtracts so(7) subalgebra

$$\begin{aligned}\mathrm{so}(7) &= \{b \in \Lambda^2E;\ b = b^{ij}E_i \wedge E_j,\ E_i,E_j \perp v;\ 1 \leqslant i,j \leqslant 7\}\\ &= \Lambda^2E(7,0).\end{aligned}$$

Let

$$\Lambda^pE(7,0) \ni C^{(p)} \to C^{(p)*} \in \Lambda^{7-p}E(7,0)$$

be the Hodge operator in $E(7,0)$. We define the three-vector ϵ

$$\Lambda^3E(7,0) \ni \epsilon := \epsilon^{ijk}E_i \wedge E_j \wedge E_k, \quad 1 \leqslant i < j < k \leqslant 7,$$

where ϵ^{ijk} is that of Lemma 2.6.

Proposition 3.3:

$$G_2 := \{b \in \Lambda^2E(7,0);\ b - b\epsilon^* = 0\}.$$

Proof: The condition $b\eta = 0$ for $b \in G_2$ is equivalent to $b\eta\bar{\eta} = 0$. The self-dual [see (2.1)] spin tensor has the following decomposition in $\mathscr{C}(8,0)$ Clifford algebra:

$$\begin{aligned}\eta\bar{\eta} &= \tfrac{1}{16}(\bar{\eta}\eta + \bar{\eta}\eta J + F^{abcd}E_{abcd}) \quad 1 \leqslant a < b < x < d \leqslant 8\\ &= \tfrac{1}{16}(1 + J + F),\end{aligned}$$

with $F^{abcd} = \bar{\eta}E^aE^bE^cE^d\eta$.

The four-vector F can be decomposed with respect to $E(7,0)$ Euclidean subspace of $E(8,0)$ as follows:

$$F = \epsilon v + f, \quad \epsilon \in \Lambda^2E(7,0),$$

where

$$\Lambda^4E(7,0) \ni f := \bar{\eta}E^iE^jE^kE^l\eta E_{ijkl} \quad 1 \leqslant i < j < k < 1 \leqslant 7.$$

From self-duality of $\eta\bar{\eta}$ with respect to the Hodge star of $E(8,0)$ one gets a seven-dimensional duality relation

$$\epsilon = f^* \text{ or } f = -\epsilon^*.$$

Thus,

$$\eta\bar{\eta} = \tfrac{1}{8}(1 + J + \epsilon v - \epsilon^*)$$

and

$$0 = 8b\eta\bar{\eta} = b + b^*v + b\epsilon v - b\epsilon^*,$$

which is equivalent to $b - b\epsilon^* = 0$. Due to reversibility of all steps of the proof we get the thesis. ■

IV. FINAL REMARKS

Let us focus our attention on "spinorial and vectorial octonions" $(S^{(+)},\Phi^{(+)})$, (E,ϕ). For any $(v,\eta)\,S^7\times S^7$ we have such a pair of algebras. In order to classify somewhat distinct cases, let us introduce three equivalence relations on the 14-dimensional manifold $S^7\times S^7$.

Definition 5:

(1) $(v,\eta)\sim(v',\eta')$ iff $\Phi_{v,\eta} = \Phi_{v',\eta'}$,

(2) $(v,\eta)\approx(v',\eta')$ iff $\phi_{v,\eta} = \phi_{v',\eta'}$,

(3) $(v,\eta)\underset{\approx}{\sim}(v',\eta')$ iff $\phi_{v,\eta} = \phi_{v',\eta'} \wedge \Phi_{v,\eta} = \Phi_{v',\eta'}$.

Introducing then

Definition 6:

(1) $\mathrm{Oct}(S^{(+)}$: $= S^7\times S^7/\sim$,

(2) $\mathrm{Oct}(E)$: $= S^7\times S^7/\approx$,

(3) $\mathrm{Oct}(S^{(+)},E)$: $= S^7\times S^7/\underset{\approx}{\sim}$,

three sets of octgonion structures, one immediately gets

Proposition 4.1:

(1) $\mathrm{Oct}(S^{(+)}) = RP(7)\times S^7$,

(2) $\mathrm{Oct}(E) = S^7\times RP(7)$,

(3) $\mathrm{Oct}(S^{(+)},E) = S^7\times S^7$,

and an obvious corollary.

Corollary 4.1:

$\mathrm{Spin}(8)/\mathrm{Aut}(S^{(+)}) = RP(7)\times S^7$,

$\mathrm{Spin}(8)/\mathrm{Aut}(E) = S^7\times RP(7)$,

$\mathrm{Spin}(8)/\mathrm{Aut}(S^{(+)},E) = S^7\times S^7$,

$\mathrm{Spin}(7)/\mathrm{Aut}(S^{(+)},E) = S^7$. ■

ACKNOWLEDGMENTS

The authors would like to thank A. Z. Jadczyk for drawing their attention to Ref. 9 and to J. Lukierski for useful comments. We also thank P. Morawiec for spontaneous discussions and pointing out Ref. 4.

[1] J. A. Wolf, J. Differential Geom. **7**, 19–44 (1972).

[2] H. Freudenthal, *Oktaven, Ausnahmen Gruppen und Oktavengeometrie* (Mathematisch Institut der Rijksuniversiteit te Utrecht, Utrecht, 1957).

[3] E. Weiss, Math. Z. **44**, 580–611 (1933).

[4] B. A. Rozenfeld, *Noneuclidean Geometries* (Moscow, 1955) (in Russian).

[5] E. Cartan, Bull. Sci. Math. **49**, 361–374 (1925).

[6] Z. Hasiewicz, A. K. Kwasniewski, and P. Morawiec, "On parallelizable spheres, division algebras and Clifford Algebras," Wrocław University preprint, No. 573 (1983).

[7] C. Chevalley, *The Algebraic Theory of Spinors* (Columbia University, New York, 1954).

[8] M. P. Atiyah, R. Bott, and A. Shapiro, Topology **3**, (1), 3–38 (1964).

[9] L. Porteous, *Topological Geometry* (Van Nostrand–Reinhold, London, 1969).

[10] P. Lounesto, Found. Phys. **11**, 721–740 (1981).

[11] J. Lukierski and P. Minneart, "Seven-spheres from Octonions and Geometric Torsion," PTB-128 (April, 1983); Phys. Lett. B (in press).

[12] F. Gursey and H. C. Tze, "Octonionic torsion on s^7 and Englert's compactification of $d = ll$ supergravity," Yale preprint (April, 1983).

[13] F. Englert, Phys. Lett. B **119**, 339 (1982).

[14] M. A. Awada, M. J. Duff, and C. N. Pope, Phys. Rev. Lett. **50**, 294 (1983).

[15] R. d'Auria and P. Fre, Phys. Lett. B **121**, 141 (1983).

[16] T. Dereli *et al.*, "Octonionic Geometry and Simple Supergravity in Eleven Dimensions," University of York preprint (1983); Phys. Lett. B **126**, 33 (1983).

[17] P. van Nieuwenhuizen, CERN preprint Th. 3497 (1983).

The Gel'fand realization and the exceptional representations of SL(2,R)

Debabrata Basu and T. Bhattacharya
Department of Physics, Indian Institute of Technology, Kharagpur 721302, West Bengal, India

(Received 21 March 1984; accepted for publication 1 June 1984)

It is shown that the canonical representation space of Gel'fand and co-workers is particularly appropriate for problems requiring explicit reduction under the noncompact SO(1,1) and E(1) bases for both the principal and exceptional series of representations of SL(2,R). We use this realization to set up complete orthonormal sets of eigendistributions corresponding to the three subgroup reductions, namely, SL(2,R)$\supset$SO(1,1), SL(2,R)$\supset$E(1), and SL(2,R)$\supset$SO(2), and evaluate the unitary transformations connecting these reductions. These overlap matrix elements appear as the applications of these distributions to a set of well-defined test functions. Using the rigorous theory of analytic continuation we show that the results for the exceptional representations have the same analytic forms as the corresponding results for the principal series. Some of these results are essential prerequisites for the solution of the Clebsch–Gordan problem (series and coefficients) of SL(2,R) in the SO(1,1) basis.

I. INTRODUCTION

The complete classification of the unitary irreducible representations (UIR's) of SL(2,R) was given by Bargmann.[1] This group, its covering group $\overline{\mathrm{SL}(2,R)}$ and its representations were further studied by Pukanszky,[2] Barut and Fronsdal,[3] Sally, Jr.,[4] and Lang.[5]

The UIR's of SL(2,R) contain two continuous series: first, the principal series labeled by a pair of indices (ϵ, j), where ϵ is a reflection label and takes on the values 0 and 1, and j is a complex number of the form $j = -\frac{1}{2} + is$, $-\infty < s < \infty$; and second, the exceptional series characterized by a single real parameter j varying over a finite interval $-1 < j < 0$. The UIR's, in addition, contain discrete spectra characterized by negative integral or half-integral values of j.

The exceptional UIR's of SL(2,R) contain some unfamiliar features and are believed to be really exceptional insofar as they have no analog in the representation theory of compact groups. These representations do not appear in the decomposition of the left regular representation into irreducible representations. "It is, as if, nature has made this object, found no use for it and, therefore, discarded it."[6]

The representations of the principal series were realized by Bargmann in the Hilbert space $L^2(\phi)$ of functions defined on the unit circle. The exceptional series was defined in a similar function space but with a nonlocal metric. Although Bargmann's realization was used in the past in connection with the reduction of the representations of SL(2,R) under the hyperbolic SO(1,1) and the parabolic E(1) subgroups,[7–11] it called for a complicated set of transformations and the resulting bases turn out to be too complicated for further use. This is because Bargmann's realization is really suited to problems requiring reduction under the elliptic SO(2) subgroup.[12] Some simplification has been achieved more recently by using the oscillator realization proposed by Holman and Biedenharn,[13] Barut and Bohm,[14] Moshinsky,[15] Wolf,[16] and Mukunda and co-workers[17], amongst others.[18] In contrast to the canonical realization of Bargmann, the oscillator realization leads to a second-order operator realization of the Lie algebra which on exponentiation yields a parametrized continuum of integral transforms. In a number of previous papers[19] we have used these integral transform realizations of SL(2,R) [as well as of SL(2,C)] to evaluate the matrix elements of operators, Clebsch–Gordan coefficients, etc.

The oscillator realization unfortunately does not yield the exceptional UIR's of SL(2,R). Consequently we could deal only with the principal and discrete series of representations in our previous papers.[19] The object of the present investigation is to supplement the previous results by parallel results for the exceptional UIR's of SL(2,R). For this purpose the representation space of Gel'fand, Graev, and Vilenkin[20] seems to be particularly appropriate. This realization of the principal and exceptional UIR's of SL(2,R) turns out to be a convenient starting point for many practical calculations, particularly those requiring explicit reduction under the hyperbolic SO(1,1) or the parabolic E(1) bases. The reductions lead to a class of distributions which have been studied extensively by Gel'fand and Shilov.[21]

We set up complete orthonormal sets of eigendistributions corresponding to the three subgroup reductions and evaluate the various overlap matrix elements which appear as the applications of these distributions to a set of well-defined test functions. We use the rigorous theory of analytic continuation of classical analysis to evaluate the SO(2)–SO(1,1) and SO(2)–E(1) overlaps which are given in terms of Gauss' hypergeometric and Whittaker functions, respectively. These results for the exceptional representations turn out to have the same analytic form as the corresponding results for the principal series previously obtained by us,[19(a),(b),(d)] by Kalnins,[22] and by Montgomery and O'Raifeartaigh.[23] The SO(1,1)–E(1) overlaps, on the other hand, turn out to be distributions of the type $x_{\pm}^{\mu}$.

Some of the results of this paper are essential prerequisites for the solution of the Clebsch–Gordan problem of SL(2,R) in the hyperbolic SO(1,1) basis involving the coupling of the exceptional UIR's which has not, so far, been treated in the literature. This problem is currently under investigation.

II. THE GEL'FAND–GRAEV–VILENKIN REALIZATION

The group SL(2,R) is the group of real matrices

$$g=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$$

with determinant 1 in two dimensions. In the realization of Gel'fand and co-workers the representations of SL(2,R) are constructed in the space $D_{(j,\epsilon)}$ of functions $\phi(x_1,x_2)$ of two real variables x_1 and x_2 satisfying the following requirements: (a) $\phi(x_1,x_2)$ is homogeneous of degree $2j$; (b) $\phi(x_1,x_2)$ has a definite parity $(-1)^\epsilon$, where $\epsilon=0$ or 1; and (c) $\phi(x_1,x_2)$ is infinitely differentiable everywhere except at the origin. The requirement of homogeneity ensures that $D_{(j,\epsilon)}$ can be realized as the space of functions $f(x)$ of a single real variable $x=x_1/x_2$. Further restrictions on j and ϵ are obtained if the representations are required, in addition, to be unitary. The condition of unitarity leads to the following classes of representations.

(*a*) *Principal series*: The representations of the principal series are defined by

$$T_g^{(j,\epsilon)}f(x)=|\beta x+\delta|^{2j}\,\mathrm{sgn}^\epsilon(\beta x+\delta)f\left(\frac{\alpha x+\gamma}{\beta x+\delta}\right), \tag{2.1}$$

where $j=-\frac{1}{2}+is$, $-\infty<s<\infty$ and $\epsilon=0$ or 1. These representations are unitary in $L^2(R)$, i.e., with respect to the scalar product

$$(f_1,f_2)=\int f_1(x)\bar{f}_2(x)dx\,. \tag{2.2}$$

(*b*) *Exceptional series*: The representations of the exceptional series correspond to $\epsilon=0$ and these are defined by

$$T_g^{(j)}f(x)=|\beta x+\delta|^{2j}f\left(\frac{\alpha x+\gamma}{\beta x+\delta}\right), \tag{2.3}$$

where j is a real number lying between 0 and -1. These representations are unitary in the Hilbert space $L^j(R)$ in which the scalar product involves a nonlocal metric:

$$(f_1,f_2)=\frac{1}{\Gamma(-2j-1)}\int f_1(x_1)|x_1-x_2|^{-2j-2}\times\bar{f}_2(x_2)dx_1\,dx_2\,. \tag{2.4}$$

The above integral converges in the usual sense for $-1<j<-\frac{1}{2}$. For $-\frac{1}{2}<j<0$ it is to be understood in the sense of its regularization. These two sectors, however, are unitarily equivalent.

(*c*) *Discrete series*: The representations of the discrete series or the so-called analytic representations are characterized by $j=-\frac{1}{2},-1,\ldots$. These representations are unitary in the Hilbert space of functions analytic in the half-planes $\mathrm{Im}\,z>0$ and $\mathrm{Im}\,z<0$. The former are called the positive discrete series and the latter the negative discrete series. Since, in this paper, we shall be mainly concerned with the exceptional representations we do not give the explicit operator form of these representations.

The generators of SL(2,R), in this realization, are first-order differential operators in $L^j(R)$ and these are given by

$$\begin{aligned}J_1&=-i\left[\frac{(1-x^2)}{2}\frac{d}{dx}+jx\right],\\ J_2&=i\left[x\frac{d}{dx}-j\right],\\ J_3&=i\left[\frac{(1+x^2)}{2}\frac{d}{dx}-jx\right].\end{aligned} \tag{2.5}$$

Here J_1 and J_2 generate the boost hyperbolic subgroups SO(1,1), J_3 generates the compact elliptic subgroup SO(2), and $T_\pm=J_1\pm J_3$ generate the Euclidean parabolic subgroups E(1).

The representations are unitary and consequently the generators (2.5) are Hermitian under the scalar product (2.4). It is sometimes more convenient to express the scalar product in the "momentum space" by introducing the Fourier transform

$$\begin{aligned}f(x)&=\frac{1}{2\pi}\int h(\rho)e^{-ix\rho}\,d\rho\,,\\ h(\rho)&=\int f(x)e^{ix\rho}\,dx\,.\end{aligned} \tag{2.6}$$

We follow here the definition of Fourier transform as adopted by Gel'fand and Shilov[24] because we shall frequently deal with the Fourier transforms of the distributions that have been extensively treated by them.

In the momentum space the nonlocal metric disappears and the scalar product takes the form

$$(f_1,f_2)=-\frac{\sin\pi j}{\pi}\int h_1(\rho)\bar{h}_2(\rho)|\rho|^{2j+1}\,d\rho\,, \tag{2.7}$$

where h_1 and h_2 are the Fourier transforms of f_1 and f_2, respectively. Since $-1<j<0$ for the exceptional UIR's the squared norm (f,f) is positive definite.

Following Gel'fand and Shilov[21] we now give the definitions and basic properties of some distributions which will appear in this work. We start with the functions $x_\pm^\mu$. Let us first consider x_+^μ which is defined by

$$x_+^\mu=\begin{cases}x^\mu, & \text{for } x>0\,,\\ 0, & \text{for } x\leqslant 0\,.\end{cases} \tag{2.8}$$

We now define the distribution x_+^μ as follows. For $\mathrm{Re}\,\mu>-1$

$$(x_+^\mu,\phi)=\int_0^\infty x^\mu\phi(x)dx\,, \tag{2.9}$$

where $\phi(x)$ is a sufficiently smooth test function with a compact support. For other values of μ this distribution is defined by the appropriate regularization procedure. In this work, however, the restriction $\mathrm{Re}\,\mu>-1$ is obeyed and the regularization of the integral will not be needed.

Similarly we consider the distribution corresponding to

$$x_-^\mu=\begin{cases}|x|^\mu, & \text{for } x<0\,,\\ 0, & \text{for } x\geqslant 0\,.\end{cases} \tag{2.10}$$

For $\mathrm{Re}\,\mu>-1$ this function defines a regular functional

$$(x_-^\mu,\phi)=\int_{-\infty}^0|x|^\mu\phi(x)=\int_0^\infty x^\mu\phi(-x)dx\,. \tag{2.11}$$

We now introduce the distributions $(x\pm i0)^\mu$ in the following way. We define

$$(x+i0)^\mu=\lim_{y\to+0}(x+iy)^\mu=\begin{cases}e^{i\mu\pi}|x|^\mu, & \text{for } x<0\,,\\ x^\mu, & \text{for } x>0\,,\end{cases} \tag{2.12a}$$

$$(x-i0)^\mu=\lim_{y\to+0}(x-iy)^\mu=\begin{cases}e^{-i\mu\pi}|x|^\mu, & \text{for } x<0\,,\\ x^\mu, & \text{for } x>0\,.\end{cases} \tag{2.12b}$$

These functions exist for all μ and define regular functionals for $\mathrm{Re}\,\mu>-1$.

In this work we shall have occasion to deal with the Fourier transforms of these distributions which are given by[25]

$$h_\mu^\pm(\rho)=\int_{-\infty}^{\infty} e^{ix\rho}(x\pm i0)^\mu\,dx$$
$$=\frac{2\pi e^{\pm i\pi\mu/2}}{\Gamma(-\mu)}\rho_\mp^{-\mu-1}. \tag{2.13}$$

III. SUBGROUP-ADAPTED EIGENDISTRIBUTIONS

A. SL(2,R)⊃SO(1,1)

The simplest operator in the hyperbolic orbit is J_2 and its eigenfunctions can be shown to be the distributions

$$f_\lambda^\pm=N_\lambda^\pm(x\pm i0)^{j-i\lambda}, \tag{3.1}$$

where

$$N_\lambda^\pm=\left[-\frac{\Gamma(-j+i\lambda)\Gamma(-j-i\lambda)}{8\pi^2\sin\pi j e^{\pm\pi\lambda}}\right]^{1/2}. \tag{3.2}$$

These are, as shown below, Dirac orthonormal and complete.

(*a*) *Orthonormality*: To establish the orthonormality we introduce the Fourier transforms of the eigendistributions (3.1) as given by Eq. (2.13) in the scalar product (2.7). Thus,

$$(f_\lambda^\epsilon,f_{\lambda'}^{\epsilon'})=\delta_{\epsilon\epsilon'}\delta(\lambda-\lambda'), \tag{3.3}$$

where $\epsilon=\pm$.

(*b*) *Completeness*: We shall now show that the normalized eigendistributions (3.1) form a complete set. The completeness condition, however, involves the nonlocal metric and requires

$$\int d\lambda\frac{1}{\Gamma(-2j-1)}\int dx''|x'-x''|^{-2j-2}\times\sum_{\epsilon=\pm}\bar f_\lambda^\epsilon(x'')f_\lambda^\epsilon(x)=\delta(x-x'). \tag{3.4a}$$

This can be written as

$$\int d\lambda\sum_{\epsilon=\pm}\theta_\lambda^\epsilon(x')f_\lambda^\epsilon(x)=\delta(x-x'), \tag{3.4b}$$

where

$$\theta_\lambda^\pm(x')=\frac{N_\lambda^\pm}{\Gamma(-2j-1)}\times\int dx''|x'-x''|^{-2j-2}(x''\mp i0)^{j+i\lambda}. \tag{3.5}$$

Inserting the Fourier transforms of the eigendistributions $(x\mp i0)^{j+i\lambda}$ we obtain after some calculations

$$\theta_\lambda^\pm(x')=-\frac{2N_\lambda^\pm e^{\mp(i\pi/2)(j+i\lambda)}\sin\pi j}{\Gamma(-j-i\lambda)}\times\int\rho_\pm^{j-i\lambda}e^{-ix'\rho}\,d\rho. \tag{3.6}$$

Comparison of Eq. (3.6) with Eq. (2.13) reveals that the rhs, apart from constant factors, is the Fourier transform of $(x'\mp i0)^{-j-1+i\lambda}$. The nonlocal metric, therefore, effectively plays the same role that the complex conjugation does in the principal series. Both change j to $-j-1$. We thus obtain

$$\theta_\lambda^\pm(x')=\frac{-2\pi\sin\pi j e^{\mp(i\pi/2)(2j+1)}N_\lambda^\pm(x'\mp i0)^{-j-1+i\lambda}}{\sin\pi(j+1-i\lambda)\Gamma(-j-i\lambda)\Gamma(-j+i\lambda)}. \tag{3.7}$$

Inserting Eq. (3.7) in the lhs of the completeness condition (3.4b) we obtain

$$\int d\lambda\sum_{\epsilon=\pm}\theta_\lambda^\epsilon(x')f_\lambda^\epsilon(x)=\frac{1}{2\pi}\left[e^{i\pi(2j+1)}\int d\lambda\frac{(x'-i0)^{-j-1+i\lambda}(x+i0)^{j-i\lambda}}{e^{2\pi\lambda}+e^{-i\pi(2j+1)}}+e^{i\pi(2j+1)}\int d\lambda\frac{(x'+i0)^{-j-1+i\lambda}(x-i0)^{j-i\lambda}}{e^{-2\pi\lambda}+e^{i\pi(2j+1)}}\right]. \tag{3.8}$$

Taking $x,x'>0$ and setting $x=e^t$, $x'=e^{t'}$, we obtain

$$\int d\lambda\sum_{\epsilon=\pm}\theta_\lambda^\epsilon(x')f_\lambda^\epsilon(x)=e^{j(t-t')-t'}\int d\lambda\left[\frac{e^{-i\pi(2j+1)}}{e^{2\pi\lambda}+e^{-i\pi(2j+1)}}+\frac{e^{i\pi(2j+1)}}{e^{-2\pi\lambda}+e^{i\pi(2j+1)}}\right]e^{-i\lambda(t-t')}. \tag{3.9}$$

Since the term in the bracket sums up to unity we have

$$\int d\lambda\sum_{\epsilon=\pm}\theta_\lambda^\epsilon(x')f_\lambda^\epsilon(x)=e^{-t}\delta(t-t')=\delta(x-x').$$

The same result holds when x and x' are both negative. When one of x and x' is positive and the other negative the rhs of (3.8) is zero. This completes the proof of the completeness condition (3.4).

B. SL(2,R)⊃E(1)

The two operators generating conjugate E(1) subgroups are $T_\pm=J_1\pm J_3$. The simpler operator in the parabolic orbit is, however, $T_-=J_1-J_3$ and the normalized eigendistribution of T_- is given by

$$e_\sigma(x)=N_\sigma e^{i\sigma x}, \tag{3.10}$$

where

$$N_\sigma=[-(|\sigma|^{-2j-1}/4\pi\sin\pi j)]^{1/2}. \tag{3.11}$$

These are also Dirac orthonormal and complete.

(*a*) *Orthonormality*: To establish the orthonormality we notice

$$(e_{\sigma'},e_\sigma)=N_\sigma N_{\sigma'}\int e^{i\sigma' x}\theta_\sigma(x)dx, \tag{3.12}$$

where

$$\theta_\sigma(x)=\frac{1}{\Gamma(-2j-1)}\int|x-x'|^{-2j-2}e^{-i\sigma x'}dx' \tag{3.13a}$$
$$=-2\sin\pi j|\sigma|^{2j+1}e^{-i\sigma x}. \tag{3.13b}$$

Inserting Eq. (3.13b) in Eq. (3.12) we obtain

$$(e_{\sigma'},e_\sigma)=\delta(\sigma'-\sigma). \tag{3.14}$$

(*b*) *Completeness*: The completeness condition now requires

$$\int d\sigma\,N_\sigma\theta_\sigma(x')e_\sigma(x)=\delta(x'-x), \tag{3.15}$$

where $\theta_\sigma(x)$ is given by (3.13a). Inserting θ_σ from Eq. (3.13b) in the lhs of (3.15) and using Eqs. (3.10) and (3.11) the completeness condition can be immediately obtained.

C. SL(2,R)⊃SO(2)

The compact SO(2) subgroup is generated by J_3 and the eigenfunctions (proper) are given by

$$g_m = N_m(1+ix)^{j-m}(1-ix)^{j+m}, \tag{3.16}$$

where

$$N_m = \left[-\frac{\Gamma(m-j)2^{-2j-2}}{\pi \sin \pi j \Gamma(m+j+1)} \right]^{1/2}. \tag{3.17}$$

(*a*) *Orthonormality*: To establish the orthonormality of g_m under the scalar product (2.4) we note that

$$(g_{m'}, g_m) = N_{m'} N_m \int \psi^j_{m'}(x) \theta^j_m(x) dx, \tag{3.18}$$

where

$$\psi^j_m(x) = (1+ix)^{j-m}(1-ix)^{j+m}, \tag{3.19}$$

and

$$\theta^j_m(x) = \frac{1}{\Gamma(-2j-1)} \int |x-x'|^{-2j-2} \bar{\psi}^j_m(x') dx'.$$

Inserting the Fourier transform of ψ^j_m,

$$\psi^j_m(x) = \frac{1}{2\pi} \int e^{-ix\rho} h^j_m(\rho) d\rho, \tag{3.20}$$

where

$$h^j_m(\rho) = \int (1+ix)^{j-m}(1-ix)^{j+m} e^{ix\rho}\, dx, \tag{3.21}$$

we obtain after some calculations

$$\theta^j_m(x) = -\frac{\sin \pi j}{\pi} \int \bar{h}^j_m(\rho) |\rho|^{2j+1} e^{ix\rho}\, d\rho. \tag{3.22}$$

The Fourier transform $h^j_m(\rho)$ which is calculated in the next section [see Eq. (4.23)] is real and satisfies

$$h^j_m(\rho)|\rho|^{2j+1} = 2^{2j+1} \frac{\Gamma(m+j+1)}{\Gamma(m-j)} h^{-j-1}_m(\rho). \tag{3.23}$$

Thus

$$\theta^j_m(x) = -2^{2j+2} \sin \pi j \frac{\Gamma(m+j+1)}{\Gamma(m-j)} \bar{\psi}^{-j-1}_m(x). \tag{3.24}$$

Substituting Eq. (3.24) in (3.18) we have

$$(g_{m'}, g_m) = -2^{2j+2} \frac{\sin \pi j \Gamma(m+j+1)}{\Gamma(m-j)} N_{m'} N_m \times \int \frac{dx}{1+x^2} \left(\frac{1+ix}{1-ix} \right)^{m-m'} = \delta_{m'm}. \tag{3.25}$$

(*b*) *Completeness*: The completeness of the compact eigenbases requires that

$$\sum_m \frac{1}{\Gamma(-2j-1)} \int dx'' |x'-x''|^{-2j-2} \bar{g}_m(x'') g_m(x) = \delta(x'-x). \tag{3.26a}$$

This requirement can be written as

$$\sum_{m=-\infty}^{\infty} N^2_m \theta^j_m(x') \psi^j_m(x) = \delta(x'-x). \tag{3.26b}$$

Inserting Eq. (3.24) (with x replaced by x') into the lhs of (3.26b) and setting $x = \tan(\phi/2)$, $x' = \tan(\phi'/2)$ we obtain after some calculation

$$\begin{aligned} \sum N^2_m \theta^j_m(x') \psi^j_m(x) &= \left(\sec^2 \frac{\phi'}{2} \right)^{-j-1} \left(\sec^2 \frac{\phi}{2} \right)^j \sum_{m=-\infty}^{\infty} e^{im(\phi'-\phi)} \\ &= 2\cos^2(\phi/2) \delta(\phi'-\phi) = \delta(x'-x). \end{aligned}$$

This completes the proof of the completeness of the elliptic SO(2) bases.

IV. UNITARY TRANSFORMATIONS CONNECTING THE THREE REDUCTIONS

A. SO(2)–SO(1,1) overlap

This overlap is the coefficient of expansion of the compact basis (3.16) in terms of the eigendistributions (3.1) and is given by

$$E^{\pm}_{m\lambda} = (g_m, f^{\pm}_\lambda) \tag{4.1a}$$

$$= N_m \int (1+ix)^{j-m}(1-ix)^{j+m} \theta^{\pm}_\lambda(x) dx, \tag{4.1b}$$

where the distribution $\theta^{\pm}_\lambda(x)$ is defined by Eq. (3.5) and is, therefore, given by (3.7). We, therefore, obtain

$$E^{\pm}_{m\lambda} = \frac{2\pi \sin \pi j e^{\mp (i\pi/2)(2j+1)} N_m N^{\pm}_\lambda}{\sin \pi(j-i\lambda) \Gamma(-j-i\lambda) \Gamma(-j+i\lambda)} K^{\pm}_{m\lambda}, \tag{4.2}$$

where

$$K^{\pm}_{m\lambda} = \int_{-\infty}^{\infty} (1+ix)^{j-m}(1-ix)^{j+m}(x \mp i0)^{-j-1+i\lambda}\, dx. \tag{4.3}$$

This integral may be thought of as representing the application of $(x \mp i0)^{-j-1+i\lambda}$ to the test function $(1+ix)^{j-m}(1-ix)^{j+m}$.

Taking $K^{+}_{m\lambda}$, for example, we note that on introducing $z = \frac{1}{2}(1-ix)$ the rhs of (4.3) can be written as a line integral along $\operatorname{Re} z = \frac{1}{2}$. We, therefore, obtain

$$K^{+}_{m\lambda} = -i2^{2j+1} J_{m\lambda}, \tag{4.4}$$

where

$$J_{m\lambda} = \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} z^{j+m}(1-z)^{j-m} \times [-i(1+0-2z)]^{-j-1+i\lambda}\, dz. \tag{4.5}$$

The only singularities of the integrand are the branch points at $z=0$, $z=1$, and $z = \frac{1}{2}+0$. We note that the branch point at $z = \frac{1}{2}+0$ is away from the path of integration ($\operatorname{Re} z = \frac{1}{2}$). The integrand is single valued and analytic in the z-plane assumed cut along the negative real axis from 0 to $-\infty$ and along the positive real axis from $\frac{1}{2}+0$ to ∞. The cut for the second factor of the integrand is likewise taken from 1 to ∞. If we, therefore, choose, as shown in the Fig. 1, a closed contour Σ, by Cauchy's theorem

$$\oint_\Sigma z^{j+m}(1-z)^{j-m}[-i(1+0-2z)]^{-j-1+i\lambda}\, dz = 0. \tag{4.6}$$

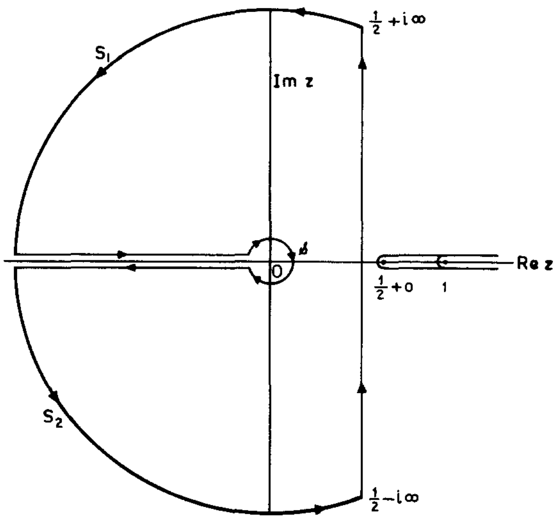


FIG. 1. The contour Σ.

Since the integrals over the large circular arcs S_1 and S_2 vanish, we have

$$J_{m\lambda} = -e^{-(i\pi/2)(-j-1+i\lambda)} I_{m\lambda}, \tag{4.7}$$

where

$$I_{m\lambda} = \int_C z^{j+m}(1-z)^{j-m}(1-2z)^{-j-1+i\lambda}\,dz. \tag{4.8}$$

Here C stands for the part of Σ formed by the small circle s of radius ϵ around the origin and the branch cut from $-\epsilon$ to $-\infty$.

We now use the rigorous theory of analytic continuation to show that $I_{m\lambda}$ is essentially a contour integral representation of a hypergeometric function of argument -1. The procedure is an adaptation of Cauchy's method of analytic continuation of functions defined by definite integrals outside the region of convergence of the integrals. We start with[26]

$$\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^\infty s^{b-1}(1+s)^{a-c}(1+2s)^{-a}\,ds = F(a,b;c;-1), \tag{4.9}$$

which is valid for $\operatorname{Re} c > \operatorname{Re} b > 0$.

The condition $\operatorname{Re} b > 0$ can be relaxed if we recast the above equation in terms of a contour integral over C. Although the line integral does not converge for $\operatorname{Re} b \leqslant 0$, the contour integral

$$\int_C z^{b-1}(1-z)^{a-c}(1-2z)^{-a}\,dz \tag{4.10}$$

continues to define an analytic function of b because the divergence at the origin is canceled by the contribution of the small circle s around the origin. The contour C in Eq. (4.8), therefore, removes all restrictions on m. Now, for $\operatorname{Re} b > 0$, the contour integral (4.10) reduces to the line integral

$$\int_C z^{b-1}(1-z)^{a-c}(1-2z)^{-a}\,dz = 2i\sin\pi(b-1)\int_0^\infty s^{b-1}(1+s)^{a-c}(1+2s)^{-a}\,ds. \tag{4.11}$$

Since the contour integral defines an analytic function of b and reduces to the line integral [rhs of (4.11)] for $\operatorname{Re} b > 0$, we obtain, by the principle of analytic continuation,

$$F(a,b;c;-1) = \frac{\Gamma(c)}{2i\sin\pi(b-1)\Gamma(b)\Gamma(c-b)} \times \int_C z^{b-1}(1-z)^{a-c}(1-2z)^{-a}\,dz, \tag{4.12}$$

for $\operatorname{Re} c > \operatorname{Re} b$ and arbitrary b. Comparing (4.12) and (4.8) we finally have

$$I_{m\lambda} = \frac{2i\sin\pi(j+m)\Gamma(j+m+1)\Gamma(-j-i\lambda)}{\Gamma(m+1-i\lambda)} \times F(j+1-i\lambda, j+m+1; m+1-i\lambda; -1). \tag{4.13}$$

Using Eqs. (4.2)–(4.8) and Eq. (4.13) we can now immediately write down the overlap matrix element

$$E^{+}_{m\lambda} = \frac{(-)^m \pi^{-1/2} 2^{j+1/2}\sin\pi j e^{-i\pi j/2}}{\sin\pi(j-i\lambda)\Gamma(m+1-i\lambda)} \times \left[\frac{\Gamma(m-j)\Gamma(m+j+1)\Gamma(-j-i\lambda)}{\Gamma(-j+i\lambda)}\right]^{1/2} \times F(j+1-i\lambda; j+m+1; n+1-i\lambda; -1). \tag{4.14}$$

Similarly

$$E^{-}_{m\lambda} = e^{i\pi j} E^{+}_{-m\lambda}. \tag{4.15}$$

The SO(2)–SO(1,1) overlap matrix elements, therefore, involve only one hypergeometric function and are simpler than those for the principal series appearing in our previous paper.[19(d)] They are, of course, unitarily equivalent.

B. SO(2)–E(1) overlap

This overlap matrix element is given by

$$F_{m\sigma} = (g_m, e_\sigma) = N_m N_\sigma \int (1+ix)^{j-m}(1-ix)^{j+m}\theta_\sigma(x)\,dx, \tag{4.16}$$

where the distribution $\theta_\sigma(x)$ is defined by Eq. (3.13a). Using Eq. (3.13b) we obtain after simplification

$$E_{m\sigma} = -2\sin\pi j|\sigma|^{2j+1} N_m N_\sigma I_{m\sigma}, \tag{4.17}$$

where

$$I_{m\sigma} = \int_{-\infty}^{\infty} (1-ix)^{j-m}(1+ix)^{j+m} e^{i\sigma x}\,dx. \tag{4.18}$$

For the evaluation of $I_{m\sigma}$ we consider the two cases $\sigma > 0$ and $\sigma < 0$ separately. We first consider $\sigma > 0$. Putting as before, $\frac{1}{2}(1+ix) = z$, the integral can be written as a line integral along $\operatorname{Re} z = \frac{1}{2}$ and we have

$$I_{m\sigma} = -i2^{2j+1}e^{-\sigma}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} e^{2\sigma z} z^{j+m}(1-z)^{j-m}\,dz. \tag{4.19}$$

The only singularities of the integrand in the finite part of the z plane are the branch points at $z=0$ and $z=1$. The integrand is, therefore, single values and analytic in the z plane assumed cut along the negative real axis from 0 to $-\infty$ and along the positive real axis from 1 to ∞. If we, therefore, choose a closed contour Σ similar to the one of the previous subsection we obtain

$$I_{m\sigma} = i2^{2j+1}e^{-\sigma}\int_C e^{2\sigma z}z^{j+m}(1-z)^{j-m}\,dz\,, \tag{4.20}$$

where C, as before, consists of a small circle of radius ϵ centered at the origin and the branch cut from $-\epsilon$ to $-\infty$.

Following the method of the previous subsection and using[27]

$$\int_0^\infty e^{-tx}t^{\alpha}(1+t)^{\beta}\,dt = \Gamma(\alpha+1)e^{x/2}x^{-[(\alpha+\beta)/2]-1}W_{(\beta-\alpha)/2,(\alpha+\beta+1)/2}(x)\,, \tag{4.21}$$

it can be shown that Eq. (4.20) now gives us a contour integral representation of the Whittaker function and this is valid for all values of m. We therefore obtain

$$I_{m\sigma} = -2^{j+1}\sin\pi(j+m)\Gamma(j+m+1) \times\sigma^{-j-1}W_{-m,j+1/2}(2\sigma)\,, \tag{4.22}$$

for $\sigma>0$. For $\sigma<0$ we may proceed similarly and combining the two cases we obtain the general result

$$I_{m\sigma} = -2^{j+1}\sin\pi(j+\operatorname{sgn}\sigma m)\Gamma(j+\operatorname{sgn}\sigma m+1) \times|\sigma|^{-j-1}W_{-\operatorname{sgn}\sigma m,j+1/2}(2|\sigma|)\,. \tag{4.23}$$

It should be noted that the Fourier transform $h^{j}_{m}(\rho)$ defined by Eq. (3.21) of the previous section is equal to $I_{-m\rho}$ and satisfies the identity (3.23).

The SO(2)–E(1) overlap can now be computed and is given by

$$F_{m\sigma} = \frac{(-)^{m+1}}{\pi}2^{j+1}\sin\pi j\left[\frac{\Gamma(m-j)}{\Gamma(m+j+1)}\right]^{1/2} \times\Gamma(j+\operatorname{sgn}\sigma m+1)|\sigma|^{-1/2}W_{-\operatorname{sgn}\sigma m,j+1/2}(2|\sigma|)\,. \tag{4.24}$$

C. SO(1,1)–E(1) overlap

This overlap matrix element is given by

$$G^{\pm}_{\sigma\lambda} = (e_\sigma, f^{\pm}_{\lambda})\,. \tag{4.25}$$

Following the method of Sec. III it can be verified that the $G^{\pm}_{\sigma\lambda}$, apart from certain constant factors, are the Fourier transforms of the distributions $(x\mp i0)^{-j-1+i\lambda}$ and these are given by

$$G^{\pm}_{\sigma\lambda} = \frac{e^{\mp i\pi j/2}}{\sqrt{2\pi}}\left[\frac{\Gamma(-j+i\lambda)}{\Gamma(-j-i\lambda)}\right]^{1/2}\sigma_{\pm}^{-1/2-i\lambda}. \tag{4.26}$$

These are essentially Mellin transform kernels.

ACKNOWLEDGMENT

We thank Dr. G. P. Sastry for many helpful discussions.

[1] V. Bargmann, Ann. Math. **48**, 568 (1947).
[2] L. Pukanszky, Math. Anal. **156**, 96 (1964).
[3] A. O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965).
[4] P. J. Sally, Jr., Bull. Am. Math. Soc. **72**, 269 (1966).
[5] S. Lang, $sl_2(R)$ (Addison–Wesley, Reading, MA, 1975).
[6] L. C. Biedenharn, *Address at the Conference on Noncompact Groups in Particle Physics*, 1966, University of Wisconsin, Milwaukee Proceedings, edited by Y. Chow (Benjamin, New York, 1966), discussion after lecture.
[7] N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 50, 417 (1968); J. G. Kuriyan, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **9**, 2100 (1968).
[8] A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8**, 52 (1968).
[9] N. Mukunda, J. Math. Phys. **10**, 2086 (1969).
[10] N. Mukunda, J. Math. Phys. **10**, 2092 (1969).
[11] G. Lindblad and B. Nagel, Ann. Inst. H. Poincaré **13**, 27 (1970).
[12] L. Pukanszky, Trans. Am. Math. Soc. **100**, 116 (1961); S. D. Majumdar, J. Math. Phys. **17**, 193 (1976); D. Basu and S. D. Majumdar, J. Math. Phys. **20**, 492 (1979).
[13] W. J. Holman and L. C. Biedenharn, Ann. Phys. (N.Y.) **39**, 1 (1966).
[14] A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970); see also A. O. Barut and R. Raczka, *Theory of Groups Representations and Applications* (PWN–Polish Scientific Publisher, Warsaw, 1977); and A. O. Barut, C. K. E. Schneider, and R. Wilson, J. Math. Phys. **20**, 2244 (1979).
[15] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772,1780 (1971); M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901,1634 (1972); M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973); P. Kramer, M. Moshinsky, and T. H. Seligman, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. III.
[16] K. B. Wolf, J. Math. Phys. **15**, 1295,2102 (1974); **17**, 601 (1976); *Integral Transforms in Science and Engineering* (Plenum, New York, 1979); C. P. Boyer and K. B. Wolf, J. Math. Phys. **16**, 1493,2215 (1975); K. B. Wolf, J. Math. Phys. **21**, 680 (1980); K. B. Wolf, in *Proceedings of IX International Colloquium on Group Theoretical Methods in Physics, Lecture Notes in Physics 135* (Springer, New York, 1980).
[17] N. Mukunda and B. Radhakrishnan, J. Math. Phys. **14**, 254 (1973); **15**, 1320,1332,1643,1656 (1974).
[18] L. Infeld and J. Plebansky, Acta Phys. Polon. **14**, 41 (1955); C. Itzykson, J. Math. Phys. **10**, 109 (1969); G. Burdet and M. Perrin, J. Math. Phys. **16**, 1692,2172 (1975); M. Perround, Helv. Phys. Acta **50**, 233 (1977); G. Burdet, M. Perrin and M. Perroud, Commun. Math. Phys. **58**, 241 (1978); E. G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1263 (1974); M. Kashiwara and M. Vergne, Inv. Math. **44**, 1 (1978).
[19] (a) D. Basu, J. Math. Phys. **19**, 1667 (1978); (b) D. Basu and D. Mitra, J. Math. Phys. **21**, 636 (1980); (c) **22**, 946 (1981); (d) D. Basu and K. B. Wolf, J. Math. Phys. **23**, 189 (1982); (e) **24**, 478 (1983).
[20] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, Chap. VII.
[21] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, Chap. I, Secs. 3.2 and 3.6.
[22] E. G. Kalnins, J. Math. Phys. **14**, 654 (1973).
[23] W. Montgomery and L. O'Raifeartaigh, J. Math. Phys. **15**, 380 (1974); see also G. Burdet, M. Perrin, and M. Perroud in Ref. 18.
[24] See Ref. 21, Chap. II, Secs. 1.1 and 2.3.
[25] See Ref. 21, Chap. II, Sec. 2.3.
[26] N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (AMS Translations, Providence, RI, 1968), Vol. 22, Chap. VII, Sec. 1, Subsec. 3.
[27] See Ref. 26, Chap. VIII, Sec. 1, Subsec. 1.

A note on the multiplicity-free unitary irreps of $\overline{\mathrm{SL}(3,R)}$

L. Weaver
Department of Physics, Kansas State University, Manhattan, Kansas 66506
L. C. Biedenharn[a)]
Center for Particle Theory, Department of Physics, University of Texas at Austin, Austin, Texas 78712

(Received 24 January 1984; accepted for publication 22 June 1984)

Unitary irreps of $\overline{\mathrm{SL}(3,\mathbf{R})}$ which contain a representation D^j of SU(2) at most once are known to exist with SU(2) content $j = k + 2n$ ($n = 0,1,2,..$ and $k = 0,\frac{1}{2},1$). Güler, in a recent paper, claims that there also exist multiplicity-free unitary representations with SU(2) content $j = k_0, k_0 + 1, k_0 + 2,...,$ for $k_0 > 3$. We show that such representations do *not* exist for $k_0 > 1$.

The group $\overline{\mathrm{SL}(3,R)}$ is of considerable importance in both nuclear and particle physics as well as in mathematics. In particle physics, the group $\overline{\mathrm{SL}(3,R)}$ appears as a form of dynamical symmetry underlying hadronic Regge bands,[1] implementing thereby a suggestion of Gell-Mann's to enlarge symmetry considerations to encompass Hamiltonians which are functions of transition operators that generate the symmetry. The same concept, applied to nuclear physics,[2] is responsible for nuclear rotational bands [$\overline{\mathrm{SL}(3R)}$ being the group of rotations and volume-preserving deformations of the nucleus]. These applications have led to several investigations to determine all unitary irreps of $\overline{\mathrm{SL}(3R)}$; a summary has been given by Sijacki.[3]

The present paper is concerned with unitary irreps of $\overline{\mathrm{SL}(3,R)}$ that contain a representation D^j of SU(2) at most once. We call these "multiplicity-free representations." This note was prompted by a recent paper by Güler[4] who claims to have constructed a new set of multiplicity-free (unitary) irreps having the SU2 content: $j = k, k + 1, k + 2, ...,$ with $k > 3$. The purpose of the present note is to prove that these new irreps do not exist.

To begin let us recall the approach to the SL(3,R) algebra suggested by Gell-Mann (this will lead to a considerable simplification which will be pointed out below). Consider the angular momentum operators **J**, obeying the commutation relations:

$$\mathbf{J}\times\mathbf{J} = i\mathbf{J}, \tag{1}$$

and adjoin the physical mass quadrupole operator **Q**. Since **Q** is a quadrupole tensor operator it obeys the commutation rule[5]

$$[J_m,Q_q] = \sum_{q'} C^{2\,1\,2}_{q\,m\,q'} Q_{q'} \tag{2}$$

(where $C^{\cdots}_{\cdots}$ are Wigner coefficients[5]). Moreover **Q** (as a function of position operators) obeys the commutation rule

$$[Q_q,Q_{q'}] = 0. \tag{3}$$

The commutation rules (1)–(3) lead to the algebra of the group T^5 @ SU2 [a contraction of $\overline{\mathrm{SL}(3R)}$] for which the multiplicity-free, unitary, "ladder" irreps: $j = j_0, j_0 + 1, ...,$ $j_0 = \frac{1}{2}, 1, \frac{3}{2}, ...,$ certainly do exist.[6] In Ref. 1, an elegant method, inverse to contraction, was given whereby one can obtain the $\overline{\mathrm{SL}(3,R)}$ algebra. For this purpose one defines a new quadrupole operator **T**, which is the *time derivative* of **Q**.

The resulting commutation relations for J and T are

$$\mathbf{J}\times\mathbf{J} = i\mathbf{J}, \tag{1}$$

$$[J_m,T_q] = \sum_{q'} C^{2\,1\,2}_{q\,m\,q'} T_{q'}, \tag{2'}$$

and

$$[T_{+2},T_{-2}] = -4J_0. \tag{4}$$

Equation (4) is the crucial new relation, all other commutators of **T** with **T** follow from this one using Eq. (1) and (2').

To solve these commutation relations, we introduce an angular momentum basis $|JM\rangle$, and explicitly assume that the representation is *multiplicity-free*, that is, each angular momentum J occurs at most once. With respect to this basis, the operator **T** takes the form

$$\langle J'M'|T_q|JM\rangle = (2J'+1)^{-1/2} C^{J\,2\,J'}_{M\,q\,M'} \langle J'\|J\rangle, \tag{5}$$

where we have abbreviated the reduced matrix element $\langle J'\|T\|J\rangle$ by $\langle J'\|J\rangle$.

There is an important physical constraint on the matrix elements of **T** which comes from the interpretation of **T** as the time derivative of **Q**. From this interpretation one has

$$\langle J'\|T\|J\rangle \equiv \langle J'\|\dot{Q}\|J\rangle = (i/\hbar)(E(J') - E(J))\langle J'\|Q\|J\rangle, \tag{6}$$

where $E(J)$ is the energy of the state $|JM\rangle$. It follows that the diagonal matrix elements of **T** must vanish if the diagonal matrix elements of the mass quadrupole operator **Q** are to be finite, that is

$$\langle J\|J\rangle = 0. \tag{7}$$

[This is the simplification mentioned earlier; it is crucial for known physical applications of $\overline{\mathrm{SL}(3,R)}$ symmetry.]

We can now determine the constraints on $\langle J'\|J\rangle$ resulting from the commutation relation, Eq. (4). There are two equations:

$$\Sigma_{\bar{J}}\langle J'\|\bar{J}\rangle\langle\bar{J}\|J\rangle W(J\,2\,J'\,2;\bar{J}\,3) = 0, \tag{8}$$

[a)]On leave from the Department of Physics, Duke University, Durham, North Carolina 27706.

 0022-2488/85/010018-03$02.50

and

$$\Sigma_{\bar{J}} \langle J' \| \bar{J} \rangle \langle \bar{J} \| J \rangle W(J\,2\,J'\,2;\bar{J}\,1) = \sqrt{10/3}\,[J(J+1)(2J+1)]^{1/2} \delta_J^{J'}. \tag{9}$$

[The $W(\cdots)$ in (8) and (9) are Racah coefficients.[5]]

(To understand the meaning of constraints (8) and (9), note that the commutator in Eq. (4)—as the product of two tensor operators of angular momentum 2—can in principle contain angular momenta 0, 1, 2, 3, and 4 (from the abstract product 2×2); the antisymmetry of the product implies that only angular momenta 1[Eq. (9)] and 3[Eq. (8)] can actually occur. Finally, the explicit forms on the right-hand sides of Eqs. (8) and (9) show that the spin 3 operator vanishes and the spin 1 operator is proportional to **J** itself.)

The final constraint on the representation results from unitarity. This implies the relation

$$\langle J' \| J \rangle^* = (-1)^{J'-J} \langle J \| J' \rangle. \tag{10}$$

All multiplicity-free unitary irreps of $\overline{\mathrm{SL}(3,R)}$ in the Gell-Mann realization are determined by solving the constraints: Eqs. (7)–(10). Since the new irreps claimed by Güler[4] actually do satisfy Eq. (7), our use of this simplification is permissible.

One set of solutions is easily obtained: put $\langle J+1 \| J \rangle = 0$ (no $\Delta J = 1$ transitions). Then the solution[7] is

$$\langle J \| J-1 \rangle = 0,$$

$$|\langle J \| J-2 \rangle|^2 = \tfrac{1}{4}(2J)(2J-1)(2J-2). \tag{11}$$

If J_0 is the minimum angular momentum in the representation, then J_0 must be 0, $\frac{1}{2}$, or 1. These three representations[8,9] are

$$0^+: J = 0,2,4,\ldots,$$

$$0^-: J = 1,3,5,\ldots, \tag{12}$$

$$\tfrac{1}{2}^+: J = \tfrac{1}{2},\tfrac{5}{2},\ldots \quad \text{(the "quarkel" irrep).}$$

Now let us look for multiplicity-free unitary irreps for which the lowest angular momentum, call it K, obeys $K > 1$.

Assuming then that $J = K$ is the minimum angular momentum, we write out Eqs. (8) and (9) for $J = J' = K$ and find ($\bar{J} = k-1$ and $k-2$ do not occur in the sum because K is the minimum J; $J = K$ does not occur because $\langle K \| K \rangle = 0$):

$$(-)|\langle K+1 \| K \rangle|^2 W(K\,2\,K\,2;K+1\;3) + |\langle K+2 \| K \rangle|^2 W(K\,2\,K\,2;K+2\;3) = 0, \tag{8'}$$

and

$$(-1)|\langle K+1 \| K \rangle|^2 W(K\,2\,K\,2;K+1\;1) + |\langle K+2 \| K \rangle|^2 W(K\,2\,K\,2;K+2\;1) = [(10)(K)(K+1)(2K+1)/3]^{1/2}. \tag{9'}$$

The explicit algebraic forms for the W's are given in the Appendix. Using these forms we find that Eqs. (8') and (9') determine unique, nonzero, and positive values for $|\langle K+1 \| K \rangle|^2$ and $|\langle K+2 \| K \rangle|^2$:

$$|\langle K+1 \| K \rangle|^2 > 0, \tag{13}$$

$$|\langle K+2 \| K \rangle|^2 > 0. \tag{14}$$

Now write out Eqs. (8) and (9) again, this time for $J = K$, $J' = K+1$. One finds

$$\langle K+1 \| K+2 \rangle \langle K+2 \| K \rangle W(K\,2\,K+1\;2;K+2\;3) = 0 \tag{8''}$$

and

$$\langle K+1 \| K+2 \rangle \langle K+2 \| K \rangle W(K\,2\,K+1\;2;K+2\;1) = 0. \tag{9''}$$

The Racah functions appearing in Eqs. (8'') and (9'') are of special form[5]: $W(a\,b\,c\,d;a+b\,f)$. For W's of this form, the coefficient is *nonzero* and *positive*, provided only that the angular momenta involved obey the triangle conditions [for the W in (8'') these are satisfied because $K > 1$].

Thus for $K > 1$ we conclude

$$\langle K+1 \| K+2 \rangle \langle K+2 \| K \rangle = 0. \tag{15}$$

It follows that

$$\langle K+1 \| K+2 \rangle = 0, \tag{16}$$

since otherwise $\langle K+2 \| K \rangle$ would necessarily be zero which contradicts conclusion (14). [Put differently, Eq. (8') for $K > 1$ implies that if $\langle K+2 \| K \rangle = 0$ then $\langle K+1 \| K \rangle = 0$ which then implies (since $\langle K \| K \rangle = 0$) that the state K is not in the respresentation, contrary to hypothesis.]

Our argument is nearly done. We work out Eq. (8) once again, this time for $J = J' = K+1$. Using Eqs. (7), (10), and (16) we find

$$(-1)|\langle K+1 \| K \rangle|^2 W(K+1\;2\;K+1\;2;K\;3) + |\langle K+1 \| K+3 \rangle|^2 \times W(K+1\;2\;K+1\;2;K+3\;3) = 0. \tag{8'''}$$

The first W in Eq. (8''') is given explicitly in the Appendix and is *nonzero* and negative for $K > 1$; the second W has the special form, hence it is *nonzero* and *positive*. Thus Eq. (8''') implies that both matrix elements in Eq. (8''') are zero. In particular,

$$\langle K+1 \| K \rangle = 0. \tag{17}$$

But if we now recall Eq. (8a) we see that Eq. (17) implies that

$$\langle K+2 \| K \rangle = 0. \tag{18}$$

Since, from Eq. (7), we have $\langle K \| K \rangle = 0$, it follows from Eqs. (17) and (18) *that the state* with $J = K$ *is not in the representation at all*, contrary to our assumption that it was the lowest state.

We conclude *multiplicity-free unitary irreps* obeying the Gell-Mann condition Eq. (7) *with SU(2) content $J=K, K+1, K+2, \ldots$ and $K>1$ do not exist.* Noting once again that the irreps claimed by Güler[4] satisfy Eq. (7) and have the form $J = K, K+1, \ldots, K > 3$ we conclude these irreps cannot exist.

Concluding remark: This work was done in 1971, and was carried out at the suggestion of Professor Gell-Mann. The nonexistence of a class of unirreps did not then seem publishable (after all, there are many nonexistent representations). We believe our argument on nonexistence is now of interest since Güler has published a claim to the contrary.

APPENDIX

We collect here the Racah coefficients[5] used in the text:

$$W(K\,2\,K\,2;K+1\;3)$$
$$=(K+3)$$
$$\times\left[\frac{4(2K-1)(2K-2)}{35(K)(2K+1)(2K+2)(2K+3)(2K+4)}\right]^{1/2}, \tag{A1}$$

$$W(K\,2\,K\,2;K+2\;3)$$
$$=\left[\frac{(2K)(2K-1)(2K-2)}{70(2K+1)(2K+2)(2K+3)(2K+4)}\right]^{1/2}, \tag{A2}$$

$$W(K\,2\,K\,2;K+1\;1)=\frac{(-)(K-2)}{[30K(K+1)(2K+1)]^{1/2}}, \tag{A3}$$

$$W(K\,2\,K\,2;K+2\;1)=\left[\frac{2K}{15(K+1)(2K+1)}\right]^{1/2}, \tag{A4}$$

$$W(K+1\;2\;K+1\;2;K\;3)$$
$$=(-)(K-1)$$
$$\times\left[\frac{8(2K+5)(2K+6)}{35(2K)(2K+1)(2K+2)(2K+3)(2K+4)}\right]^{1/2}. \tag{A5}$$

ACKNOWLEDGMENTS

One of the authors (L. C. B.) was supported in part by the National Science Foundation.

[1]Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965).

[2]L. Weaver and L. C. Biedenharn, Nucl. Phys. A **185**, (1972).

[3]D. Sijacki, J. Math. Phys. **16**, 269 (1975).

[4]Y. Güler, J. Math. Phys. **26**, 21 (1985).

[5]L. C. Biedenharn and J. D. Louck, "Angular Momentum in Quantum Physics," in *Encyclopedia of Mathematics and Its Applications*, Vol. 8, edited by G-C. Rota (Addison-Wesley, Reading, MA, 1981).

[6]L. Weaver, L. C. Biedenharn, and R. Y. Cusson, Ann. Phys. (N. Y.) **77**, 250 (1973).

[7]L. Weaver, thesis, Duke University, 1970.

[8]D. W. Joseph (private communication, 1970); L. Weaver (Ref. 7—but $\frac{1}{2}^+$ irrep was not explicitly mentioned). The constraint $\langle J \| J \rangle = 0$ can be relaxed for these representations, and although the matrix elements change slightly the J-content stays the same.

[9]L. C. Biedenharn, R. Y. Cusson, M. Y. Han, and L. Weaver, Phys. Lett. B **42**, 257 (1972) and Erratum, Phys. Lett. B **73**, 503 (1978). (The Erratum—submitted in 1974 but lost by the editors—withdrew an erroneous "3/2 irrep.")

Multiplicity-free, unitary and nonunitary irreducible representations of $\overline{SL(3,R)}$

Y. Güler
Physics Department, Middle East Technical University, Ankara, Turkey

(Received 22 October 1982; accepted for publication 22 June 1984)

Multiplicity-free, irreducible representations of the group $\overline{SL(3,R)}$ are obtained from SU(2) subgroup representations by a constructive method. It is observed that there exist two series of unitary representations with k contents $\{k_0,k_0+1,k_0+2,...\}k_0\geqslant 3$, $\{k_0,k_0+2,k_0+4,...\}$ $k_0=0, 1, \frac{1}{2}$ and finite-dimensional representations with k content $\{k_0+1,k_0+3,...,k_0+2n+1\}$ $k_0=1, \frac{1}{2}$ $\{2,4,6,...,2n\}$, $n=1,2,...,k_0=0$.

I. INTRODUCTION

The problem of determination of all representations of semisimple, real, noncompact groups has not been fully solved yet. There are mainly two approaches in this respect. The first one is a method used by Gel'fand and Graev.[1] They use the functions defined on the coset spaces of the group as the representation space and determine principal series of unitary, irreducible representations of the noncompact group SL(n,R). The second method is initiated by Chandra[2] who determines all irreducible, unitary representations of a noncompact group using the representations induced from a maximal compact subgroup. Since the representations of the compact group are known, the problem is reduced to the determination of all unitary kernels of the compact subgroup.

As a specific case, the problem of determination of representations of the noncompact group $\overline{SL(3,R)}$ has been attacked by several authors.[3–8] Although various methods are used to obtain several series of representations, they are far from being complete. Even in the most exhaustive works some representations are missing and calculations are based on some assumptions. In this work we obtain all representations, unitary and nonunitary, of $\overline{SL(3,R)}$ by a method initiated by Naimark.[9] This method is simple and direct. Our future research will be on the determination of all representations of SL(n,R) by the same method.

II. THE DETERMINATION OF THE REPRESENTATION SPACE

As it is very well known all irreducible representations of SU(2) subgroup are labeled by a positive integer or half-integer k. Eigenvectors f^k_v, $(v=-k,-k+1,...,k)$ of the Hermitian infinitesimal operators of the SU(2) Lie algebra H_+ and H_- act on canonical basis vectors as

$$H_3 f^k_v = v f^k_v, \tag{2.1}$$

$$H_+ f^k_v = \alpha^k_{v+1} f^k_{v+1}, \tag{2.2}$$

$$H_- f^k_v = \alpha^k_v f^k_{v-1}, \tag{2.3}$$

where

$$\alpha^k_v = [(k+v)(k-v+1)]^{1/2}. \tag{2.4}$$

The representation space M_k is invariant under SU(2) subalgebra. Any representation of the SU(2) subgroup is the direct sum of irreducible representations given by an integer or half-integer k.

Since any representation of $\overline{SL(3,R)}$ contains a representation of the SU(2) subgroup one should consider the space

$$R=\sum_k M_k \tag{2.5}$$

as the multiplicity-free representation space of $\overline{SL(3,R)}$. Hence, the problem of determination of irreducible representations of $\overline{SL(3,R)}$ induced from SU(2) subgroup is reduced to the problem of determination of possible k values contained in R. For this purpose let us consider the irreducible space $M_k\neq 0$. Since all irreducible representations of SU(2) subgroup are finite dimensional, α^k_v should be zero for some definite v. In fact for $v=k$

$$H_+ f^k_k = 0. \tag{2.6}$$

In general, it is proved as a theorem[4] that any eigenvector of H_3 corresponding to the eigenvalue v and satisfying the condition $H^p_+ f=0$ can be written as a linear combination of the eigenstates f^k_v, $k=v,v+1,...,v+p-1$, where p is a nonzero positive integer.

The next step is to determine the possible k values in R using the commutation relations, which are given in the Appendix, and the above theorem. The commutation relation

$$[H_3,T_{+2}] = 2T_{+2} \tag{2.7}$$

gives the vector $T_{+2} f^k_v$ as a multiple of the eigenvector f^k_{v+2}. Besides making use of the commutation relation

$$[H_+,T_{+2}]=0 \tag{2.8}$$

one obtains

$$H^p_+(T_{+2}f^k_v) = T_{+2}(H^p_+ f^k_v). \tag{2.9}$$

Noticing $H^p_+ f^k_v=0$ for $p=k-v+1$, the vector $T_{+2}f^k_v$ can be written as

$$T_{+2}f^k_v = \sum_{j=v+2}^{k+2} \beta^k_{v+2,j} f^j_{v+2}. \tag{2.10}$$

Following the same procedure for the commutation relations

$$[H_-,T_\mu] = [6-\mu(\mu-1)]^{1/2}T_{\mu-1}, \quad \mu=0,\pm 1,\pm 2, \tag{2.11}$$

one determines

$$T_{+1}f_v^k = \frac{1}{2}[H_-,T_{+2}]f_v^k = \sum_{j=v+1}^{k+2} \delta_{v+1,j}^k f_{v+1}^j, \tag{2.12}$$

$$T_0 f_v^k = \frac{1}{6^{1/2}}[H_-,T_{+1}]f_v^k = \sum_{j=v}^{k+2} \eta_{v,j}^k f_v^j, \tag{2.13}$$

$$T_{-1}f_v^k = [H_-,T_0]f_v^k = \sum_{j=v-1}^{k+2} \xi_{v-1,j}^k f_{v-1}^j, \tag{2.14}$$

$$T_{-2}f_v^k = [H_-,T_-]f_v^k = \sum_{j=v-2}^{k+2} \gamma_{v-2,j}^k f_{v-2}^j, \tag{2.15}$$

where the coefficients $\delta_{v+1,j}^k$, $\eta_{v,j}^k$, $\xi_{v-1,j}^k$, and $\gamma_{v-2,j}^k$ are defined as

$$\begin{aligned} &\delta_{v+1,v+1}^k = -\alpha_v^k \beta_{v+1,v+1}^k, \\ &\delta_{v+1,j}^k = \beta_{v+2,j}^k \alpha_{v+2}^j - \alpha_{v+2}^j - \alpha_v^k \beta_{v+1,j}^k, \\ &j = v+2,\dots,k+2, \end{aligned} \tag{2.16}$$

$$\begin{aligned} &\eta_{v,v}^k = -\alpha_v^k \delta_{v,v}^k, \\ &\eta_{v,j}^k = \delta_{v+1,j}^k \alpha_{v+1}^j - \alpha_v^k \delta_{v,j}^k, \quad j = v+1,\dots,k+2, \end{aligned} \tag{2.17}$$

$$\begin{aligned} &\xi_{v-1,v-1}^k = -\alpha_v^k \eta_{v-1,v-1}^k, \\ &\xi_{v-1,j}^k = \eta_{v,j}^k \alpha_v^j - \alpha_v^k \eta_{v-1,j}^k, \quad j = v,\dots,k+2, \end{aligned} \tag{2.18}$$

$$\begin{aligned} &\gamma_{v-2,v-2}^k = -\alpha_v^k \xi_{v-2,v-2}^k, \\ &\gamma_{v-2,j}^k = \alpha_{v-1}^j \xi_{v-1,j}^k - \alpha_v^k \xi_{v-2,j}^k, \\ &j = v-1,\dots,k+2. \end{aligned} \tag{2.19}$$

In order to be able to construct a representation of $\overline{SL(3,R)}$ all the commutation relations given in the Appendix should be satisfied. Hence, our task is now to find all nonzero coefficients such that all the commutation relations except (2.11) are satisfied.

The commutation relation $[T_{-2},H_-]f_v^k = 0$ and Eq. (2.15) gives

$$\sum_{j=v-3}^{k+2} \alpha_v^k \gamma_{v-3,j}^k f_{v-3}^j = \sum_{j=v-2}^{k+2} \alpha_{v-2}^j \gamma_{v-2,j}^k f_{v-3}^j. \tag{2.20}$$

The linear independence of the basis vectors f_v^k requires

$$\alpha_v^k \gamma_{v-3,v-3}^k = 0, \tag{2.21}$$

$$\alpha_v^k \gamma_{v-3,j}^k = \gamma_{v-2,j}^k \alpha_{v-2}^j, \quad j = v-2,\dots,k+2. \tag{2.22}$$

Equation (2.21) implies that

$$\gamma_{v-3,v-3}^k = 0, \quad v < k+1, \tag{2.23}$$

$$\gamma_{v-2,v-2}^k = 0, \quad v < k, \tag{2.24}$$

especially for $v = k-1$

$$\gamma_{k-3,k-3}^k = 0. \tag{2.25}$$

Besides, Eq. (2.22) for $j = v-2$ gives

$$\alpha_v^k \gamma_{v-3,v-2}^k = \gamma_{v-2,v-2}^k \alpha_{v-2}^{v-2} = 0, \tag{2.26}$$

$$\gamma_{v-2,v-1}^k = 0 \quad \text{for} \quad v < k-1, \tag{2.27}$$

and especially for $v = k-2$

$$\gamma_{k-4,k-3}^k = \gamma_{k-4,k-4}^k = 0. \tag{2.28}$$

Following the same reasoning for $j = v-1, v,\dots,k-3$ one concludes that the only nonzero $\gamma_{v-2,j}^k$ are for $j > k-3$. Therefore,

$$T_{-2}f_v^k = \sum_{j=k-2}^{k+2} \gamma_{v-2,j}^k f_{v-2}^k. \tag{2.29}$$

Equations (2.16)–(2.19) imply that the coefficients $\beta_{v+2,j}^k$, $\delta_{v+1,j}^k$, $\eta_{v,j}^k$, and $\xi_{v-1,j}^k$ are all nonzero for $j > k-3$. So

$$T_{-1}f_v^k = \sum_{j=k-2}^{k+2} \xi_{v-1,j}^k f_{v-1}^j, \tag{2.30}$$

$$T_0 f_v^k = \sum_{j=k-2}^{k+2} \eta_{,v,j}^k f_v^j, \tag{2.31}$$

$$T_{+1}f_v^k = \sum_{j=k-2}^{k+2} \delta_{v,j}^k f_{v+1}^j, \tag{2.32}$$

$$T_{+2}f_v^k = \sum_{j=k-2}^{k+2} \beta_{v+2,j}^k f_{v+2}^j. \tag{2.33}$$

The commutation relation $[T_{+2},H_+] = 0$ reduces the number of independent coefficients. In fact,

$$[T_{+2},H_+]f_v^k = \alpha_{v+1}^k \sum_{j=v+3}^{k+2} \beta_{v+3,j}^k f_{v+3}^j - \sum_{j=v+2}^{k+2} \beta_{v+2,j}^k \alpha_{v+3}^j f_{v+3}^j = 0. \tag{2.34}$$

The linear independence of the canonical basis vectors f_v^k implies that

$$\alpha_{v+3}^{v+2} \beta_{v+2,v+2}^k = 0, \tag{2.35}$$

$$\beta_{v+3j}^k \alpha_{v+1}^k = \beta_{v+2,j}^k \alpha_{v+3}^j, \quad j = k-2, k-1,\dots,k+2, \tag{2.36}$$

$$\beta_{v+2,j}^k = \frac{\Pi_{l=v+1}^{j-2} \alpha_l^k}{\Pi_{l=v+3}^j \alpha_l^j} \beta_{j,j}^k, \quad j = k-2,\dots,k+2. \tag{2.37}$$

Keeping in mind that the coefficients $\delta_{v+1,j}^k$, $\eta_{v,j}^k$, $\xi_{v-1,j}^k$, and $\gamma_{v-2,j}^k$ are functions of $\beta_{v+2,j}^k$, all nonzero coefficients are expressed in terms of five coefficients $\beta_{j,j}^k$ $j = k-2, k-1, k, k+1, k+2$,

$$\beta_{v+2,k+2}^k = K^{(1)}(k,v)A_{k+2}, \quad K^{(1)}(k,v) = [(k+v+1)(k+v+2)(k+v+3)(k+v)]^{1/2}, \tag{2.38}$$

$$\beta_{v+2,k+1}^k = K^{(2)}(k,v)B_{k+1}, \quad K^{(2)}(k,v) = [(k-v)(k+v+1)(k+v+2)(k+v+3)]^{1/2}, \tag{2.39}$$

$$\beta_{v+2,k}^k = K^{(3)}(k,v)C_k, \quad K^{(3)}(k,v) = [(k-v)(k-v-1)(k+v+1)(k+v+2)]^{1/2}, \tag{2.40}$$

$$\beta_{v+2,k-1}^k = K^{(4)}(k,v)D_k, \quad K^{(4)}(k,v) = [(k+v+1)(k-v)(k-v-1)(k-v-2)]^{1/2}, \tag{2.41}$$

$$\beta_{v+2,k-2}^k = K^{(5)}(k,v)E_k, \quad K^{(5)}(k,v) = [(k-v)(k-v-1)(k-v-2)(k-v-3)]^{1/2}, \tag{2.42}$$

$$\delta_{v+1,k+2}^k = L^{(1)}(k,v)A_{k+2}, \quad L^{(1)}(k,v) = 4[(k+v+1)(k+v+2)(k+v+3)(k-v+1)]^{1/2}, \tag{2.43}$$

$$\delta_{v+1,k+1}^k = L^{(2)}(k,v)B_{k+1}, \quad L^{(2)}(k,v) = (2k-4v)[(k+v+1)(k+v+2)]^{1/2}, \tag{2.44}$$

$$\delta^k_{v+1,k} = L^{(3)}(k,v)C_k, \quad L^{(3)}(k,v) = -(4v+2)[(k-v)(k+v+1)]^{1/2}, \tag{2.45}$$
$$\delta^k_{v+1,k-1} = L^{(4)}(k,v)D_k, \quad L^{(4)}(k,v) = -(2k+4v-2)[(k-v)(k-v-1)]^{1/2}, \tag{2.46}$$
$$\delta^k_{v+1,k-2} = L^{(5)}(k,v)E_k, \quad L^{(5)}(k,v) = -4[(k-v)(k-v-1)(k-v-2)(k+v)]^{1/2} \tag{2.47}$$
$$\eta^k_{v,k+2} = M^{(1)}(k,v)A_{k+2}, \quad M^{(1)}(k,v) = 12[(k+v+1)(k+v+2)(k-v+2)(k-v+2)]^{1/2} \tag{2.48}$$
$$\eta^k_{v,k+1} = M^{(2)}(k,v)B_{k+1}, \quad M^{(2)}(k,v) = -12v[(k+v+1)(k-v+1)]^{1/2}, \tag{2.49}$$
$$\eta^k_{v,k} = M^{(3)}(k,v)C_k, \quad M^{(3)}(k,v) = 12v^2 - 4k^2 - 4k, \tag{2.50}$$
$$\eta^k_{v,k-1} = M^{(4)}(k,v)D_k, \quad M^{(4)}(k,v) = 12v[(k+v)(k-v)]^{1/2}, \tag{2.51}$$
$$\eta^k_{v,k-2} = M^{(5)}(k,v)E_k, \quad M^{(5)}(k,v) = 12[(k-v)(k-v-1)(k+v)(k+v-1)]^{1/2}, \tag{2.52}$$
$$\xi^k_{v-1,k+2} = N^{(1)}(k,v)A_{k+2}, \quad N^{(1)}(k,v) = 24[(k+v+1)(k-v+2)(k-v+1)(k-v+3)]^{1/2}, \tag{2.53}$$
$$\xi^k_{v-1,k+1} = N^{(2)}(k,v)B_{k+1}, \quad N^{(2)}(k,v) = -12(k+2v)[(k-v+1)(k-v+2)]^{1/2}, \tag{2.54}$$
$$\xi^k_{v-1,k} = N^{(3)}(k,v)C_k, \quad N^{(3)}(k,v) = (24v-12)[(k+v)(k-v+1)]^{1/2}, \tag{2.55}$$
$$\xi^k_{v-1,k+1} = N^{(4)}(k,v)D_k, \quad N^{(4)}(k,v) = 12(k-2v+1)[(k+v)(k+v-1)]^{1/2}, \tag{2.56}$$
$$\xi^k_{v-1,k-2} = N^{(5)}(k,v)E_k, \quad N^{(5)}(k,v) = -24[(k-v)(k+v-1)(k+v)(k+v-2)]^{1/2}, \tag{2.57}$$
$$\gamma^k_{v-2,k+2} = P^{(1)}(k,v)A_{k+2}, \quad P^{(1)}(k,v) = 24[(k-v+2)(k-v+1)(k-v+3)(k-v+4)]^{1/2}, \tag{2.58}$$
$$\gamma^k_{v-2,k+1} = P^{(2)}(k,v)B_{k+1}, \quad P^{(2)}(k,v) = -24[(k-v+1)(k-v+2)(k+v)(k-v+3)]^{1/2}, \tag{2.59}$$
$$\gamma^k_{v-2,k} = P^{(3)}(k,v)C_k, \quad P^{(3)}(k,v) = 24[(k+v)(k-v+1)(k+v-1)(k-v+2)]^{1/2}, \tag{2.60}$$
$$\gamma^k_{v-2,k-1} = P^{(4)}(k,v)D_k, \quad P^{(4)}(k,v) = -24[(k+v-1)(k+v-2)(k+v)(k-v+1)]^{1/2}, \tag{2.61}$$
$$\gamma^k_{v-2,k-2} = P^{(5)}(k,v)E_k, \quad P^{(5)}(k,v) = 24[(k+v)(k+v-1)(k+v-2)(k+v-3)]^{1/2}, \tag{2.62}$$

where

$$A_{k+2} = \frac{\beta^k_{k+2,k+2}}{[6.7.8.9(2k+1)(2k+2)(2k+3)(2k+4)]^{1/2}}, \tag{2.63}$$
$$B_{k+1} = \frac{\beta^k_{k+1,k+1}}{[84k(2k+1)(2k+2)]^{1/2}}, \quad C_k = \frac{\beta^k_{k,k}}{[4k(2k-1)]^{1/2}}, \tag{2.64}$$
$$D_k = \frac{\beta^k_{k-1,k-1}}{[\frac{3}{10}(2k-2)]^{1/2}}, \quad E_k = \sqrt{5}\beta^k_{k-2,k-2}. \tag{2.65}$$

Hence, with the aid of some of the commutation relations we determine all nonzero coefficients in terms of five arbitrary complex numbers, $A_{k+2}, B_{k+1}, C_k, D_k$, and E_k. Equations (2.29)–(2.33) in terms of these complex numbers are determined as

$$T_{-2}f^k_v = P^{(5)}(k,v)E_k f^{k-2}_{v-2} + P^{(4)}(k,v)D_k f^{k-1}_{v-2} + P^{(3)}(k,v)C_k f^k_{v-2} + P^{(2)}(k,v)B_{k+1} f^{k+1}_{v-2} + P^{(1)}(k,v)A_{k+2} f^{k+2}_{v-2}, \tag{2.66}$$
$$T_{-1}f^k_v = N^{(5)}(k,v)E_k f^{k-2}_{v-1} + N^{(4)}(k,v)D_k f^{k-1}_{v-1} + N^{(3)}(k,v)C_k f^k_{v-1} + N^{(2)}(k,v)B_{k+1} f^{k+1}_{v-1} + N^{(1)}(k,v)A_{k+2} f^{k+2}_{v-1}, \tag{2.67}$$
$$T_0 f^k_v = M^{(5)}(k,v)E_k f^{k-2}_v + M^{(4)}(k,v)D_k f^{k-1}_v + M^{(3)}(k,v)C_k f^k_v + M^{(2)}(k,v)B_{k+1} f^{k+1}_v + M^{(1)}(k,v)A_{k+2} f^{k+2}_v, \tag{2.68}$$
$$T_{+1}f^k_v = L^{(5)}(k,v)E_k f^{k-2}_{v+1} + L^{(4)}(k,v)D_k f^{k-1}_{v+1} + L^{(3)}(k,v)C_k f^k_{v+1} + L^{(2)}(k,v)B_{k+1} f^{k+1}_{v+1} + L^{(1)}(k,v)A_{k+2} f^{k+2}_{v+2}, \tag{2.69}$$
$$T_{+2}f^k_v = K^{(5)}(k,v)E_k f^{k-2}_{v+2} + K^{(4)}(k,v)D_k f^{k-1}_{v+2} + K^{(3)}(k,v)C_k f^k_{v+2} + K^{(2)}(k,v)B_{k+1} f^{k+1}_{v+1} + K^{(1)}(k,v)A_{k+2} f^{k+2}_{v+2}. \tag{2.70}$$

III. THE DETERMINATION OF A_k, B_k, C_k, D_k, AND E_k

To determine the representation space R, five coefficients A_k, B_k, C_k, D_k, and E_k should be calculated as functions of k only using the rest of the commutation relations. These relations will give equations containing the arbitrary coefficients. Simultaneous solution of the equations will determine A_k, B_k, C_k, D_k, and E_k in terms of k only. For the sake of completeness we will demonstrate the method for only one commutation relation. All equations resulting from other commutation relations will be listed.

Using Eq. (2.29) and Eq. (2.33), the commutation relation

$$[T_{-2},T_{+2}]f^k_v = 4H_3 f^k_v \tag{3.1}$$

gives

$$T_{-2}\left[\sum_{j=k-2}^{k+2}\beta^{k}_{v+2,j}f^{j}_{v+2}\right]-T_{+2}\left[\sum_{j=k-2}^{k+2}\gamma^{k}_{v-2,j}f^{j}_{v-2}\right]=4vf^{k}_{v}, \tag{3.2}$$

$$\sum_{j=k-2}^{k+2}\beta^{k}_{v+2,j}\left[\sum_{m=j-2}^{j+2}\gamma^{j}_{v,m}f^{m}_{v}\right]-\sum_{m=j-2}^{k+2}\left[\sum_{m=j-2}^{j+2}\beta^{j}_{v,m}f^{m}_{v}\right]=4vf^{k}_{v}. \tag{3.3}$$

The linear independence of f^k_v imply the following equations:

$$P^{(5)}(k,v)K^{(5)}(k-2,v-2)E_kE_{k-2}=K^{(5)}(k,v)P^{(5)}(k-2,v+2)E_kE_{k-2}, \tag{3.4}$$

$$\begin{aligned}&P^{(5)}(k,v)K^{(4)}(k-2,v-2)E_kD_{k-2}=P^{(4)}(k,v)K^{(5)}(k-1,v-2)E_{k-1}D_k\\&\quad=K^{(5)}(k,v)P^{(4)}(k-2,v+2)E_kD_{k-2}=K^{(4)}(k,v)P^{(5)}(k-1,v+2)E_{k-1}D_k,\end{aligned} \tag{3.5}$$

$$\begin{aligned}&P^{(5)}(k,v)K^{(3)}(k-2,v-2)E_kC_{k-2}+P^{(4)}(k,v)K^{(4)}(k-1,v-2)D_kD_{k-1}\\&\quad+P^{(3)}(k,v)K^{(5)}(k,v-2)C_kE_k=K^{(5)}(k,v)P^{(3)}(k-2,v+2)E_kC_{k-2}\\&\quad+K^{(4)}(k,v)P^{(4)}(k-1,v+2)D_kD_{k-1}+K^{(3)}(k,v)P^{(5)}(k,v+2)C_kE_k,\end{aligned} \tag{3.6}$$

$$\begin{aligned}&P^{(5)}(k,v)K^{(2)}(k-2,v+2)E_kB_{k-1}+P^{(4)}(k,v)K^{(3)}(k-1,v-2)D_kC_{k-1}\\&\quad+P^{(3)}(k,v)K^{(4)}(k,v-2)C_kD_k+P^{(2)}(k,v)K^{(5)}(k+1,v-2)B_{k+1}E_{k+1}\\&\quad=K^{(5)}(k,v)P^{(2)}(k-1,v+2)E_kB_{k-1}K^{(4)}(k,v)P^{(3)}(k-1,v+2)C_{k-1}D_k\\&\quad+K^{(3)}(k,v)P^{(4)}(k,v+2)C_kD_k+K^{(2)}(k,v)P^{(5)}(k+1,v+2)B_{k+1}E_{k+1},\end{aligned} \tag{3.7}$$

$$\begin{aligned}&P^{(5)}(k,v)K^{(1)}(k-2,v-2)E_kA_k+P^{(4)}(k,v)K^{(2)}(k-1,v-2)B_kD_k+P^{(3)}(k,v)K^{(3)}(k,v-2)C^2_k\\&\quad+P^{(2)}(k,v)K^{(4)}(k+1,v-2)B_{k+1}D_{k+1}+P^{(1)}(k,v)K^{(5)}(k+2,v-2)A_{k+2}E_{k+2}\\&\quad=K^{(5)}(k,v)P^{(1)}(k-2,v+2)E_kA_k+K^{(4)}(k,v)P^{(2)}(k-1,v+2)B_kD_k\\&\quad+K^{(3)}(k,v)P^{(3)}(k,v-2)C^2_k+K^{(2)}(k,v)P^{(4)}(k+1,v+2)B_{k+1}D_{k+1}\\&\quad+K^{(1)}(k,v)P^{(5)}(k+2,v+2)A_{k+2}E_{k+2}+4v,\end{aligned} \tag{3.8}$$

$$\begin{aligned}&P^{(4)}(k,v)K^{(1)}(k-1,v-2)A_{k+1}D_k+P^{(4)}(k,v)K^{(2)}(k,v-2)C_kB_{k+1}+P^{(2)}(k,v)\\&\quad\times K^{(3)}(k+1,v-2)B_{k+1}C_{k+1}+P^{(1)}(k,v)K^{(4)}(k\,12,v-2)A_{k+2}D_{k+2}\\&\quad=K^{(4)}(k,v)P^{(1)}(k-1,v+2)A_{k+1}D_k+K^{(3)}(k,v)P^{(2)}(k,v+2)C_kB_{k+1}\\&\quad+K^{(2)}(k,v)P^{(3)}(k+1,v+2)B_{k+1}C_{k+1}+K^{(1)}(k,v)P^{(4)}(k+2,v+2)A_{k+2}D_{k+2},\end{aligned} \tag{3.9}$$

$$\begin{aligned}&P^{(3)}(k,v)K^{(1)}(k,v-2)A_{k+2}C_k+P^{(2)}(k,v)K^{(2)}(k+1,v-2)B_{k+1}B_{k+2}\\&\quad+P^{(1)}(k,v)K^{(3)}(k+2,v-2)A_{k+2}C_{k+2}=K^{(3)}(k,v)P^{(1)}(k,v+2)A_{k+2}C_k\\&\quad+K^{(2)}(k,v)P^{(2)}(k+1,v-2)B_{k+1}B_{k+2}+K^{(1)}(k,v)P^{(3)}(k+2,v+2)A_{k+2}C_{k+2},\end{aligned} \tag{3.10}$$

$$\begin{aligned}&P^{(2)}(k,v)K^{(1)}(k+1,v-2)B_{k+1}a_{k+3}+P^{(1)}(k,v)K^{(2)}(k+2,v-2)A_{k+2}B_{k+3}\\&\quad=K^{(2)}(k,v)P^{(1)}(k+1,v+2)B_{k+1}A_{k+3}+K^{(1)}(k,v)P^{(2)}(k+2,v+2)A_{k+2}B_{k+3},\end{aligned} \tag{3.11}$$

$$P^{(1)}(k,v)K^{(1)}(k+2,v-2)A_{k+2}A_{k+4}=K^{(1)}(k,v)P^{(1)}(k+2,v+2)A_{k+2}A_{k+4}. \tag{3.12}$$

Repeating the same procedure for the commutation relations $[T_{-1},T_{-2}]=0$, $[T_{-2},T_{+1}]=2H_-$, $[T_0,T_{-1}]=\sqrt{6}H_-$, $[T_{+1},T_{-1}]=2H_3$, $[T_{+2},T_{-1}]=2H_+$, $[T_0,T_{+1}]=\sqrt{6}H_+$, $[T_0,T_{+2}]=0$, and $[T_{+1},T_{+2}]=0$, one obtains equations containing k,v and five arbitrary functions A_k, B_k, C_k, D_k, and E_k. All equations which are functions of k only are listed below:

$$kA_{k+3}B_{k+1}-(k+4)A_{k+2}B_{k+3}=0, \tag{3.13}$$

$$(2k-1)A_{k+2}C_k-3B_{k+1}B_{k+2}-(2k+7)A_{k+2}C_{k+2}=0, \tag{3.14}$$

$$3(k-1)A_{k+1}D_k+(k-3)B_{k+1}C_k-(k+5)B_{k+1}C_{k+1}-3(k+3)A_{k+2}D_{k+2}=0, \tag{3.15}$$

$$(-2k+3)A_kE_k+(-k+2)B_kD_k+2C^2_k+(k+3)B_{k+1}D_{k+1}+(2k+5)A_{k+2}E_{k+2}=0, \tag{3.16}$$

$$3(k-2)B_{k-1}E_k+(k-4)C_{k-1}D_k-(k+4)C_kD_k-3(k+2)B_{k+1}E_{k+1}=0, \tag{3.17}$$

$$(-2k+5)C_{k-2}E_k+3D_kD_{k-1}+(2k+3)C_kE_k=0, \tag{3.18}$$

$$(k-3)E_kD_{k-2}-(k+1)E_{k-1}D_k=0. \tag{3.19}$$

Since we are free to define a new set of basis vectors $f'^k_v=\omega(k)f^k_v$, the number of independent coefficients may be decreased by choosing $\omega(k)$ in such a way that $A'_k=E'_k$, $B'_k=D'_k$. To be able to write Eqs. (2.66)–(2.70) in terms of new basis vectors f'^k_v and new coefficients A'_k, B'_k, C'_k one has the restriction

$$E'_k=\frac{\omega(k)}{\omega(k-2)}E_k,\quad D'_k=\frac{\omega(k)}{\omega(k-1)}D_k,\quad C'_k=C_k,$$

$$B'_{k+1}=\frac{\omega(k)}{\omega(k+1)}B_{k+1},\quad A'_{k+2}=\frac{\omega(k)}{\omega(k+2)}A_{k+2}. \tag{3.20}$$

The function $\omega(k)$ which satisfies $A'_k=E'_k$ and $B'_k=D'_k$ is

$$\omega(k)=\left[\prod_{j=k_0+1}^{k+1}\frac{A_j/E_j}{B_j/D_j}\right]^{1/2}. \tag{3.21}$$

Here k_0 is the minimum of k values included in R. Under these conditions Eqs. (3.13)–(3.19) in terms of new coefficients A'_k, B'_k, C'_k read as

$$kA'_{k+3}B'_{k+1}-(k+4)A'_{k+2}B'_{k+3}=0, \tag{3.22}$$

$$(2k-1)A'_{k+2}C'_k-3B'_{k+1}B'_{k+2}-(2k+7)A'_{k+2}C'_{k+2}=0, \tag{3.23}$$

$$3(k-1)A'_{k+1}B'_k+(k-3)B'_{k+1}C'_k-(k+5)B'_{k+1}C'_{k+1}-3(k+3)A'_{k+2}B'_{k+2}=0, \tag{3.24}$$

$$(-2k+3)A'^2_k+(-k+2)B'^2_k+2C^2_k+(k-3)B'^2_{k+1}+(2k+5)A'^2_{k+2}=0. \tag{3.25}$$

The last step is to solve Eqs. (3.22)–(3.25) simultaneously for the unknowns A'_k, B'_k, C'_k. For this purpose let us write Eq. (3.22) for $k=k-2$

$$(k-2)A'_{k+1}B'_{k-1}=(k+2)A'_kB'_{k+1}. \tag{3.26}$$

Multiplying both sides by k and letting

$$\sigma_k=k(k-2)B'_{k-1}, \tag{3.27}$$

Eq. (3.22) takes the form

$$A'_{k+1}=(\sigma_{k+2}/\sigma_1)A'_k. \tag{3.28}$$

Since the minimum value of k is k_0, Eqs. (2.66)–(2.70) impose the restrictions $A'_{k_0}=A'_{k_0+1}=B'_{k_0}=0$. Hence,

$$A'_k=Zk(k-2)(k-1)(k+1)B'_kB'_{k-1},\quad k=k_0+2,k_0+3,\ldots,k_0\neq 0, \tag{3.29}$$

where

$$Z=\frac{A'_{k_0+2}}{k_0(k_0+1)(k_0+2)(k_0+3)B'_{k_0+1}B'_{k_0+2}}. \tag{3.30}$$

Multiplying Eq. (3.23) by $(2k+3)$ and inserting the expression for A'_{k+2} one obtains

$$(2k+3)(2k-1)C'_k-(2k+3)(2k+7)C'_{k+2}=3(2k+3)/Zk(k+1)(k+2)(k+3). \tag{3.31}$$

Letting

$$\sigma'_k=(2k+3)(2k-1)C'_k \tag{3.32}$$

C'_k's are obtained as

$$C'^+_k=\frac{(2k_0+3)(2k_0-1)C'_{k_0}-F(k_0,n)}{(2k-1)(2k+3)},\quad k=k_0+2n,\ n=1,2,\ldots,k_0\neq 0, \tag{3.33}$$

$$C'^-_k=\frac{(2k_0+5)(2k_0+1)C'_{k_0+1}-F(k_0+1,n)}{(2k-1)(2k+3)},\quad k=k_0+2n+1,\ n=1,2,\ldots,k_0\neq 0, \tag{3.34}$$

where

$$F(k_0,n)=\sum_{j=1}^{n}\frac{3[2(k_0+2j-2)+3]}{Z(k_0+2j-2)(k_0+2j-1)(k_0+2j)(k_0+2j+1)}. \tag{3.35}$$

Multiplying Eq. (3.24) by $(3k+3)$, inserting expressions for A_{k+1}, A_{k+2} and letting

$$\sigma''_k=9(k-1)^2(k+1)^2B'^2_k, \tag{3.36}$$

one obtains

$$\sigma''_k-\sigma''_{k+2}=(3(k+1)/Zk(k+2))[(k+5)C'_{k+1}-(k-3)C'_k],\quad k\neq 0. \tag{3.37}$$

Eq. (3.37) gives nonzero B'_k's as

$$B'_{k_0+2}=\frac{[(k_0-3)C'_{k_0}-(k_0+5)C'_{k_0+1}]B'_{k_0+1}}{3(k_0+3)A'_{k_0+2}}, \tag{3.38}$$

$$B'_{k_0+2n}=-\frac{[F'(k_0,n)]^{1/2}}{3(k_0+2n-1)(k_0+2n+1)},\quad n=0,1,2,\ldots,k_0\neq 0, \tag{3.39}$$

$$B'_{k_0+2n+1}=\frac{[9k_0^2(k_0+2)^2B'^2_{k_0+1}-F'(k_0+1,n)]^{1/2}}{3(k_0+2n)(k_0+2n+2)},\quad n=1,2,\ldots,k_0\neq 0, \tag{3.40}$$

where

$$F'(k_0,n) = \sum_{j=1}^{n} \frac{3(k_0+2j-1)(k_0+2j+3)}{Z(k_0+2j-2)(k_0+2j)} C_{k_0+2j-1} - \frac{3(k_0+2j-1)(k_0+2j-5)}{Z(k_0+2j-2)(k_0+2j)} C_{k_0+2j-2}, \tag{3.41}$$

$$F'(k_0,0) = 0.$$

Equation (3.25) contributes only for $k = k_0$ which reads as

$$2C'^2_{k_0} + (k_0-3)B'^2_{k_0+1} + (2k_0+5)A'^2_{k_0+2} = 0. \tag{3.42}$$

Therefore, all nonzero coefficients A'_k, B'_k, and C'_k are determined by nonzero integer or half-integer k_0 and three independent complex numbers B'_{k_0+1}, C'_{k_0}, and C'_{k_0+1}.

The case $k_0 = 0$ should be treated in the same manner. Equation (3.22) gives

$$A'_0 = A'_1 = A'_2 = 0, \tag{3.43}$$

$$A'_k = Z_0 k(k-2)(k-1)(k+1)B'_k B'_{k-1}, \quad k = 3,4,..., k_0 = 0, \tag{3.44}$$

where

$$Z_0 = A'_3/24B'_2B'_3. \tag{3.45}$$

Inserting Eq. (3.44) for $k = k+2$ in Eq. (3.23) one gets

$$B'_1 = 0, \quad C'_0 = \text{any complex number}, \tag{3.46}$$

$$C'^-_k = \frac{5C'_1 - F_0(n)}{(2k-1)(2k+3)}, \quad k = 2n+1, n = 1,2,..., k_0 = 0, \tag{3.47}$$

$$C'^+_k = \frac{21C'_2 - G_0(n)}{(2k-1)(2k+3)}, \quad k = 2n+2, \ n = 1,2,..., k_0 = 0, \tag{3.48}$$

where

$$F_0(n) = \sum_{j=1}^{n} \frac{3(4j+1)}{Z_0(2j-1)(2j)(2j+1)(2j+2)}, \tag{3.49}$$

$$G_0(n) = \sum_{j=1}^{n} \frac{3(4j+3)}{Z_0 2j(2j+1)(2j+2)(2j+3)}, \tag{3.50}$$

Eq. (3.37) takes the form

$$\sigma''_k - \sigma''_{k+2} = \frac{3(k+1)}{Z_0 k(k+2)} [(k+5)C'_{k+1} - (k-3)C'_k], \quad k = 2,3,4,..., k_0 = 0. \tag{3.51}$$

So

$$B'_3 = \frac{(-2C'_1 - 6C'_2)B'_2}{12A'_3}, \tag{3.52}$$

$$B'^+_k = \frac{[81B'^2_2 - F'_0(n)]^{1/2}}{3(k-1)(k+1)}, \quad k = 2n+2, n = 1,2,..., k_0 = 0, \tag{3.53}$$

$$B'^-_k = \frac{[576B'^2_3 - G'_0(n)]^{1/2}}{3(k-1)(k+1)}, \quad k = 2n+3, n = 1,2,..., k_0 = 0, \tag{3.54}$$

where

$$F'_0(n) = \sum_{j=1}^{n} \frac{3(2_j+1)}{Z_0 2_j(2_j+2)} \times [(2_j+5)C'_{2_j+1} - (2j-3)C'_{2_j}], \tag{3.55}$$

$$G'_0(n) = \sum_{j=1}^{n} \frac{3(2_j+2)}{Z_0(2_j+1)(2_j+2)} \times [(2_j+6)C'_{2_j+2} - (2_j-2)C'_{2_j+1}]. \tag{3.56}$$

For $k = k_0$ and $k = k_0+1$ Eq. (3.25) implies that

$$C'_0 = 0, \tag{3.57}$$

$$2C'^2_1 - 2B'^2_2 + 7A'^2_3 = 0. \tag{3.58}$$

Hence, for $k_0 = 0$ case all nonzero coefficients A'_k, B'_k, and C'_k are determined by three independent complex numbers B'_2, C'_1, and C'_2.

Letting $B'_{k_0+1} = a+ib$, $C'_{k_0} = a_0 + i\beta_0$, $C'_{k_0+1} = \alpha_1 + i\beta_1$ and using Eqs. (3.30), (3.33), (3.34), and (3.38), explicit expressions for Z and C'^+_k, C'^-_k are obtained as

$$Z = \psi(k_0)\left[\frac{QU-VY}{U^2+Y^2} - i\frac{UV+QY}{U^2Y^2}\right], \tag{3.59}$$

$$C'_{k_0+2j} = \left[\eta(k_0,j)\alpha_0 - \frac{\xi(k_0,j)[QU-VY]}{\psi(k_0)[Q^2+V^2]}\right] + i\left[\eta(k_0,j)\beta_0 - \frac{\xi(k_0,j)[UV+QY]}{[Q^2+V^2]\Psi(k_0)}\right] \tag{3.60}$$

$$C'_{k_0+2j+1} = \left[\eta(k_0+1,j)\alpha_1 - \frac{\xi(k_0+1,j)[QU-VY]}{\psi(k_0)[Q^2+V^2]}\right] + i\left[\eta(k_0+1,j)\beta_1 - \frac{\xi(k_0+1,j)[UV+QY]}{\Psi(k_0)[Q^2+V^2]}\right], \tag{3.61}$$

where

$$Q = 2(\beta_0^2 - \alpha_0^2) + (k_0-3)(b^2-a^2), \tag{3.62}$$

$$V = 4\alpha_0\beta_0 + 2ab(k_0-3), \tag{3.63}$$

$$U = (a^2-b^2)[(k_0-3)\alpha_0 - (k_0+5)\alpha_1] - 2ab[(k_0-3)\beta_0 - (k_0+5)\beta_1], \tag{3.64}$$

$$Y = (a^2-b^2)[(k_0-3)\beta_0 - (k_0+5)\beta_1] + 2ab[(k_0-3)\alpha_0 - (k_0+5)\alpha_1], \tag{3.65}$$

$$\psi(k_0) = \frac{3(k_0+3)}{k_0(2k_0+5)(k_0+1)(k_0+2)(k_0+3)}, \tag{3.66}$$

$$\eta(k_0,j) = \frac{(2k_0+3)(2k_0-1)}{[2(k_0+2j)+3][2(k_0+2j)-1]}, \tag{3.67}$$

$$\xi(k_0,j) = \frac{F(k_0,j)}{[2(k_0+2j)+3][2(k_0+2j)-1]}.$$

Expressions of A'_k's and B'_k's are easily determined in terms of real numbers $\alpha_0, \beta_0, \alpha_1, \beta_1, a$ and b, by using Eqs. (3.37) and (3.29).

IV. THE MULTIPLICITY-FREE, IRREDUCIBLE, UNITARY REPRESENTATIONS

The unitary condition requires

$$(T_\mu f_v^k, f_{v'}^{k'}) = -(f_v^k, T_\mu f_{v'}^{k'}), \tag{4.1}$$

where

$$\mu = 0, \pm 1, \pm 2, \quad k = k_0 + n, \quad n = 0,1,2,3,..., \quad k_0 = 0, \tfrac{1}{2}, 1, ..., \quad v = -k, -k+1, ..., k. \tag{4.2}$$

Here the scalar product is well defined in the space R and the canonical basis vectors f_v^k are orthonormal with respect to the scalar product, that is

$$(f_v^k, f_{v'}^{k'}) = \delta_{kk'}\delta_{vv'}. \tag{4.3}$$

Having determined $T_\mu f_v^k$ in terms of A'_k, B'_k, and C'_k [Eqs. (2.66)–(2.70)] the unitary condition imposes restrictions on the coefficients A'_k, B'_k, C'_k; especially

$$\begin{aligned}(T_0 f_v^k, f_v^k) &= -(f_v^k, T_0 f_v^k),\\ (T_0 f_v^k, f_v^{k-2}) &= -(f_v^k, T_0 f_v^{k-2}),\\ (T_0 f_v^k, f_v^{k-1}) &= -(f_v^k, T_0 f_v^{k-1}),\end{aligned} \tag{4.4}$$

imply

$$C'_k = -C'^*_k, \quad A'_k = -A'^*_k, \quad B'_k = B'^*_k, \tag{4.5}$$

where * shows complex conjugation. All other conditions are satisfied identically. These restrictions lead to

$$B'_{k_0+1} = a, \quad C'_{k_0} = i\beta_0, \quad C'_{k_0+1} = i\beta_1, \tag{4.6}$$

$$Q = 2\beta_0^2 - a^2(k_0-3), \quad V = U = 0,$$

$$Y = a^2[(k_0-3)\beta_0 - (k_0+5)\beta_1], \tag{4.7}$$

$$B'^2_{k_0+2} = -\frac{(2k_0+5)Y^2}{9(k_0+3)^2 Q_{a^2}}, \quad A^2_{k_0+2} = -\frac{Q}{2_{k_0+5}}, \tag{4.8}$$

$$B'^2_{k_0+2n} = \frac{(Y/Q\psi(k_0))\Sigma_{j=1}^n \beta_1 A(k_0,j)\eta(k_0+1,j-1) - \beta_0 B(k_0,j)\eta(k_0,j-1)}{9(k_0+2n-1)^2(k_0+2n+1)^2} + \frac{(Y^2/Q^2\psi^2(k_0))\Sigma_{j=1}^n B(k_0,j)\xi(k_0,j-1) - A(k_0,j)\xi(k_0+1,j-1)}{9(k_0+2n-1)^2(k_0+2n+1)^2}, \tag{4.9}$$

$$B'^2_{k_0+2n+1} = \frac{9k_0^2(k_0+2)^2a^2 + (Y/Q\psi(k_0))\Sigma_{j=1}^n \beta_0 A(k_0+1,j)\eta(k_0,j) - \beta_1 B(k_0+1,j)\eta(k_0+1,j-1)}{9(k_0+2n)^2(k_0+2n+2)^2} + \frac{(Y^2/Q^2\psi^2(k_0))\Sigma_{j=1}^n B(k_0+1,j)\xi(k_0+1,j-1) - A(k_0+1,j)\xi(k_0,j)}{9(k_0+2n)^2(k_0+2n+2)^2}, \tag{4.10}$$

where

$$A(k_0,j) = \frac{(k_0+2j-1)(k_0+2j+3)}{(k_0+2j-2)(k_0+2j)},$$

$$B(k_0,j) = \frac{(k_0+2j-1)(k_0+2j-5)}{(k_0+2j-2)(k_0+2j)}. \tag{4.11}$$

Equations (4.9) and (4.10) imply that the unitarity condition, B'_k are real numbers for every k, is satisfied in $Y=0$. This constraint is satisfied for the following two cases:

(1) $(k_0-3)\beta_0 - (k_0+5)\beta_1 = 0, \quad a \neq 0, k_0 \neq 0;$ (4.12)

(2) $a = 0, k_0 \neq 0.$ (4.13)

Since the parameters β_0 and β_1 are independent, constraint (4.12) is valid if

(a) $\beta_0 = \beta_1 = 0, \quad a \neq 0, k_0 \neq 0, k_0 \neq 3$; or (4.14)

(b) $\beta_1 - 0, \quad a \neq 0, k_0 = 3.$ (4.15)

In the general approach the aim was to obtain A'_k, B'_k, and C'_k by solving fundamental equations (3.22)–(3.25). But now the problem is to determine those A'_k, B'_k, and C'_k which satisfy the fundamental equations and are consistent with the constraints imposed by the unitarity. All possible cases are listed below.

(1a) For $\beta_0 = \beta_1 = 0, \quad a \neq 0, k_0 \neq 0$

$$C'_{k_0+2n} = C'_{k_0+2n+1} = 0, \quad n = 0,1,2,..., \tag{4.16}$$

$$B'_{k_0+2n} = 0, \quad n = 1,2,3,..., \tag{4.17}$$

$$B'_{k_0+2n+1} = \frac{k_0(k_0+2)a}{(k_0+2n)(k_0+2n+2)}, \quad n = 0,1,2,..., \tag{4.18}$$

$$A'^{2}_{k_02} = -(k_0-3)a^2/(2k_0+5), \quad k_0>3, \tag{4.19}$$

$$A'_{k_0+2n+1} = A'_{k_0+2n} = A'_{k_0+2}, \quad n=1,2,3,\ldots . \tag{4.20}$$

(1b) For $\beta_1 = 0$, $a \neq 0$, $k_0 = 3$

$$B'_{k_0+2n} = 0, \quad n=1,2,3,\ldots, \tag{4.21}$$

$$C'_{k_0+2n+1} = 0, \quad n=0,1,2,\ldots, \tag{4.22}$$

$$B'_{k_0+2n+1} = \frac{15a}{(3+2n)(5+2n)}, \quad n=0,1,2,\ldots, \tag{4.23}$$

$$A'_5 = A'_6 = 0,$$

$$A'_{3+2n+1} = A'_{3+2n} = A'_7, \quad n=2,3,4,\ldots . \tag{4.24}$$

(2) For $a=0$, $k_0 \neq 0$

$$B'_{k_0+2n+1} = B'_{k_0+2n} = 0, \quad n=0,1,2,\ldots, \tag{4.25}$$

$$C'_{k_0+2n} = 0, \quad n=0,1,2,\ldots, \tag{4.26}$$

$$A'_{k_0+2} = A'_{k_0+2n+1} = 0, \quad n=1,2,3,\ldots . \tag{4.27}$$

(3) For $k_0 = 0$

$$B'_{2n+1} = B'_{2n+2} = 0, \tag{4.28}$$

$$C'_{2n+1} = C'_{2n+2} = 0. \tag{4.29}$$

Finite-dimensional representations are characterized by conditions $B'_{k+1} = A'_{k+2} = 0$ for every k. Thus, constraints with the fundamental equations are

$$A'_{k_0+n} = 0, \quad n=2,3,4,\ldots, k_0 \neq 0, \tag{4.30}$$

$$B'_{k_0+2n} = B'_{k_0+2n+1} = 0, \quad n=0,1,2,\ldots, k_0 \neq 0, \tag{4.31}$$

$$C'_{k_0+2n} = 0, \quad n=0,1,2,\ldots, k_0 \neq 0, \tag{4.32}$$

$$C'_{2n+1} = 0, \quad n=0,1,2, k_0 = 0, \tag{4.33}$$

$$B'_{2n+1} = B'_{2n+2} = 0, \quad n=1,2,3,\ldots, k_0 = 0. \tag{4.34}$$

V. CONCLUSION

Irreducible, multiplicity-free representations of the noncompact group $\overline{SL(3,R)}$ are classified by a constructive method. The results are listed below.

A. Nonunitary representations

(a) Infinite-dimensional representations are labeled by k_0 and three complex numbers. The k contents are

$$\{0,1,2,3,\ldots\}, \quad k_0 = 0, \tag{5.1}$$

$$\{k_0, k_0+1, k_0+2, \ldots\}, \quad k_0 = \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots . \tag{5.2}$$

(b) Finite-dimensional representations are labeled by k_0 and a complex number. The k contents are

$$\{2,4,6,\ldots,2n\}, \quad k_0 = 0, n = 1,2,\ldots, \tag{5.3}$$

$$\{k_0+1, k_0+3, \ldots, k_0+2n+1\}, \quad k_0 = \tfrac{1}{2}, 1, n = 0,1,2,\ldots . \tag{5.4}$$

B. Unitary representations

There exist two series of unitary representations.

(a) The k content is

$$\{k_0, k_0+1, k_0+2, \ldots\}, \quad k_0 \geqslant 3. \tag{5.5}$$

(b) The k contents are

$$\{0,2,4,\ldots\}, \quad k_0 = 0, \tag{5.6}$$

$$\{\tfrac{1}{2}, \tfrac{5}{2}, \tfrac{9}{2}, \ldots\}, \quad k_0 = \tfrac{1}{2}, \tag{5.7}$$

$$\{1,3,5,7,\ldots\}, \quad k_0 = 1. \tag{5.8}$$

ACKNOWLEDGMENT

This research was partially supported by the Scientific and Technical Research Council of Turkey.

APPENDIX: $\overline{SL(3,R)}$ COMMUTATION RELATIONS

All commutation relations of $\overline{SL(3,R)}$ algebra are

$$[H_3, H_\pm] = \pm H_\pm, \quad [H_+, H_-] = 2H_3,$$
$$[T_{+2}, T_{-2}] = 4H_3,$$
$$[H_3, T_\mu] = \mu T_\mu, \quad \mu = 0, \pm 1, \pm 2,$$
$$[H_\pm, T_\mu] = [6 - \mu(\mu \pm 1)]^{1/2} T_{\mu \pm 1},$$
$$[T_0, T_{+2}] = [T_{+1}, T_{+2}] = [T_{-1}, T_{-2}] = 0,$$
$$[T_0, T_{+1}] = \sqrt{6} H_+, \quad [T_0, T_{-1}] = \sqrt{6} H_-,$$
$$[T_{+1}, T_{-1}] = 2H_3,$$
$$[T_{+1}, T_{-2}] = -2H_-, \quad [T_{-1}, T_{+2}] = -2H_+.$$

[1] I. M. Gel'fand and M. I. Graev, Amer. Math. Soc. Transl. Ser. 2 **2**, 147 (1956).
[2] H. Chandra, Proc. Natl. Acad. Sci. (USA) **37**, 362 (1951).
[3] Y. Dothan, M. Gell-Mann, and Y. Ne'man, Phys. Lett. **17**, 148 (1965).
[4] R. Y. Cusson, Nucl. Phys, A **114**, 289 (1968).
[5] L. Weaver and L. C. Biedenharn, Nucl. Phys. A **185**, 1 (1972).
[6] Dj. Sijacki, J. Math. Phys. **16**, 298 (1975).
[7] Y. Güler, J. Math. Phys. **18**, 413 (1977).
[8] Y. Güler, J. Math. Phys. **19**, 508 (1978).
[9] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

Unitary representations of the (4 + 1) de Sitter group on unitary irreducible representation spaces of the Poincaré group: Equivalence with their realizations as induced representations

P. Moylan
Max Planck Institut für Physik und Astrophysik, Werner Heisenberg Institut für Physik, Munich, Federal Republic of Germany

(Received 13 June 1984; accepted for publication 20 July 1984)

In a previous work we have constructed realizations of the principal continuous series of unitary irreducible representations of the simply connected covering group of the (4 + 1) de Sitter group on unitary irreducible representation spaces of the simply connected covering group of the Poincaré group. In this work we demonstrate the equivalence of the representations constructed in the previous work with their realizations as induced representations.

I. INTRODUCTION

In a previous article,[1] hereafter denoted by I, we have given a construction of the principal continuous series of unitary irreducible representations (UIR's) of the simply connected covering group of the (4 + 1) de Sitter group on UIR spaces of the simply connected covering group of the Poincaré group. The method of construction presented there is a generalization to arbitrary spin of the classical description given by Bargmann[2] for the construction of $SO_0(n,1)$ multiplier representations on S_{n-1}. The usual description of these representations is the description obtained by inducing representations of certain subgroups to the whole of the group.[3] In this paper we prove the equivalence of the realizations of the UIR's of $SO_0(4,1)$ presented in I with their realizations by induced representations. We also consider the standard description by induced representations of the complementary series of UIR's of $SO_0(4,1)$ and show how they can be given a description analogous to the one given for the principal series in I.

In Sec. II we review some general facts concerning the $SO_0(n,1)$ groups and their standard decompositions used in the description of the induced representations.[3,4] We give our definition of a multiplier representation and present a classification of all continuous unitary irreducible representations of $SO_0(4,1)$. In Sec. III we present the inducing construction of the induced unitary irreducible representations of $SO_0(4,1)$ and describe several equivalent realizations. In Sec. IV the realization of principal series UIR's of $SO_0(4,1)$ on UIR spaces of the simply connected covering groups of the Euclidean and Poincaré groups, which was presented in I, is briefly reviewed. In Sec. V the equivalence of the realizations described in Sec. IV with their counterparts, which were described in Sec. III is proven, and the analogous construction of the complementary series UIR's on UIR spaces of the Euclidean and Poincaré group is briefly sketched. We also briefly mention the existence of a similar description of the UIR's of $SO_0(4,1)$ on UIR spaces of the Galilei group. Also we point out the equivalence of the UIR's with irreducible representations of $SO_0(4,1)$ occurring in the left regular representations on certain homogeneous spaces of $SO_0(4,1)$. With regard to notation, we adapt that used in I, with a few minor changes, which will be made clear when they occur. One noteworthy new convention is that a point in Minkowski momentum space $M^{3,1}$ is labeled by the numbers (p_1, p_2, p_3, p_4) instead of (p_0, p_1, p_2, p_3) as in I.

II. PROPERTIES OF THE SO(n,1) GROUPS AND MULTIPLIER REPRESENTATIONS

The group $O(n,1)$ is defined as the set of all linear transformations of $(n+1)$-dimensional Minkowski space $M^{n,1}$ which leaves invariant the quadratic form

$$\Omega(x) = -x_1^2 - x_2^2 - \cdots - x_n^2 + x_{n+1}^2. \tag{2.1}$$

The component connected to the identity of $O(n,1)$ is denoted by $SO_0(n,1)$. The simply connected covering group of $SO_0(n,1)$ is denoted by $\overline{SO_0(n,1)}$. In general we denote the simply connected covering group of an arbitrary group G by $\bar{G}$. The Lie algebra of $\overline{SO_0(n,1)}$ is denoted by so(n,1) and the Hermitian generators A_{ab} of the Lie algebra obey the commutation relations

$$[A_{ab}, A_{cd}] = -i(\eta_{ac} A_{bd} + \eta_{bd} A_{ac} - \eta_{bc} A_{ad} - \eta_{ad} A_{bc}) \tag{2.2}$$

$$[\eta_{ab} = \mathrm{diag}(-1, -1, -1, \ldots, -1, 1)].$$

The generators I_{ab}, which are most frequently used by the mathematicians, are related to the A_{ab}'s by $iI_{ab} = A_{ab}$. The I_{ab} constitute a basis for a representation of the Lie algebra so(n,1).

The following important subgroups play a crucial role in the analysis of the principal series representations of $\overline{SO_0(n,1)}$[3,4]: $K = \overline{SO_0(n)}$, the maximal compact subgroup whose generators are $I_{\mu\nu}$ $(\mu,\nu = 1,\ldots,n)$; $A = \overline{SO_0(1,1)}$, a one-dimensional noncompact subgroup generated by $I_{n,n+1}$ $M = \overline{SO_0(n-1)}$, the centralizer of A in K, whose generators are I_{ij} $(i,j = 1,\ldots,n-1)$; N, a nilpotent, abelian subgroup with generators $I_{i,n+1} - I_{i,n}$; $\tilde{N}$, a nilpotent, abelian subgroup (translation subgroup) with generators $I_{i,n+1} + I_{i,n}$; and $H = \overline{SO_0(n-1,1)}$.

Let M' be the normalizer of A in K. The Weyl group is defined to be the finite group $W = M'/M$. It has the two elements[5]

$$W = \{e,w\} \quad [w = \mathrm{diag}(-1, -1, \ldots, -1, 1) \text{ for } n \text{ even}].$$

We have the following decompositions of $\overline{SO_0(n,1)}$: (1) the Iwasawa decomposition

 0022-2488/85/010029-12$02.50

$$\overline{\mathrm{SO}_0(n,1)} = KAN; \tag{2.3}$$

(2) the Gelfand–Naimark–Bruhat decomposition

$$\overline{\mathrm{SO}_0(n,1)} = \tilde{N}MAN \cup S; \tag{2.4}$$

(3) Hannabuss' decomposition (n even)

$$\overline{\mathrm{SO}_0(n,1)} = (H \cup Hw)AN \cup S' \tag{2.5}$$

[S and S' are excluded subsets of $\overline{\mathrm{SO}_0(n,1)}$ with group invariant measure zero which are described in Ref. 3]. The corresponding coset spaces are[3,4]:

$$(1) \quad \overline{\mathrm{SO}_0(n,1)}/MAN \simeq S_{n-1}, \tag{2.3'}$$

$$(2) \quad (\overline{\mathrm{SO}_0(n,1)} \setminus S)/MAN \simeq \mathbb{R}^{n-1}, \tag{2.4'}$$

$$(3) \quad (\overline{\mathrm{SO}_0(n,1)} \setminus S')/MAN \simeq T_{n-1} \quad (n \text{ even}). \tag{2.5'}$$

Here S_{n-1} and T_{n-1} are the $(n-1)$ sphere and the $(n-1)$-dimensional two-sheeted hyperboloid in $\mathbb{R}^n$, respectively. Note the above results (2.3)–(2.5) for the covering group of $\mathrm{SO}_0(n,1)$ follow from the corresponding decompositions for $\mathrm{SO}_0(n,1)$ (which are given in Ref. 3) and the observation that the fundamental group of a Cartesian product of topological spaces is isomorphic to the direct product of the fundamental group of the individual spaces, together with the fact that the fundamental groups of S_{n-1} $(n>2)$, A, and N are all isomorphic to the trivial group consisting only of the identity element. Thus

$$\overline{\mathrm{SO}_0(n,1)} \simeq S_{n-1}\ \mathrm{Spin}(n-1)\,AN, \tag{2.6}$$

since[6]

$$\mathrm{Spin}(n-1) = \overline{\mathrm{SO}_0(n-1)} = M \quad (n \geqslant 3). \tag{2.7}$$

Now we make the following definition of a multiplier representation of an arbitrary Lie group G. Let

$$\mathscr{H} = \mathscr{H}_0(M) \otimes V \quad [V = \mathbb{C}^{\alpha(s)};\ \alpha(s) \in \mathbb{Z}_+] \tag{2.8}$$

be a Hilbert space of complex-valued measurable functions over a manifold M with inner product

$$(f_1, f_2) = \int_M d\Omega(\xi, \xi')\langle f_1(\xi), f_2(\xi')\rangle_V, \tag{2.9}$$

where

$$\langle f_1(\xi), f_2(\xi')\rangle_V = \sum_{n=1}^{\alpha(s)} \overline{f_1^n(\xi)}\, f_2^n(\xi')$$

is the usual inner product on V. We then have

Definition: A multiplier representation $U^{(s,\nu)}$ of a Lie group G on $\mathscr{H}$ is a bounded, continuous representation $U^{(s,\nu)}(G)$ of G on $\mathscr{H}$ given by

$$[U^{(s,\nu)}(g)f](\xi) = [\mu(g^{-1}, \xi)]^{\nu} D^{\alpha(s)}(g, \xi) f(g \cdot \xi), \quad (f \in \mathscr{H}, g \in G), \tag{2.10}$$

where $g \cdot \xi$ is a global action of G on M and $\mu(g^{-1}, \xi)$, $D^{\alpha(s)}(g, \xi)$ are a.e. (almost everywhere) continuous, differentiable quantities such that

$$\text{(i)} \quad \mu(e, \xi) = 1 \quad \forall \xi \in M \quad (e = id \text{ in } G), \tag{2.11}$$

$$\text{(ii)} \quad \mu(g_1 g_2, \xi) = \mu(g_2, \xi)\mu(g_1, g_2 \xi), \quad (g_1, g_2 \in G, \ \xi \in M), \tag{2.12}$$

$$\text{(i')} \quad D^{\alpha(s)}(e, \xi) = \mathbb{1} \quad \forall \xi \in M \quad (\mathbb{1} = id \text{ on } V), \tag{2.11'}$$

$$\text{(ii')} \quad D^{\alpha(s)}(g_1 g_2, \xi) = D^{\alpha(s)}(g_1, \xi) D^{\alpha(s)}(g_2, g_1^{-1} \xi) \quad (g_1, g_2 \in G, \ \xi \in M) \tag{2.12'}$$

[$D^{\alpha(s)}$ are $\alpha(s) \times \alpha(s)$ matrices].

Let us denote the action $g \cdot \xi$ on M by $g^{-1}\xi$. The multiplier $\mu^{\nu}(\ )D^{\alpha(s)}(\ )$ define a finite-dimensional representation of a subgroup, from which the induced representation of the group is obtained. The cocycle conditions (2.12) and (2.12') insure that the representation property

$$[U^{(s,\nu)}(g_1 g_2) f](\xi) = [U^{(s,\nu)}(g_1)(U^{(s,\nu)}(g_2) f)](\xi) \tag{2.13}$$

is satisfied. Since we wish to consider only unitary representations,

$$(U^{(s,\nu)}(g) f_1, U^{(s,\nu)}(g) f_2) = (f_1, f_2) \tag{2.14}$$

yields the condition of unitarity

$$\overline{[\mu(g^{-1}, g\xi)]^{\nu}}[\mu(g^{-1}, g\xi')]^{\nu} d\Omega(g\xi, g\xi') = d\Omega(\xi, \xi'). \tag{2.15}$$

For $d\Omega(\xi, \xi') = d\Omega_M(\xi)\delta_M(\xi, \xi')$ the condition implies

$$\frac{d\Omega_M(g^{-1}\xi)}{d\Omega_M(\xi)} = |\mu(g^{-1}, \xi)^{\nu}|^2 \tag{2.16}$$

and the Hilbert space reduces to the Hilbert space of all $\alpha(s)$-dimensional complex vector-valued functions which are square integrable with respect to the measure $d\Omega_M$.[7] We will shortly see that this case describes the principal series of unitary representations of $\overline{\mathrm{SO}_0(n,1)}$ for $n=4$. For $d\Omega(\xi, \xi') = K(1 - \xi \cdot \xi')\ d\Omega_M(\xi) d\Omega_M(\xi')$, with $\lambda = \sigma + n - 1$ and $-(n-1) < \sigma < 0$, we obtain the cases of the spin-zero complementary series of unitary representations of $\overline{\mathrm{SO}_0(n,1)}$.[8]

Finally we conclude this section with a classification of all continuous unitary irreducible representations of $\overline{\mathrm{SO}_0(4,1)}$.[9,10] They are in the notation of Ref. 9: (1) principal series: $D(\rho, l_{5,1}, l_{5,2})$

$$z_{52} + \tfrac{3}{2} = l_{5,2} = i\rho, \quad \rho > 0, \qquad l_{5,1} = 0, \tfrac{1}{2}, 1, \ldots; \tag{2.17}$$

(2) discrete series: $D(s; l_{5,1}, l_{5,2})$

$$l_{5,2} + 2 = 1 \ \text{or}\ l_{5,2} + 2 = 0, \qquad l_{5,1} = 1, 2, 3, \ldots, \qquad z_{5,2} = l_{5,2}; \tag{2.18}$$

(3) exceptional or complementary series: $D(e; l_{5,1}, l_{5,2})$

$$\text{(a)} \quad 0 \leqslant l_{5,2} < \tfrac{1}{2}, \quad l_{5,1} = 1, 2, 3, \ldots, \quad z_{5,2} + \tfrac{3}{2} = l_{5,2}; \tag{2.19}$$

$$\text{(b)} \quad 0 \leqslant l_{5,2} < \tfrac{3}{2}, \qquad l_{5,1} = 0, \qquad z_{5,2} + \tfrac{3}{2} = l_{5,2}; \tag{2.20}$$

(4) discrete series: $D(+; l_{5,1}, l_{5,2})$ and $D(-; l_{5,1}, l_{5,2})$

$$z_{5,2} = l_{5,2}, \qquad l_{5,2} + 2 = l_{5,1}, l_{5,1} - 1, \ldots, 1 \ \text{or}\ \tfrac{1}{2}, \qquad l_{5,1} = \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots, \tag{2.21}$$

$D(-;l_{5,1},l_{5,2})$ is the same as for $D(+;l_{5,1},l_{5,2})$. The second- and fourth-order Casimir operators of $\overline{\mathrm{SO}_0(4,1)}$ for these representations are given by the following multiples of the identity (cf. Böhm, Ref. 10):

$$
\begin{aligned}
-C_2 &= \{s(s+1)+c^2-\tfrac{9}{4}\}\cdot I;\\
-C_4 &= \{s(s+1)[c^2-\tfrac{1}{4}]\}\cdot I,
\end{aligned}
\tag{2.22}
$$

where

$$z_{5,2}+\tfrac{3}{2}=c \quad\text{and}\quad l_{5,1}=s. \tag{2.23}$$

III. CERTAIN EQUIVALENT REALIZATIONS OF THE UNITARY INDUCED IRREDUCIBLE REPRESENTATIONS OF $\overline{\mathrm{SO}_0(4,1)}$

Let $\mathbb{C}^{2s+1}$ be a $(2s+1)$-dimensional complex Hilbert space, in which a UIR, D^s,[11] of $M=\overline{\mathrm{SO}_0(3)}$ is realized. Consider the collection $\mathscr{C}^s$ of all functions on $\bar{G}=\overline{\mathrm{SO}_0(4,1)}$ which take values in $\mathbb{C}^{2s+1}$ and satisfy the following covariance condition[3,11]:

$$\mathscr{F}(g\tau)=\mathscr{F}(gman)=|a|^{3/2+c}D^s(m^{-1})\mathscr{F}(g), \tag{3.1}$$

$$(\mathscr{F}\in\mathscr{C}^s,\ g\in\bar{G},\ m\in M,\ a(t)\in A,$$
$$n\in N,\ \tau=man,\ |a|e^t).$$

The representation $U^{(s,c)}$ of $\bar{G}$, induced by the finite-dimensional representation $|a|^{-3/2-c}D^s(m)$ of MAN, is defined by

$$(U^{(s,c)}(g)\mathscr{F})(g')=\mathscr{F}(g^{-1}g');\quad g,g'\in\bar{G},\quad \mathscr{F}\in\mathscr{C}^s_\infty. \tag{3.2}$$

[A subspace with an appropriate square summability condition will later be specified. Temporarily we call it $\mathscr{C}^s_\infty\subset\mathscr{C}^s$. We take the representation (3.2) to be defined on this space.]

A function $\mathscr{F}\in\mathscr{C}^s_\infty$ [satisfying (3.1)] is essentially completely determined by its values on the subgroup $\tilde{N}=\mathbb{R}^3$.[12] Thus we obtain from (3.2) a representation on functions on $\mathbb{R}^3$ which are related to functions $\mathscr{F}$ by

$$\tilde{f}(x)=\mathscr{F}(\tilde{n}_x)\quad(\mathscr{F}\in\mathscr{C}^s_\infty,\quad \tilde{n}_x\in\tilde{N};\quad x\in\mathbb{R}^3).$$

The Bruhat decomposition[13]

$$g^{-1}\tilde{n}_x=\tilde{n}_{x'}m^{-1}(g,x)\,a^{-1}(g,x)\,n^{-1}(g,x) \tag{3.3}$$

induces an action of $\bar{G}$ on $\mathbb{R}\cup\{\infty\}$ [14] given by (see Appendix C)

$$\mathbb{R}^3\ni y^i\xrightarrow{g^{-1}}y^{i'}=(g^{-1}y)^i=\frac{g_j^{-1i}y^j+g_4^{-1i}y^2+g_5^{-1i}}{g_j^{-15}y^j+g_4^{-15}y^2+g_5^{-15}};$$

$$g=\begin{vmatrix} g^i_j & g^i_4 & g^i_5\\ g^4_j & g^4_4 & g^4_5\\ g^5_j & g^5_4 & g^5_5\end{vmatrix}\qquad(y=\sqrt{2}x).$$

The transformation property of the measure $d\Omega_x$ on $\mathbb{R}^3$ under this action of $\overline{\mathrm{SO}_0(4,1)}$ is given by[15]

$$d\Omega_{x'}=|a(g,x)|^{-3}\,d\Omega_x. \tag{3.4}$$

(Note: the transformed measure may be infinite for certain x's and g's.) The representation $\tilde{U}^{(s,c)}$ is found with the help of (3.1) to be

$$[\tilde{U}^{(s,c)}(g)\tilde{f}](x)=|a(g,x)|^{-3/2-c}D^s(m(g,x))\tilde{f}(g^{-1}x), \tag{3.5}$$

where[16]

$$n(g,x)\,a(g,x)\,m(g,x)=\tilde{n}_x^{-1}g\tilde{n}_{g\cdot x}. \tag{3.6}$$

Next we transfer these representations over to spaces of functions on the three-sphere [Eq. (2.3′) for $n=4$]. Embed $\mathbb{R}^3$ into a four-dimensional Euclidean space $\mathbb{R}^4$ in such a way that $\mathbb{R}^3$ is a hyperplane passing through the point $\mathring{u}=(0,0,0,-1)$ (see Fig. 1). Then perform stereographic projection onto the unit sphere minus the north pole in $\mathbb{R}^4$, $S_3-\{N\}$, as shown in Fig. 1. The equations of stereographic projection read[17]

$$
\begin{aligned}
x_i &= u_i/(1+u_4)=\hat{\tau}_i^{-1}(u),\\
\left.\begin{aligned}
u_i &= 2x_i/(x^2+1),\quad i=1,2,3,\\
u_4 &= (x^2-1)/(x^2+1)
\end{aligned}\right\} &= \hat{\tau}_\mu(x).
\end{aligned}
\tag{3.7}
$$

The measure on S_3 is in terms of these stereographic projection coordinates[17]:

$$d\Omega_u=[1/(1+x^2)^3]\,d\Omega_x. \tag{3.8}$$

Now we define the following $\mathbb{C}^{2s+1}$ valued functions in terms of corresponding functions on $\mathbb{R}^3$:

$$\hat{f}(u)=[1+x^2]^{3/2+c}\tilde{f}(x). \tag{3.9}$$

The stereographic projection (3.7) induces, through the action of $\bar{G}$ on $\mathbb{R}^3\cup\{\infty\}$ [Eq. (3.3)], an action of $\bar{G}$ on S_3 given by[1,8,18]

$$u^\mu\xrightarrow{g}u^{\mu'}=(\hat{g}^\mu_5+\hat{g}^\mu_\nu u^\nu)/(\hat{g}^5_5+\hat{g}^5_\mu u^\mu), \tag{3.10}$$

with

$$\hat{g}=\hat{Q}_0^{-1}g\hat{Q}_0=\begin{vmatrix} \hat{g}^i_j & \hat{g}^i_4 & \hat{g}^i_5\\ \hat{g}^4_j & \hat{g}^4_4 & \hat{g}^4_5\\ \hat{g}^5_j & \hat{g}^5_4 & \hat{g}^5_5\end{vmatrix}\in\mathrm{SO}_0(4,1)$$

[$\hat{g}$ corresponds to $g\in G$ through an automorphism of $\mathrm{SO}_0(4,1)$ induced by stereographic projection—compare Eq. (3.17) and see Appendix C.]. The measure $d\Omega_u$ on S_3 transforms as follows under this action of $\bar{G}$ on S_3 (see I):

$$
\begin{aligned}
d\Omega_{gu} &= |[\hat{g}^5_5+\hat{g}^5_\mu u^\mu]|^{-3}\,d\Omega_u\\
&= |\mu(\hat{g},u)|^{-3}\,d\Omega_u.
\end{aligned}
\tag{3.11}
$$

Using (3.4), (3.8), and (3.11) we obtain

$$|\mu(\hat{g}^{-1},u)|=[((1+x'^2)/(1+x^2))|a(g,x)|]. \tag{3.12}$$

Combining this result with (3.5) and (3.9) we obtain for the action of $\bar{G}$ on $\hat{f}$

$$
\begin{aligned}
&(\hat{U}(g)\hat{f})(u)\\
&\quad=|\mu(\hat{g}^{-1},u)|^{-3/2-c}D^s(m(\hat{g},u))\hat{f}(\hat{g}^{-1}u),
\end{aligned}
\tag{3.13}
$$

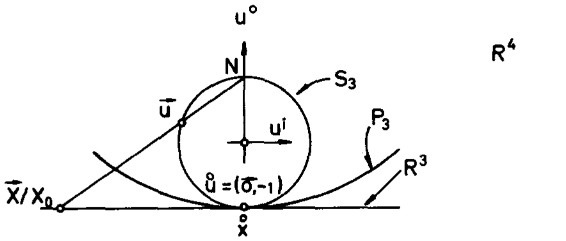


FIG. 1. Projections of the unit sphere in $\mathbb{R}^4$, S_3 onto $\mathbb{R}^3$, and P_3, the unit parabola in $\mathbb{R}^4$. Equations (3.7) follow for $x_0=\frac{1}{2}$.

where $D^s(m(\hat{g}, u)) = D^s(m(g, x))$. [(2.12′) is satisfied for $D^s(m(\hat{g},u))$ because $\hat{g}_1^{-1}u = \hat{\tau}(g_1^{-1}x)$—see Appendix C.]

Finally we described the representation of $\overline{\mathrm{SO}_0(4,1)}$ acting on functions on the hyperboloid T_3. Let

$$T_3 = \{p_\mu \in M^{3,1} \mid p^\mu p_\mu = -p^{1^2} - p^{2^2} - p^{3^2} + p^{4^2} = 1\}$$

be the unit hyperboloid of two sheets in $M^{3,1}$. Define a projection τ, mapping T_3 into S_3 as follows:

$$S_3 \ni u \to p = \tau u \in T_3\colon\ p^\mu[1/u^4, (-u^i/u^4)], \tag{3.14}$$

$$T_3 \ni p \to u = \tau^{-1}p \in S_3\colon\ u^\mu = [1/p^4, (-p^i/p^4)]. \tag{3.15}$$

The mapping is exhibited in Fig. 2. It induces an action of $\bar{G}$ on T_3 which is given by

$$p^{\mu'} = (\bar{g}_5^\mu + \bar{g}_\nu^\mu p^\nu)/(\bar{g}_5^5 + \bar{g}_\mu^5 p^\mu), \tag{3.16}$$

where (see I)

$$\bar{g} = \begin{vmatrix} \hat{g}_j^i & -\hat{g}_5^i & \hat{g}_4^i \\ -\hat{g}_j^5 & \hat{g}_5^5 & \hat{g}_4^5 \\ -\hat{g}_j^4 & \hat{g}_5^4 & \hat{g}_4^4 \end{vmatrix}. \tag{3.17}$$

The measure $d\Omega_p$ on T_3 transforms as follows under the action of $\bar{G}$ on T_3:

$$d\Omega_{\bar{g}p} = |\bar{g}_5^5 + \bar{g}_\mu^5 p^\mu|^{-3}\, d\Omega_p. \tag{3.18}$$

Certainly Eqs. (3.16) and (3.18) are not defined for certain p^μ's and $\bar{g}$'s. This means that in order to consider the action of $\bar{G}$ on T_3 we must first compactify it by the adjunction of a surface at infinity. Still Eq. (3.18) will be undefined on certain sets of measure zero. But this does not affect the unitary representations of G which we will construct on T_3.

Now we consider the mapping $\bar{\Pi}$ from functions on S_3 to functions on T_3 given by

$$\begin{aligned} \hat{f}(u) &\to \bar{f}(p) = (\bar{\Pi}\hat{f})(p) = (1/|p^4|^{3/2+c})\hat{f}(\tau^{-1}p), \\ \bar{f}(p) &\to \hat{f}(u) = (\bar{\Pi}^{-1}\bar{f})(u) = (1/|u^4|^{3/2+c})\bar{f}(\tau u). \end{aligned} \tag{3.19}$$

The measures on S_3 and T_3 transform as follows under τ (see I):

$$d\Omega_u = (1/|p^4|^3)\, d\Omega_p. \tag{3.20}$$

We obtain for the representation $U^{(s,c)}$ on T_3 the following expression:

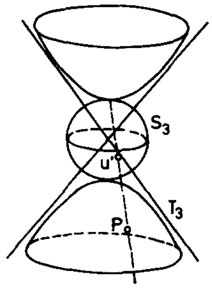


FIG. 2. Projection of the unit sphere S_3 onto the two-sheeted unit hyperboloid T_3.

$$[\bar{U}(g)\bar{f}](p) = |\mu(\bar{g}^{-1}, p)|^{-3/2-c} D^s(m(\bar{g}, p))\bar{f}(\bar{g}^{-1}p), \tag{3.21}$$

where

$$\mu(\bar{g}^{-1}, p) = (\bar{g}_5^{5^{-1}} + \bar{g}_\mu^{5^{-1}} p^\mu), \tag{3.22}$$

$\bar{g}^{-1}p$ is given by (3.16) and $m(\bar{g}, p)$ is given by (3.6) for $x = x(u(p))$ obtained from Eqs. (3.7) and (3.15). [We note that the $\bar{g}$ in $m(\bar{g}, p)$ is an element of the covering group $\bar{G}$.] Using the above realizations of $U^{(s,c)}$ and T_3 given by Eqs. (3.13) and (3.21) we can obtain the principal, discrete, and complementary series of UIR's of $\overline{\mathrm{SO}_0(4,1)}$ on certain Hilbert spaces of functions defined on S_3 and T_3.

A. Principal series

Let

$$\hat{\mathscr{H}} = \mathscr{L}^2(S_3) \otimes \mathbb{C}^{2s+1} \tag{3.23}$$

and let $c = i\rho$ in (3.13). Then $\hat{U}^{(s,i\rho)}(g)\hat{f}(u)$ $(\hat{f} \in \hat{\mathscr{H}})$ in (3.13) defines a UIR of $\bar{G}$ on $\hat{\mathscr{H}}$, in which the inner product is given by

$$(\hat{f}, \hat{g}) = \int_{S_3} \sum_{i=-s}^{s} \overline{\hat{f}_i(u)}\, \hat{g}_i(u)\, d\Omega_u. \tag{3.24}$$

[Unitarity of (3.13) follows from the unitarity of $D^s(m(\hat{g},u))$ and the transformation property of the Jacobian; see I.] Likewise $U_{(\bar{g})}^{(s,i\rho)}\bar{f}(p)$ in (3.21) defines a unitary representation of $\overline{\mathrm{SO}_0(4,1)}$ on

$$\bar{\mathscr{H}} = \mathscr{L}^2(T_3) \otimes \mathbb{C}^{2s+1}. \tag{3.25}$$

The mapping $\bar{\Pi}$ [Eq. (3.19)] is an isometric isomorphism from $\hat{\mathscr{H}}$ onto $\bar{\mathscr{H}}$ which intertwines the two representations $\hat{U}^{(s,+i\rho)}(\bar{G})$ and $\bar{U}^{(s,+i\rho)}(\bar{G})$. It follows from their unitary equivalence with $\tilde{U}^{(s,+i\rho)}(\bar{G})$ [Eq. (3.5)] on

$$\tilde{\mathscr{H}} = \mathscr{L}^2(\mathbb{R}^3) \otimes \mathbb{C}^{2s+1} \tag{3.26}$$

that they are realizations of the principal series

$$D(\rho; s, i\rho)$$

which satisfy the definition in Sec. II. [Unitary equivalence of $\hat{U}(\bar{G})$ with $\tilde{U}(\bar{G})$ follows in a similar way as equivalence of $\hat{U}(\bar{G})$ with $\bar{U}(\bar{G})$, i.e., use (3.5) and (3.9). For a proof of irreducibility see Ref. 4.]

B. Complementary series

Now let $\hat{\mathscr{H}}_c$ be the space of all measurable $2s+1$ complex-valued functions on S_3 which are finite with respect to the following scalar product:

$$(\hat{f}, \hat{g}) = \int_{S_3}\int_{S_3} \sum_{i=-s}^{s} \overline{\hat{f}_i(u_1)}\, \hat{K}_{ij}^c(u_1, u_2) \times \hat{g}_j(u_2)\, d\Omega_{u_1}\, d\Omega_{u_2}. \tag{3.27}$$

Using (3.8) and (3.9) we obtain equivalently

$$(\tilde{f}, \tilde{g}) = \int_{\mathbf{R}_3}\int_{\mathbf{R}_3} \sum_{i=-s}^{s} \overline{\tilde{f}(x_1)}\, \tilde{K}_{ij}^c(x_1, x_2) \times \tilde{g}_j(x_2)\, d\Omega_{x_1}\, d\Omega_{x_2}, \tag{3.28}$$

where

$$\tilde{K}^c_{ij}(x_1, x_2) = [1 + x_1^2]^{+3/2+c}[1 + x_2^2]^{+3/2+c} \times \hat{K}^c_{ij}(u_1(x_1), u_2(x_2)) \quad (3.29)$$

and $u(x)$ is given by (3.7). $\tilde{K}^c_{ij}(x_1, x_2)$ is constructed so that (3.5) is unitary with respect to (3.28).[19] This requirement imposes restrictions on $\tilde{K}^c_{ij}(x_1, x_2)$ and it can be shown to have explicitly the following form[20]:

$$\tilde{K}^c_{ij}(x_1, x_2) = \tilde{K}^c_{ij}(x_1 - x_2) = \sum_{l=-s}^{s} \frac{K(s,c)}{(x_{12}^2)^{3/2-c}} D^s_{il}((\mathbf{x}_1 - \mathbf{x}_2)\cdot\boldsymbol{\sigma}) D^s_{lj}(i\pi) \quad (\mathbf{x}_{12} = \mathbf{x}_1 - \mathbf{x}_2)$$

[see Ref. 19 for the definition of $D(i\pi)$ and $\boldsymbol{\sigma}$]. The normalization factor $K(c,s)$ for s integral is given in Ref. 4. This expression is the configuration space form of the conformally invariant propagator for a spin s field in three-dimensional Euclidean space-time (see Ref. 3). In order that (3.27) be finite and positive, it is required that[21,22]

$$-\tfrac{3}{2} < c < \tfrac{3}{2} \quad \text{for } s = 0, \qquad -\tfrac{1}{2} < c < \tfrac{1}{2} \quad \text{for } s = 1,2,3,\dots . \quad (3.30)$$

Concerning the discrete series, we note that they can also be described as multiplier representations given by Eq. (3.5).[4,23] Mackey has conjectured that, at least for certain semisimple Lie groups, the discrete series representations have a correspondence with projective representations of an associated semidirect product which is a limiting form of the semisimple Lie group (group contraction).[24] Perhaps a study of the conjectured correspondence would provide a more geometrical explanation for the absence of half-integral spins for the complementary series, which can be viewed as analytical continuations of the principal series, characterized by the number $(i\rho, s)$, to real values of $i\rho$ (Ref. 3, 4, and 8), and also for the absence of the discrete series in the harmonic analysis of the left regular representation of $SO_0(4,1)$ on $SO_0(4,1)/SO_0(4)$ (Ref. 23).

IV. REALIZATIONS OF THE PRINCIPAL SERIES UIR's of $\overline{SO_0(4,1)}$ ON UIR SPACES OF $\overline{E(4)}$ and $\overline{\mathscr{P}}$

In I we constructed realizations of the principal series UIR's of $\overline{SO_0(4,1)}$ on UIR spaces of the simply connected covering group of the four-dimensional Euclidean group, $\overline{E(4)}$, and on UIR spaces of the simply connected covering group of the Poincaré group $\overline{\mathscr{P}}$. We recall the constructions here. (Here S_3 and T_3 denote a sphere and hyperboloid of radii m.)

A. Realizations on UIR spaces of $\overline{E(4)}$

Consider the generators of the Clifford algebra, corresponding to the Riemannian spaces $\mathbb{R}^4$ defined by[25]

$$\{\hat{\gamma}^\mu, \hat{\gamma}^\nu\} = 2\delta^{\mu\nu} \quad (4.1)$$

and realized on $\mathbb{C}^4$ as 4×4 Hermitian matrices (choose $\hat{\gamma}_0 = \gamma_0$, $\hat{\gamma}_i = \alpha_i = \gamma_0\gamma_i$, where γ_0, γ_i are the customary Dirac matrices in the convention of Ref. 1). Next introduce the tensor product space

$$\otimes_{2s}\mathbb{C}^4 = \underbrace{\mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \cdots \otimes \mathbb{C}^4}_{2s \text{ times}} \quad (4.2)$$

and consider the quantities[26]

$$\hat{\gamma}^\mu_{(k)} = I \otimes I \otimes \cdots \otimes \hat{\gamma}^\mu \otimes \cdots \otimes I; \quad (4.3)$$

where $\hat{\gamma}^\mu$ occupies the kth slot. Using the properties of the tensor product of linear operators we can verify the validity of the following equations, with the help of (4.1)[27]:

$$\{\hat{\gamma}^\mu_{(k)}, \hat{\gamma}^\nu_{(l)}\} = 2\delta^{\mu\nu}\delta_{kl}. \quad (4.4)$$

Now define the operators[28]

$$\hat{\Gamma}^\mu = \frac{1}{2}\sum_{k=1}^{2s} \hat{\gamma}^\mu_{(k)} \quad (4.5)$$

and

$$\hat{S}_{\mu\nu} = \sum_k \hat{S}^{(k)}_{\mu\nu} = \frac{-i}{4}\sum_k [\hat{\gamma}_{\mu(k)}, \hat{\gamma}_{\nu(k)}]. \quad (4.6)$$

We verify

$$[\hat{S}_{\mu\nu}, \hat{\Gamma}_\rho] = -i(\delta_{\nu\rho}\hat{\Gamma}_\mu - \delta_{\mu\rho}\hat{\Gamma}_\nu),$$
$$[\hat{S}_{\mu\nu}, \hat{S}_{\rho\sigma}] = i(\delta_{\mu\rho}\hat{S}_{\nu\sigma} + \delta_{\nu\sigma}\hat{S}_{\mu\rho} - \delta_{\nu\rho}\hat{S}_{\mu\sigma} - \delta_{\mu\sigma}\hat{S}_{\nu\rho}), \quad (4.7)$$
$$[\hat{\Gamma}_\mu, \hat{\Gamma}_\nu] = i\hat{S}_{\mu\nu},$$

so that $\hat{\Gamma}_\mu$ and $\hat{S}_{\mu\nu}$ are the Hermitian generators of a representation of $\overline{SO_0(5)}_{\hat{\Gamma}_\mu, \hat{S}_{\mu\nu}}$ on $\otimes_{2s}\mathbb{C}^4$, and $-i\hat{\Gamma}_\mu, \hat{S}_{\mu\nu}$ are the generators of a finite-dimensional representation of $\overline{SO_0(4,1)}_{-i\hat{\Gamma}_\mu, \hat{S}_{\mu\nu}}$. The representation of the group, $\overline{SO_0(4,1)}_{-i\hat{\Gamma}_\mu, \hat{S}_{\mu\nu}}$, is obtained by exponentiating

$$\overline{SO_0(4,1)} \ni A \to D(\hat{A}) = \exp(-(i/2)\,\omega^{ab}\hat{S}_{ab}) \quad (4.8)$$

with $\hat{S}_{5\mu} = -i\hat{\Gamma}_\mu$ and $\hat{S}_{\mu\nu}$ given by (4.6).[29] A basis for the Lie algebra representation is given by the quantities $i\hat{S}_{ab}$.

Next consider the space $\hat{\mathscr{G}}^{(s)}$ of all C^∞ differentiable tensors of rank $2s$, $\hat{\psi}(u; \zeta_1\zeta_2\cdots\zeta_{2s})$, from S_3 into $\otimes_{2s}\mathbb{C}^4$, which are totally symmetric in their $2s$ four-valued variables $\zeta = \zeta_1\cdots\zeta_{2s}$. (Sometimes we denote the collection $\zeta_1\cdots_{2s}$ simply by ζ when no confusion arises.) The representation of $\overline{SO_0(4,1)}$ defined by (4.8) is also a representation on $\hat{\mathscr{G}}^{(s)}$ as shown in I. Let $\hat{P}_\mu$ be the operator of multiplication by u_μ on $\hat{\psi}$: $\hat{P}_\mu\hat{\psi}(u;\zeta) = u_\mu\hat{\psi}(u;\zeta)$; and introduce the generalized Dirac equation[30]

$$(\hat{\Gamma}^\mu\hat{P}_\mu - \tfrac{1}{2}m)\,\hat{\psi}(u;\zeta) = 0 \quad (\hat{P}^\mu\hat{P}_\mu\hat{\psi} = m^2\hat{\psi}). \quad (4.9)$$

Denote the subspace of all $\hat{\psi} \in \hat{\mathscr{G}}^{(s)}$ which satisfy this equation by "$\hat{\mathscr{R}}^{(s)}$." We introduce the following inner product on "$\hat{\mathscr{R}}^{(s)}$":

$$(\hat{\psi}, \hat{\phi}) = \int_{S_3} d\Omega_u \sum_\zeta \overline{\hat{\psi}(u;\zeta)}\,\hat{\phi}(u;\zeta). \quad (4.10)$$

Let $\hat{\mathscr{R}}^{(s)}$ be the Hilbert space completion of "$\hat{\mathscr{R}}^s$" with respect to this inner product. Equation (4.9) with its Euclidean mass-shell condition

$$\hat{P}_\mu\hat{P}^\mu\hat{\psi} = m^2\hat{\psi}$$

are completely equivalent to the $\overline{E(4)}$ Bargmann–Wigner equations

$$\{\hat{\gamma}^\mu_{(k)}\hat{P}_\mu - m\}\,\hat{\psi}(u;\zeta) = 0 \quad (k = 1,\dots,2s) \quad (4.11)$$

on $\hat{\mathscr{R}}^{(s)}$.

Using (4.9) we obtain the important operator identity (cf. I)

$$(1/\lambda)\, B_\mu^{(s)} = (1/m)\, u^\rho \hat{S}_{\rho\mu} = i\hat{\Gamma}_\mu - i(2s/2m)\, u_\mu \quad (4.12)$$

valid on $\hat{\mathscr{R}}^{(s)}$. With the help of this result we have constructed in I an arbitrary principal series UIR of $\overline{\mathrm{SO}_0(4,1)}$ on $\hat{\mathscr{R}}^{(s)}$. It is defined as follows:

$$\overline{\mathrm{SO}_0(4,1)} \ni g \to \hat{U}(g):(\hat{U}(g)\,\hat{\psi})(u)$$
$$= \frac{1}{|\mu(\hat{g}^{-1}, u/m)|^{3/2-s+i\rho}} D(\hat{g})\, \hat{\psi}(m\hat{g}^{-1}u/m). \quad (4.13)$$

[For simplicity we have omitted the ζ indices on the $\hat{\psi}$ and the matrix $D(\hat{g})$.] The $\hat{g}$ occurring in the $D(\hat{g})$ is an element of $\overline{\mathrm{SO}_0(4,1)}$, but the $\hat{g}^{-1}$ occurring in the multiplier factor and to the right of the $\hat{\psi}$ is an element of $\mathrm{SO}_0(4,1)$ and its matrix form is given by the expression below Eq. (3.10) of Sec. III.[31] Here, $\hat{g}^{-1}(u/m)$ is given by (3.10) with the understanding that $(\hat{g}^{-1})^a_b$ replaces $\hat{g}^a_b$ and u/m replaces u; $\mu(\hat{g}^{-1}, u/m)$ is defined in (3.11). In I it is proven that (4.13) defines a principal series of UIR of $\overline{\mathrm{SO}_0(4,1)}$ characterized by ρ and s, i.e., it defines a $D(\rho; s, i\rho)$.

We have the result that $\hat{\mathscr{R}}^{(s)}$ is the carrier space for UIR of $\overline{E(4)}$.[32] The relevant UIR's of $\overline{E(4)}$ are characterized by two numbers m and s ($m>0, s=0, \frac{1}{2}, 1,\ldots$) which are related to the second- and fourth-order Casimir operators of $\overline{E(4)}$ as follows:

$$\hat{P}_\mu \hat{P}^\mu = m^2 I, \quad \hat{W} = m^2 s(s+1) I \quad (4.14)$$

[$\hat{P}_\mu$ and $\hat{L}_{\mu\nu}$ are Hermitian generators of $\overline{E(4)}$ and $\hat{W} = -\hat{w}_\mu \hat{w}^\mu$ with $\hat{w}_\mu$ the $\overline{E(4)}$ Pauli–Lublanski four-vector: $\hat{w}_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} \hat{P}^\nu \hat{L}^{\rho\sigma}$]. The UIR of $\overline{E(4)}$ on $\hat{\mathscr{R}}^{(s)}$ is defined by

$$[\hat{U}([a, \hat{\Lambda}])\,\hat{\psi}](u) = e^{iu_\mu a^\mu} D(\hat{\Lambda})\, \hat{\psi}(\hat{\Lambda}^{-1}u) \quad (4.15)$$

[$a \in T_4$, $\hat{\Lambda} \in \overline{\mathrm{SO}_0(4)}$], where $D(\hat{\Lambda})$ is given by (4.8) for $\hat{\Lambda} \in \overline{\mathrm{SO}_0(4)}$. The representation of $\overline{E(4)}$ so constructed is the UIR characterized by the numbers m and s.

In (4.13) and (4.15) the functions $\hat{\psi}(u; \zeta_1 \cdots \zeta_{2s})$ are the components of a vector $\hat{\psi} \in \hat{\mathscr{R}}^{(s)}$ with respect to the generalized basis $|u\rangle \otimes |\zeta_1\rangle \otimes \cdots \otimes |\zeta_{2s}\rangle$ of $\mathscr{L}^2(S_3) \otimes (\otimes_{2s} \mathbb{C}^4)$. There exists a more convenient basis for the subspace $\hat{\mathscr{R}}^{(s)}$; it is the so-called spinor basis $|f_{j_3}^{jj_0}(u)\rangle$ which corresponds, in the reduction with respect to $\overline{\mathrm{SO}_0(4)}_{\hat{S}_{\mu\nu}}$, to

$$|f_{j_3}^{jj_0}(u)\rangle = |u\rangle \otimes |jj_3; (\pm j_0 c = s)\rangle. \quad (4.16)$$

The $|j_3 j; (\pm j_0 c)\rangle$ satisfy the simultaneous eigenvalue equations[33]

$$\begin{aligned}
&\tfrac{1}{2}\hat{S}_{\mu\nu}\hat{S}^{\mu\nu}|j_3 j; (\pm j_0 c)\rangle \\
&\quad = (-1 + c^2 + j_0^2)|j_3 j; (\pm j_0 c)\rangle, \\
&\hat{S}_{0i}\hat{S}_i |j_3 j; (\pm j_0 c)\rangle = \pm c j_0 |j_3 j; (\pm j_0 c)\rangle, \\
&\tfrac{1}{2}\hat{S}_{ij}\hat{S}^{ij}|j_3 j; (\pm j_0 c)\rangle = j(j+1)|j_3 j; (\pm j_0 c)\rangle, \qquad (4.17)\\
&\hat{S}_{12}|j_3 j; (\pm j_0 c)\rangle = j_3 |j_3 j; (\pm j_0 c)\rangle \\
&\quad (\hat{S}_i = \tfrac{1}{2}\epsilon_{ijk}\hat{S}_{jk}).
\end{aligned}$$

The components of a vector $\hat{\psi} \in \hat{\mathscr{R}}^{(s)}$ with respect to this symmetry-suited basis $|f_{j_3}^{jj_0}(u)\rangle$ are denoted by $\hat{\psi}(u; j_3 jj_0)$, and the actions of $\overline{\mathrm{SO}_0(4,1)}$ and $\overline{E(4)}$ [Eqs. (4.13) and (4.15)] on the $\hat{\psi}(u; j_3 jj_0)$ take the particularly simple forms[34]

$$(\hat{U}(g)\,\hat{\psi})(u; j_3 jj_0) = \frac{1}{|\mu(\hat{g}^{-1}, u/m)|^{3/2-s+i\rho}} \times D^{jj_0 j' j_0'}_{j_3 j_3'}(\hat{g})\, \hat{\psi}\Big(m\hat{g}^{-1}\frac{u}{m}; j_3' j' j_0'\Big) \quad (4.13')$$

$$[\hat{U}([a, \hat{\Lambda}])\,\hat{\psi}](u; j_3 jj_0) = e^{iu_\mu a^\mu} D^{jj'}_{j_3 j_3'}(\hat{\Lambda})\, \hat{\psi}(\hat{\Lambda}^{-1}u; j_3' j' j_0) \quad (4.15')$$

(summation over repeated indices is meant!), where

$$D^{jj_0 j' j_0'}_{j_3 j_3'}(\hat{g}) = \exp\{ i(\omega^{ab}/2)(\hat{S}_{ab})^{jj_0 j' j_0'}_{j_3 j_3'} \}. \quad (4.18)$$

We have for $s = \frac{1}{2}$ the well-known result

$$\hat{S}_i = \frac{1}{2}\begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix},$$
$$\hat{S}_{4i} = \frac{1}{2}\begin{pmatrix} \sigma_i & 0 \\ 0 & -\sigma_i \end{pmatrix}. \quad (4.19)$$

The matrix elements

$$(\hat{S}_{\mu\nu})^{jj'}_{j_3 j_3'} \quad \text{(for fixed } j_0) \quad (4.20)$$

in the general case (s arbitrary) are determined by ($c = s$)

$$\hat{J}_q |j_3 j; (\pm j_0, c)\rangle = -[j(j+1)]^{1/2} \times \langle 1qjj_3 | 1jj_3 + q\rangle |j_3 + qj, (\pm j_0, c)\rangle, \quad (4.21)$$

$$\begin{aligned}
&i\hat{N}_q |j_3 j; (\pm j_0 c)\rangle \\
&= \{[(j+1)^2 - j_0^2][(j+1)^2 - c^2]/(2j+3)(j+1)\}^{1/2} \\
&\quad \times \langle 1qjj_3 | 1jj + 1j_3 + q\rangle |j_3 + q, j+1; (\pm j_0 c)\rangle \\
&\quad \pm i\{ j_0 c/[j(j+1)]^{1/2}\} \langle 1qjj_3 | 1jjj_3 + q\rangle \\
&\quad \times |j_3 + q, j; (\pm j_0 c)\rangle \\
&\quad - [(j^2 - j_0^2)(j^2 - c^2)/(2j-1)j]^{1/2} \\
&\quad \times \langle 1qjj_3 | 1jj - 1j_3 + q\rangle |j_3 + q, j; (\pm j_0, c)\rangle. \qquad (4.22)
\end{aligned}$$

with $q = 0, \pm 1$ with $\hat{J}_0 = \hat{S}_{12}$, $\hat{J}_{\pm 1} = +(1/\sqrt{2})(\hat{S}_{23} \pm i\hat{S}_{31})$ and $\hat{N}_0 = \hat{S}_{34}$, $\hat{N}_{\pm 1} = \mp(1/\sqrt{2})\,(\hat{S}_{14} \pm i\hat{S}_{24})$ ($\langle 1qjj_3 | 1jjj_3 + q\rangle$ are the Clebsch–Gordan coefficients as given in Edmonds[35]). The matrix elements $(\hat{\Gamma}_\mu)^{jj_0 j' j_0'}_{j_3 j_3'}$ can easily be obtained from the expressions on pp. 203–205 of Ref. 10 (Böhm). For $s = \frac{1}{2}$ they are

$$\hat{\Gamma}_4 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (4.23)$$

$$\hat{\Gamma}_i = \frac{i}{2}\begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}. \quad (4.24)$$

The usual description of the positive mass, arbitrary spin UIR's of $\overline{E(4)}$ is given by induced representations: We introduce the Hilbert space

$$\hat{\mathscr{H}}(m, s) = \mathscr{L}^2(S_3) \otimes \mathbb{C}^{2s+1} \quad (4.25)$$

consisting of all complex $2s+1$ vector-valued functions $\hat{\psi}(u; s_3)$ on S_3 which satisfy

$$(\hat{\psi}, \hat{\psi}) = \sum_{s_3=-s}^{s} \int_{S_3} d\Omega_u \, \overline{\hat{\psi}(u; s_3)} \hat{\psi}(u; s_3) < \infty. \tag{4.26}$$

The representation of $\overline{E(4)}$ on $\mathscr{H}(m, s)$ is given by

$$[\hat{U}_W([a, \hat{\Lambda}])\hat{\psi}](u; s_3) = e^{iu_\mu a^\mu} \sum_{s_3'} D_{s_3 s_3'}(R(\hat{\Lambda}, u)) \hat{\psi}(\hat{\Lambda}^{-1}u; s_3'), \tag{4.27}$$

where

$$R(\hat{\Lambda}, u) = \hat{L}(u) \hat{\Lambda} \hat{L}^{-1}(\hat{\Lambda}^{-1}u) \tag{4.28}$$

is the $\overline{E(4)}$ Wigner rotation, $\hat{L}^{-1}(u)$ is the $\overline{E(4)}$ analog of the Lorentz boost, and $D_{s_3}^{s_3'}$ denotes the usual $(2s+1)\times(2s+1)$ linear matrix representation of $\overline{\mathrm{SO}_0(3)}$. The representations of $\overline{E(4)}$ on $\hat{\mathscr{R}}^{(s)}$ and $\hat{\mathscr{H}}(m, s)$, given by (4.15) [or (4.15′)] and (4.27), respectively, are equivalent unitary representations.[36] There exists an intertwining operator

$$\hat{F}: \hat{\mathscr{H}}(m, s) \to \hat{\mathscr{R}}^{(s)} \tag{4.29}$$

which intertwines the two representations (4.15) and (4.27). The intertwining operator has a particularly transparent form between components of $\hat{\psi} \in \hat{\mathscr{R}}^{(s)}$ with respect to the spinor basis (see I) and the $\hat{\psi}(u; s_3)$ of $\hat{\mathscr{H}}(m, s)$. First we write every $\hat{\psi}(u; s_3)$ as

$$\hat{\phi}(u; s_3 s n) = \begin{pmatrix} \hat{\psi}(u; s_3) \\ 0 \end{pmatrix} \begin{matrix} \}n = n_{\max} = s \\ \}n \neq n_{\max} \end{matrix} \tag{4.30}$$

[n labels the eigenvalues of $\hat{\Gamma}_4$—see Ref. (33) for details]. We then have

$$\hat{\mathscr{H}}(m, s) \ni \hat{\phi} \to (\hat{F}\hat{\phi}) \in \hat{\mathscr{R}}^{(s)},$$

$$\hat{\phi}(u; s_3 s n) \to (\hat{F}\hat{\phi})(u; j_3 j j_0) = \sum_{n s s_3} D_{j_3 s}^{js(j_0)}(\hat{L}^{-1}(u)) \, Q^{j_0}(ns) \, \hat{\phi}(u; s_3 s n), \tag{4.31}$$

and inversely

$$\hat{\mathscr{R}}^{(s)} \ni \hat{\psi} \to (\hat{F}^{-1}\hat{\psi}) \in \hat{\mathscr{H}}(m, s),$$

$$\hat{\psi}(u; j_3 j j_0) \to (\hat{F}^{-1}\hat{\psi})(u; s_3 s n) = \sum_{j j_0 j_3} D_{s_3 j_3}^{sj(j_0)}(\hat{L}(u)) \, Q^{-1 j_0}(ns) \, \hat{\psi}(u; j_3 j j_0). \tag{4.32}$$

The transformation matrix $Q^{j_0}(ns)$ is a consequence of our particular choices of coordinate systems in $\otimes_{2s} \mathbb{C}^4$ which are associated with the representation of $\overline{E(4)}$ on $\hat{\mathscr{R}}^{(s)}$ and on $\hat{\mathscr{H}}(m, s)$. Its explicit form is not needed for our purposes. For the case of $s = \frac{1}{2}$ the matrix is explicitly calculated in Ref. 37 for the SL(2, C) case.

B. Realization on UIR spaces of $\overline{\mathscr{P}}$

Now we consider the Clifford algebra corresponding to the Minkowski space, $M^{3,1}$. Four of the generators of the Clifford algebra are defined by

$$\{\gamma^\mu, \gamma^\nu\} = 2n^{\mu\nu}, \quad n^{\mu\nu} = \mathrm{diag}(-1, -1, -1, 1). \tag{4.33}$$

On $\otimes_{2s} \mathbb{C}^4$ we introduce the four quantities $\gamma_{(k)}^\mu$ defined by

$$\{\gamma_{(k)}^\mu, \gamma_{(l)}^\nu\} = 2n^{\mu\nu}\delta_{kl} \quad \text{[compare (4.3)]}, \tag{4.34}$$

and also the quantities

$$\Gamma_\mu = \frac{1}{2} \sum_{k=1}^{2s} \gamma_{(k)}^\mu.$$

The $-i\Gamma_\mu$ and

$$S_{\mu\nu} = i[\Gamma_\mu, \Gamma_\nu] \tag{4.35}$$

define a representation of $\overline{\mathrm{SO}_0(4,1)}_{-i\Gamma_\mu, S_{\mu\nu}}$ on $\otimes_{2s} \mathbb{C}^4$ given by

$$\overline{\mathrm{SO}_0(4,1)} \ni A \to D(\overline{A}) = \exp(-i\tfrac{1}{2}\omega^{ab} S_{ab}) \tag{4.36}$$

with $S_{5\mu} = -i\Gamma_\mu$ and $S_{\mu\nu}$ given by (4.35), and $\overline{A}$ defined in terms of A through (3.17).

Next consider the Hilbert space completion of the set of all normalizable functions from T_3 into $\otimes_{2s} \mathbb{C}^4$ which satisfy the equations

$$\begin{aligned} &(\Gamma^\mu P_\mu - \tfrac{1}{2} m) \, \psi(p; \xi_1 \cdots \xi_{2s}) = 0, \\ &(P^\mu P_\mu - m^2) \, \psi(p; \xi_1 \cdots \xi_{2s}) = 0, \end{aligned} \tag{4.37}$$

and which are totally symmetric in all ξ variables. Here, P_μ is multiplication by p_μ. The norm is defined as

$$(\psi, \psi) = \int_{T_3} \sum_{\xi} \bar{\psi} \gamma_{(1)}^4 \cdots \gamma_{(2s)}^4 \psi \, d\Omega_p. \tag{4.38}$$

Denote this Hilbert space by $\mathscr{R}^{(s)}$.

Using (4.37) we obtain

$$\frac{1}{\lambda} \hat{B}_\mu^{(s)} = \frac{1}{2m} \{P^\rho, S_{\rho\mu}\} = -i\Gamma_\mu + i\left[\frac{(2s)}{(2m)}\right] P_\mu, \tag{4.39}$$

which is a relation valid on $\mathscr{R}^{(s)}$. Using (4.39) it can be proven that the following defines a principal series UIR of $\overline{\mathrm{SO}_0(4,1)}$ on $\mathscr{R}^{(s)}$ characterized by m and s (see I):

$$(\bar{U}(g)\psi)(p) = (1/|\mu(\bar{g}^{-1}, p/m)|^{3/2 - s + i\rho}) \, D(\bar{g}) \, \psi(m\bar{g}^{-1}p/m). \tag{4.40}$$

We have omitted the ξ indices. Here, $\bar{g}$ is defined by (3.17) and

$$\mu(\bar{g}^{-1}, p/m) = \bar{g}^{-1}{}_5^5 + \bar{g}^{-1}{}_\mu^5 \, p^\mu/m.$$

In contrast to the $\overline{E(4)}$ case we have the result that $\mathscr{R}^{(s)}$ is the direct sum of two UIR spaces of $\overline{\mathscr{P}}$ (see I)

$$\mathscr{R}^{(s)} = \mathscr{H}(m, s; +) \oplus \mathscr{H}(m, s; -). \tag{4.41}$$

The positive mass UIR's $\mathscr{H}(m,s; \pm)$ of $\overline{\mathscr{P}}$ are characterized by the mass squared

$$P_\mu P^\mu = m^2 I \tag{4.42}$$

and the square of the Pauli–Lublanski four-vector

$$W = m^2 s(s+1) I \tag{4.43}$$

plus the sign of the energy $p_0/|p_0|$. [P_μ and $L_{\mu\nu}$ are the Hermitian generators of $\overline{\mathscr{P}}$ and W is defined in terms of P_μ and $L_{\mu\nu}$ exactly in the same way as for $\overline{E(4)}$.] A UIR of $\overline{\mathscr{P}}$ on $\mathscr{R}^{(s)}$ is given, analogous to Eq. (4.15) for $\overline{E(4)}$, by

$$[U([a, A])\psi](p; \epsilon) = e^{ia_\mu p^\mu} D(A) \, \psi(A^{-1}p; \epsilon) \quad (\epsilon = p_0/|p_0|), \tag{4.44}$$

where $D(A)$ is defined by (4.36) for $A \in \overline{\mathrm{SO}_0(3,1)}$.

The description of these UIR's of $\overline{\mathscr{P}}$ on

$$\mathscr{H}(m,s;+)=\mathscr{L}^2(T_3^+)\otimes \mathbb{C}^{2s+1} \tag{4.45}$$

or

$$\mathscr{H}(m,s;-)=\mathscr{L}^2(T_3^-)\otimes \mathbb{C}^{2s+1} \tag{4.46}$$

consisting of all complex $(2s+1)$-vector valued functions $\psi(p,s_3;\epsilon)$ on T_3^ϵ which satisfy

$$(\psi,\psi)=\sum_{s_3=-s}^{s}\int_{T_3} d\Omega_p\,\bar{\psi}(p,s_3;\epsilon)\,\psi(p,s_3;\epsilon)<\infty \tag{4.47}$$

[for $\epsilon=+$ we have T_3^+ corresponding to $\mathscr{H}(m,s;+)$ and for $\epsilon=-$ we have T_3^- corresponding to $\mathscr{H}(m,s;-)$], is given by

$$[U_W([a,\Lambda])\psi](p,s_3;\epsilon)$$
$$=e^{ip_\mu a^\mu}\sum_{s_3'} D_{s_3}^{s_3'}(R(\Lambda,p))\,\psi(\Lambda^{-1}p,s_3';\epsilon), \tag{4.48}$$

where

$$R(\Lambda,p)=L_p\Lambda L(\Lambda^{-1}p)$$
$$(R(\Lambda,p)=R(\Lambda,\Lambda^{-1}p)\text{ of } I) \tag{4.49}$$

is the $\overline{\mathrm{SO}_0(3,1)}$ Wigner rotation $L^{-1}(p)$ is the Lorentz boost, and $D_{s_3}^{s_3'}$ is the $(2s+1)\times(2s+1)$ matrix representation of $\overline{\mathrm{SO}_0(3)}$. As in the $\overline{E(4)}$ case there exists an intertwining operator

$$F:\mathscr{H}(m,s;+)\oplus\mathscr{H}(m,s;-)\to\mathscr{R}^{(s)} \tag{4.50}$$

whose restriction to $\mathscr{H}(m,s;\epsilon)$ intertwines the two representations (4.44) and (4.48). Its explicit form is given in the first paper of Ref. 36 and also in Ref. 37 for $s=\frac{1}{2}$.

V. EQUIVALENCE OF THE REALIZATIONS OF THE PRINCIPAL SERIES UIR'S OF $\overline{\mathrm{SO}_0(4,1)}$ ON $\hat{\mathscr{R}}^{(s)}$ AND $\mathscr{R}^{(s)}$ WITH THEIR REALIZATIONS AS INDUCED REPRESENTATIONS

In this section we prove the equivalence of the realizations of the principal series UIR's of $\overline{\mathrm{SO}_0(4,1)}$ constructed in Sec. III with those described in Sec. IV. Since the proofs of the equivalence for the representations on function spaces over S_3 and T_3 are so similar, we treat only the case dealing with S_3 in detail.

First we must know the infinitesimal generators of the representation given by (4.13): the e.s.a. (essentially self-adjoint) infinitesimal generators, defined by the equation

$$\{-i\hat{L}^{ab}\hat{\psi}\}(u)=\frac{d}{d\omega}\{\hat{U}(e^{\omega I^{ab}})\hat{\psi}\}(u)|_{\omega=0} \tag{5.1}$$

can be calculated, using a coordinate patch. They are expressible in terms of $\hat{P}_\mu$, $\hat{M}_{\mu\nu}=\hat{X}_\mu\hat{P}_\nu-\hat{X}_\nu\hat{P}_\mu$, and $\hat{S}_{\mu\nu}$ as follows[38]:

$$\hat{L}_{s\mu}=-(1/\lambda)\hat{B}_\mu, \tag{5.2}$$

$$\hat{L}_{\mu\nu}=\hat{M}_{\mu\nu}+\hat{S}_{\mu\nu}, \tag{5.3}$$

where $\hat{X}_\mu$ is the operator i times differentiation by u_μ on $\mathbb{R}^4$ and

$$\frac{1}{\lambda}\hat{B}_\mu=\frac{\rho}{m}\hat{P}_\mu+\frac{1}{2m}\{\hat{P}^\rho,\hat{L}_{\rho\mu}\}=\frac{1}{\lambda}\mathring{\hat{B}}_\mu+\frac{1}{\lambda}\hat{B}_\mu^{(s)}, \tag{5.4}$$

λ is related to ρ and m through the equation

$$\rho=\sqrt{m^2/\lambda^2} \tag{5.5}$$

and

$$\frac{1}{\lambda}\mathring{\hat{B}}_\mu=\frac{\rho}{m}\hat{P}_\mu+\frac{1}{2M}\{\hat{P}^\rho,\hat{M}_{\rho\mu}\}, \tag{5.6}$$

$$\frac{1}{\lambda}\hat{B}_\mu^{(s)}=\frac{1}{2M}\{\hat{P}^\rho,\hat{S}_{\rho\mu}\}. \tag{5.7}$$

With the intertwining operator $\hat{F}$ [Eq. (4.31)] we can define a representation $\hat{U}_W$ of $\overline{\mathrm{SO}_0(4,1)}$ on $\hat{\mathscr{H}}(m,s)$ as follows:

$$g\to\hat{U}_W(g):[\hat{U}_W(g)\hat{\psi}_W](u)=[\hat{F}^{-1}\hat{U}(g)\hat{\psi}](u)$$
$$(\hat{\psi}_W=\hat{F}^{-1}\hat{\psi}). \tag{5.8}$$

Under $\hat{F}$ the operators $\hat{P}_\mu$ and $\hat{L}_{\mu\nu}$ become transformed into[39]

$$\hat{P}_{\mu_W}=\hat{F}^{-1}\hat{P}_\mu\hat{F}=\hat{P}_\mu, \tag{5.9}$$

$$\hat{L}_{ij_W}=\hat{F}^{-1}\hat{L}_{ij}\hat{F}=\hat{L}_{ij}, \tag{5.10}$$

$$\hat{L}_{4i_W}=\hat{F}^{-1}\hat{L}_{4i}\hat{F}=\hat{M}_{4i}+[1/(\hat{P}_4+m)]\,\hat{P}^k\hat{S}_{ki}. \tag{5.11}$$

Thus the transformed $(1/\lambda)\hat{B}_4$ becomes

$$\frac{1}{\lambda}\hat{B}_{4_W}=\frac{\rho}{m}\hat{P}_4+\frac{1}{2m}\{\hat{P}^i,\hat{M}_{i4}\}+\frac{1}{m}\frac{1}{(\hat{P}_4+m)}\hat{P}^i\hat{P}^k\hat{S}_{ki}$$
$$=\frac{1}{\lambda}\mathring{\hat{B}}_4. \tag{5.12}$$

And the transformed $(1/\lambda)\hat{B}_i$'s are

$$\frac{1}{\lambda}\hat{B}_{i_W}=\frac{1}{\lambda}\mathring{\hat{B}}_i+\frac{1}{(\hat{P}_4+m)}\hat{P}^k\hat{S}_{ki}. \tag{5.13}$$

Using these results we can easily prove the equivalence of the representations (4.13) on $\hat{\mathscr{R}}^{(s)}$ and (3.13) (for $c=i\rho$) on the Hilbert space $\hat{\mathscr{H}}$ of (3.23). First note that

$$\hat{\mathscr{H}}\cong\hat{\mathscr{H}}(m,s), \tag{5.14}$$

where $\hat{\mathscr{H}}(m,s)$ is given by (4.25). We then have the following.

Theorem: The representation $\hat{U}_W(g)$ of $\overline{\mathrm{SO}_0(4,1)}$, defined by (5.8) on $\hat{\mathscr{H}}(m,s)$ is equivalent to the representation (3.13) $(c=i\rho)$ on $\hat{\mathscr{H}}$.

Proof: For $g_{\mu\nu}=\Lambda\in\overline{\mathrm{SO}_0(4)}$ we have by (4.27) and the intertwining property of $\hat{F}$ [see Eq. (4.29)] together with the fact that (4.13) and (4.15) agree for $\overline{\mathrm{SO}_0(4)}$ transformations

$$\hat{U}_W(\Lambda)=\hat{U}_W([0,\hat{\Lambda}]). \tag{5.15}$$

Thus

$$[\hat{U}_W(\Lambda)\hat{\psi}_W](u)=D(R(\hat{\Lambda},u))\,\psi_W(\hat{\Lambda}^{-1}u). \tag{5.16}$$

Next we need to know the form of the one-parameter group of transformations $\hat{U}_W(e^{\omega I^{45}})$ generated by $(1/\lambda)\hat{B}_{4W}$ which is defined in (5.12). The result is well known,[40] it is

$$[\hat{U}_W(e^{\omega I^{45}})\hat{\psi}_W](u)$$
$$=\frac{1}{|\mu(\hat{g}_{45}^{-1},u/m)|^{3/2-i\rho}}\hat{\psi}_W\left(m\hat{g}_{45}^{-1}\frac{u}{m}\right). \tag{5.17}$$

It is seen to be identical with (3.13), since from Appendix A, $D^s(m(\hat{g}_{45},u))=1$. Also, by comparing (A12) and (B16) we can conclude that (5.16) is identical to (3.13) for Lorentz transformations. So $\hat{U}_W$ and (3.13) agree for the one-parameter subgroups of $\overline{SO_0(4)}$ transformations and rotations in the 45 plane of $M_{4,1}$. But an arbitrary element of $\overline{SO_0(4,1)}$ has a decomposition as[41]

$$g = g_{12}\, g_{23}\, g_{34}\, g_{45}\, g_{12}\, g_{23}\, g_{34}\, g_{12}\, g_{23}\, g_{12} \tag{5.18}$$

(g_{ab} corresponds to a one-parameter group of rotations in the a-b plane of $M_{4,1}$). Therefore, by the homomorphism properties of the representations (3.13) and $\hat{U}_W$, we conclude that they are equivalent representations. [The equivalency is given by the isomorphism in Eq. (5.14).]

From the theorem follows the equivalence of the realizations of the principal series given by (3.13) and (4.13). The intertwining operator has been explicitly constructed and is given by (4.29). Another proof of the equivalence can be obtained by a comparison of the "spin parts" of Eqs. (5.10), (5.11), (5.12), and (5.13) with the Eqs. (A12), (A13), (A14), and (A15) of Appendix A. A more interesting global proof of the equivalence is also possible by calculating the explicit form of $\hat{U}_W$ from (4.13) without resorting to infinitesimal arguments, i.e., using the $\overline{SO_0(4)}$ analog of the explicit form for $\hat{F}$ provided in Ref. 36.

In order to prove that (4.40) defines a representation equivalent to that defined by (3.21) we proceed analogously as for the just-treated S_3 case. We use the Poincaré group "Foldy–Wouthuysen" transformation, as defined in Ref. 36, to construct a representation $\bar{U}_W(\overline{SO_0(4,1)})$ on $\mathscr{H}(m,s;+) \oplus \mathscr{H}(m,s;-)$ [Eq. (4.41)] equivalent to one given by (4.40) on $\mathscr{R}^{(s)}$. We verify that the representations $\bar{U}(g)$ of (3.21) and $\bar{U}_W(g)$ are equivalent for $g \in \overline{SO_0(3,1)}$. Then using the results of Ref. 36 we verify that the spin parts of $\bar{U}_W(g_{\mu 5})$ and $\bar{U}(g_{\mu 5})$ of Eq. (3.21) are the same.

Now it is clear how one obtains forms analogous to (4.13) for the complementary series. Using the intertwining operator $\hat{F}$ [Eq. (4.31)] we can unwind the Wigner rotation to obtain another representation on a space, $\hat{\mathscr{R}}_c^{(s)}$. On $\hat{\mathscr{R}}_c^{(s)}$ the representation has a form given by Eq. (4.13), in which ip is replaced by a real number c which satisfies (3.30). Using the explicit form of $\hat{F}$ (Ref. 36 or Ref. 37) we can calculate the transformed norm on $\hat{\mathscr{R}}_c^{(s)}$ from the old one [Eq. (3.27)] and obtain a characterization of $\hat{\mathscr{R}}_c^{(s)}$ in terms of totally symmetric normalizable tensors which satisfy certain constraint relations, as was the case for the principal series. The discrete series is a little more complicated, but one should also be able to obtain a form analogous to (4.13) for this case.

In concluding this section we shall briefly consider the form of the representation given by (3.5). In order to construct the analog of (4.13) or (4.40) for this case we must use the Lévy-Leblond equation for the Galilei group,[42] and again construct Bargmann–Wigner equations of the Galilei group out of this equation.[42] However, the manifold which must be used is the paraboloid P_3 in $\mathbb{R}^4$ defined by

$$P_3 = \{x^\mu \in \mathbb{R}^4 |\, x^4 = \tfrac{1}{2}(x_1^2 + x_2^2 + x_3^2) - 1\}. \tag{5.19}$$

(See Fig. 1.) Under the mapping (3.7) x^4 is given by

$$x_4 = \tfrac{1}{2}((u^4+1)/(-u^4+1)) - 1. \tag{5.20}$$

We can view (3.5) as a representation on function spaces over P_3 (see Ref. 43). We note that the representations (3.5), (3.13), and (3.21) are all equivalent to representations on a space of homogeneous functions on projective P_3, S_3, or T_3, i.e., $\mathbf{K}_+^4$.[44] The corresponding representations on homogeneous functions on $\mathbf{K}_+^4$ are all related to one another by certain $\overline{SO_0(4,1)}$ transformations of $M^{4,1}$.[45] Furthermore, through the use of a generalization (to arbitrary spins) of the Gelfand–Graev transform for $\overline{SO_0(4,1)}$,[46] we obtain equivalent representations, which occur as irreducible representations in the decomposition of the left regular representation of $G = \overline{SO_0(4,1)}$ on $\mathscr{L}^2(\bar{G}/K)$, where $\bar{G}/K$ is the Cartesian product of de Sitter or anti-de Sitter space with some discrete space (for example, $\mathbf{Z}_N$) (at least locally).

VI. CONCLUSIONS

The generalizations of the methods presented here to the $\overline{SO_0(p,q)}$ groups is straightforward. We must introduce a spin structure appropriate to $M^{p,q}$'s and consider Bargmann–Wigner equations on various higher-dimensional analogs of S_3 and T_3. The remarks at the end of Sec. V provide a slight generalization of the well-known harmonic analysis on $\bar{G}/K$, which probably can be extended with appropriate modification to $\bar{G} = \overline{SO_0(p,q)}$ (see Strichartz, Ref. 46). A more interesting problem is the harmonic analysis on $\bar{G}/H$, in which H is an arbitrary subgroup of $\bar{G}$ such that $H \subset K$ and $\dim(H) < \dim(K)$. It may be possible to utilize the methods developed here, in order to solve this problem for certain H's. Another interesting mathematical question is whether or not the method presented here can be generalized to the semisimple Lie groups which have semidirect products as group contractions.[24] The results of Ref. 24 seem to indicate that this is possible. Finally we mention that our results provide independent proofs of some of the results in Ref. 3 on equivalences and forms of the representations associated with the models (2.3′), (2.4′), and (2.5′). In particular, the result that the Wigner form [Eq. (4.48)] of the Poincaré group representation is identical to the induced form [Eq. (3.21)] of the $\overline{SO_0(4,1)}$ representation, when restricted to the $\overline{SO_0(3,1)}$ subgroup, is also obtained in Ref. 3.

Note added in proof: Bakri has also considered equations such as (4.12) and and (4.39) and has proven their equivalence with the Bargmann–Wigner equations (4.11) for SL(2,$\mathbb{C}$) for the special $\overline{SO_0(3,2)}_{\Gamma_\mu, S_{\mu\nu}}$ representations considered here. See M. M. Bakri, Nuovo Cimento A **51**, 864 (1967); M. M. Bakri, J. Math. Phys. **10**, 298 (1969).

ACKNOWLEDGMENTS

I express my sincere gratitude to Professor John Gilbert for numerous enlightening discussions which led to the solution of the problem presented here. I wish to thank Professor W. Drechsler for discussions, which were useful for the formulation of Appendix B, and the Max Planck Institut for hospitality. The work described in this paper was done during the tenure of a fellowship from Deutscher Akademischer Austauschdienst (DAAD).

APPENDIX A: CALCULATION OF $D^s(m(\hat{g},u))$ FOR $\hat{g}$ INFINITESIMAL

In this appendix we collect a number of useful properties about the matrix $D^s(m(\hat{g},u))$ defined by (3.13). From (3.6) we have

$$m(g,x)=\tilde n_{x(u)}^{-1}\, g\tilde n_{g\cdot x(u)}\, n^{-1}(g,x)\, a^{-1}(g,x), \tag{A1}$$

with $x(u)$ defined by (3.7). Since $m(g,x)\in \overline{M}=\mathrm{SU}(2)$, it can be expressed in the form

$$m(g,x)=\exp(\tfrac14 i\omega^{ij}(g,x)\,\hat\sigma_{ij}), \tag{A2}$$

with

$$\hat\sigma_{ij}=\epsilon_{ijk}\sigma_k, \tag{A3}$$

where σ_i are the 2×2 Pauli matrices satisfying

$$\sigma_i\,\sigma_j=i\epsilon_{ijk}\,\sigma_k+\delta_{ij}. \tag{A4}$$

$D^s(m(g,x))$ is then obtained through the homomorphism of Lie algebras defined by

$$\tfrac12\sigma_i\to\hat S_i=\tfrac12\epsilon_{ijk}\,\hat S_{jk}, \tag{A5}$$

where $\hat S_i$ are $(2s+1)\times(2s+1)$ Hermitian matrices satisfying the commutation relations of SU(2)

$$[\hat S_i,\hat S_j]=i\epsilon_{ijk}\,\hat S_k. \tag{A6}$$

The matrices $m(g,x)$ for the various continuous subgroups of the decomposition (2.4) can be computed and are found to be[47] translations,

$$m(\tilde n_a,x)=\mathbb{1},\quad \tilde n_a=e^{a^i(\hat I_{i4}+\hat I_{i5})}\in\tilde N$$
$$[(\hat I_{ab})^c_d=n_{ad}\,\delta^c_b-n_{bd}\,\delta^c_a]; \tag{A7}$$

SU(2) rotations,

$$m(m,x)=e^{(i/4)\,\epsilon_{ijk}\omega_{ij}\,\sigma_k},\quad m=e^{(1/2)\,\omega^{ij}\hat I_{ij}}\in\mathrm{SU}(2), \tag{A8}$$

dilations,

$$m(a,x)=\mathbb{1},\quad a=e^{\omega\hat I_{54}}; \tag{A9}$$

special conformal transformations,

$$m(n_b,x)=\frac{(1-(\mathbf{x}_u\cdot\boldsymbol\sigma)(\mathbf{b}\cdot\boldsymbol\sigma))}{(1-2\mathbf{x}\cdot\mathbf{b}+x^2b^2)^{1/2}},$$
$$n_b=e^{b^i(\hat I_{i4}-\hat I_{i5})}. \tag{A10}$$

We now use these expressions to calculate the $D^s(m(\hat g,u))$ for infinitesimal rotations in the a,b planes of $M^{4,1}$:

$$g_{ab}\cong I+\omega\hat I_{ab}. \tag{A11}$$

From (A7) and (A10) together with the cocycle conditions (2.12′) of $D^s(m(g,u))$, we obtain, with the help of (A3) and (A5),

$$D^s(m(\hat g_{i4},u))=D^s(m(g_{i4},x))$$
$$=\mathbb{1}-i[b^iu^j/(1+u_4)]\,\hat S_{ij}\quad(b^i=\omega^{i4}), \tag{A12}$$

$$D^s(m(\hat g_{i5},u))=\mathbb{1}+i[\tilde b^iu^j/(1+u_4)]\,\hat S_{ij}$$
$$(\tilde b^i=\omega^{i5}), \tag{A13}$$
$$[x_i=u_i/(1+u_4)\ \text{by}\ (3.7)].$$

For an infinitesimal rotation in the i,j plane of $M^{4,1}$ we obtain using (A8):

$$D^s(m(\hat g_{ij},u))=\mathbb{1}+(i/2)\,\omega^{ij}\hat S_{ij}. \tag{A14}$$

Finally from (A9) for an infinitesimal dilation we have

$$D^s(m(\hat g_{45},u))=\mathbb{1}, \tag{A15}$$

and so also for a finite dilation.

APPENDIX B: SOME PROPERTIES OF $SO_0(4)$

The group $SO_0(4)$ has a decomposition (Cartan decomposition) as the product of $SO_0(3)$ and a compact analog of the Lorentz boosts

$$SO_0(4)=\hat{\mathscr{L}}SO(3). \tag{B1}$$

Therefore, for each $g\in SO_0(4)$ we have

$$\hat g=\hat L_u^{-1}R, \tag{B2}$$

with $R\in SO(3)$; and $\hat L_u^{-1}$ is the boost matrix

$$(\hat L_u^{-1})^\mu_\nu=(\hat L^{-1}(u))^\mu_\nu$$
$$=\begin{pmatrix}\delta^i_j-u^iu_j/(1+u_4) & u^i\\ -u_j & u_4\end{pmatrix}, \tag{B3}$$

where $\Sigma_{i=1}^3 u^iu^i+u_4^2=1$. Note that $\hat{\mathscr{L}}$ can be identified as a manifold with the points on the three sphere S_3.

The decomposition (B1) defines a transitive action of $SO_0(4)$ on the manifold of the boosts S_3[48]:

$$S_3=SO_0(4)/SO_0(3). \tag{B4}$$

The isotropy subgroup of the point $\mathring{u}=(0,0,0,1)$ is $SO_0(3)$, and it consists of all matrices of the form

$$R^\mu_\nu=\begin{pmatrix}R^i_j & 0\\ 0 & 1\end{pmatrix}, \tag{B5}$$

where R_{ij} is the matrix of a 3×3 rotation. Notice that the boost $\hat L_u^{-1}$ takes $\mathring{u}$ into $u\in S_3$; consequently the action of $SO_0(4)$ on S_3 is seen to be transitive.

There exists another extremely useful decomposition of $SO_0(4)$. It is the Wigner decomposition[49]: For $\hat\Lambda\in SO_0(4)$, we have

$$\hat\Lambda=\hat L_{\hat\Lambda u}^{-1}R'(\hat\Lambda,u)\,\hat L_u, \tag{B6}$$

where $R'(\hat\Lambda,u)$ is a pure rotation, and $\hat L_u$ and $\hat L_{\hat\Lambda u}^{-1}$ are defined by Eq. (B3). To see that $R'(\hat\Lambda,u)$ is a pure rotation we consider the action of $R'(\hat\Lambda,u)$ on $\mathring{u}$:

$$R'(\hat\Lambda,u)\,\mathring{u}=\hat L_{\hat\Lambda u}\hat\Lambda\hat L_u^{-1}\,\mathring{u}=\hat L_{\hat\Lambda u}(\hat\Lambda u)=\mathring{u}.$$

Therefore, $R'(\hat\Lambda,u)$ is an element of the isotropy subgroup of $\mathring{u}$, i.e., $R'(\hat\Lambda,u)\in SO_0(3)$. We can easily verify from its definition that this Wigner rotation satisfies cocycle conditions.

For $\hat\Lambda=\hat L_{\bar u}^{-1}$ a pure boost the Wigner rotation becomes expressible as the product of the three boosts

$$R'(\hat L_{\bar u}^{-1},u)=\hat\Lambda_{\hat L_{\bar u}^{-1}u}\,\hat L_{\bar u}^{-1}\,\hat L_u^{-1}. \tag{B7}$$

Next we consider the product of two $SO_0(4)$ transformations $\hat\Lambda_1=\hat L_{u_1}^{-1}R$, and $\hat\Lambda_2=\hat L_{u_2}^{-1}R_2$. We have for $\hat\Lambda_1=R$ and $\hat\Lambda_2=\hat L_u^{-1}$ the following result:

$$\hat\Lambda_1\hat\Lambda_2=R\hat L_u^{-1}=\hat L_{Ru}^{-1}R. \tag{B8}$$

[Use (B3) and (B5).] Thus for $\hat\Lambda_1=R_1$ and $\hat\Lambda_2=\hat L_u^{-1}R_2$ we obtain

$$\hat\Lambda_1\hat\Lambda_2=R_1\hat L_u^{-1}R_2=\hat L_{R_1u}^{-1}R_1\,R_2=\hat L_{R_1u}^{-1}(R_1\,R_2). \tag{B9}$$

Finally for $\hat\Lambda_1=\hat L_{u_1}^{-1}R_1$ and $\hat\Lambda_2=\hat L_{u_2}^{-1}R_2$ we obtain

$$\hat{A}_1\hat{A}_2=\hat{L}_{u_1}^{-1}R_1\,\hat{L}_{u_2}^{-1}R_2=\hat{L}_{u_1}^{-1}\,\hat{L}_{R_1u_2}^{-1}(R_1\,R_2). \tag{B10}$$

Letting $\hat{L}_{u}^{-1}=\hat{L}_{(R_1u_2)}^{-1}$ and $\hat{L}_{\bar{u}}^{-1}=\hat{L}_{u_1}^{-1}$ in (B7) we can rewrite (B10) as

$$\hat{A}_1\hat{A}_2=\hat{L}^{-1}_{\hat{L}_{u_1}^{-1}R_1u_2}\,R'(\hat{L}_{u_1}^{-1},R_1u_2)(R_1\,R_2). \tag{B11}$$

This is the general formula for the product of two $SO_0(4)$ transformations in terms of a boost and rotation.

Specializing to the case of $\hat{A}_1=R$ a pure rotation and $\hat{A}_2=\hat{L}_{u}^{-1}$ a pure boost we obtain the following expression for the Wigner rotation corresponding to a rotation:

$$R'(R,u)=\hat{L}_{Ru}R\hat{L}_{u}^{-1}=R'(I,Ru)\,R. \tag{B12}$$

But since $R'(I,Ru)=\hat{L}_{Ru}I\hat{L}_{Ru}^{-1}=I$ [compare Eq. (2.11′)]

$$R'(R,u)=R. \tag{B13}$$

Thus the Wigner rotation of a rotation is the rotation itself.

Using Eqs. (B7) and (B13), we compute the $(2s+1)\times(2s+1)$ matrices of the Wigner rotation $D(R'(\hat{A},u))$ for infinitesimal rotation

$$\hat{A}_{\mu\nu}=I+\omega I_{\mu\nu} \tag{B14}$$

in the μ,ν planes of $\mathbb{R}^4$. For $\hat{A}_{ij}=R_{ij}$ a rotation in the i,j plane, we obtain from (B13)

$$D(R(R,u))=D(R'(R,Ru^{-1}))=1+i\omega\hat{S}_{ij}, \tag{B15}$$

where $\hat{S}_{ij}$ is defined in Eq. (A5). For $\hat{A}_{i4}=\hat{L}_{\tilde{u}_i}^{-1}$ an infinitesimal boost in the i-4 plane we obtain, using (B7),

$$D(R(\hat{L}_{\tilde{u}_i},u))=D(R'(\hat{L}_{\tilde{u}_i},L_{\tilde{u}_i}^{-1}u))=1-i\frac{\tilde{u}^iu^j}{1+u_4}\hat{S}_{ij}\quad(\omega^i=\tilde{u}^i). \tag{B16}$$

An easy way to prove (B16) is to recognize that it is just the $(2s+1)\times(2s+1)$ matrix representation of the Thomas rotation for the product of the two boosts $L_{\tilde{u}_i}^{-1}L_{u_i}^{-1}$.[50]

APPENDIX C: USEFUL FACTS ABOUT THE BRUHAT AND IWASAWA DECOMPOSITIONS OF G

In this appendix we derive certain facts necessary for the establishment of the equivalence of (3.5) with (3.13). First we collect the necessary information from Ref. 4 needed to prove that the Bruhat decomposition induces the usual conformal transformations on $\mathbb{R}^3$:

$$y^i\xrightarrow{g}y^{i'}=\frac{g^i_jy^j+g^i_4\,y^2+g^i_5}{g^5_j\,y^j+g^5_4\,y^2+g^5_5}\quad(y=\sqrt{2x}). \tag{C1}$$

To verify this we show it is true for translation, dilations, rotations, and conformal inversions. We have, according to (3.3), (1) translations (Ref. 4, p. 17),

$$\tilde{n}_a\,\tilde{n}_x=\tilde{n}_{x'},\qquad x'=x+a,$$

$$\hat{\tilde{n}}_a=\begin{pmatrix}1 & -a^T & a^T\\ a & 1-\tfrac{1}{2}a^2 & \tfrac{1}{2}a^2\\ a & -\tfrac{1}{2}a^2 & 1+\tfrac{1}{2}a^2\end{pmatrix}, \tag{C2}$$

(2) dilations (Ref. 4, p. 18),

$$a\tilde{n}_x=\tilde{n}_{x'}h(a,x),\quad (h=man),\quad x'=|a|\,x,\qquad \hat{a}=\begin{pmatrix}1 & 0 & 0\\ 0 & \cosh\alpha & \sinh\alpha\\ 0 & \sinh\alpha & \cosh\alpha\end{pmatrix}, \tag{C3}$$

(3) rotations (Ref. 4, p. 18),

$$m\tilde{n}_x=\tilde{n}_{x'}h(m,x),\quad x'=m_{ij}\,x_j,\qquad \hat{m}=\begin{pmatrix}m_{ij} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{C4}$$

(4) conformal inversion (Ref. 4, p. 23),

$$R\tilde{n}_x=\tilde{n}_{x'}h(R,x),\quad x'=-x/x^2,\qquad \hat{R}=\begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix}. \tag{C5}$$

Next using (3.7) and (3.10) we verify that the action g on $\mathbb{R}^3$ and the action $\hat{g}$ on S_3 are connected by the similarity transformation

$$\hat{Q}_0=\begin{pmatrix} & 0 & 0\\ 0 & 1/\sqrt{2} & 1/\sqrt{2}\\ 0 & -1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}. \tag{C6}$$

Let

$$\hat{g}=\begin{pmatrix}\hat{g}_{ij} & \hat{g}_{i4} & \hat{g}_{i5}\\ \hat{g}_{4j} & \hat{g}_{44} & \hat{g}_{45}\\ \hat{g}_{5j} & \hat{g}_{54} & \hat{g}_{55}\end{pmatrix}. \tag{C7}$$

Then

$$g=\hat{Q}_0\hat{g}\,\hat{Q}_0^{-1}=\begin{pmatrix}\hat{g}_{ij} & (1/\sqrt{2})(\hat{g}_{i4}+\hat{g}_{i5}) & (1/\sqrt{2})(-\hat{g}_{i4}+\hat{g}_{i5})\\ (1/\sqrt{2})(\hat{g}_{4j}+\hat{g}_{5j}) & \tfrac{1}{2}(\hat{g}_{44}+\hat{g}_{54}+\hat{g}_{45}+\hat{g}_{55}) & \tfrac{1}{2}(-\hat{g}_{44}-\hat{g}_{54}+\hat{g}_{45}+\hat{g}_{55})\\ (1/\sqrt{2})(-\hat{g}_{4j}+\hat{g}_{5j}) & \tfrac{1}{2}(-\hat{g}_{44}+\hat{g}_{54}-\hat{g}_{45}+\hat{g}_{55}) & \tfrac{1}{2}(\hat{g}_{44}-\hat{g}_{54}-\hat{g}_{45}+\hat{g}_{55})\end{pmatrix}. \tag{C8}$$

For translation $\hat{g}$ is given by (C2) so

$$g=\begin{pmatrix}1 & 0 & \sqrt{2}a^T\\ \sqrt{2}a & 1 & 2a^2\\ 0 & 0 & 1\end{pmatrix}. \tag{C9}$$

Use of this in (C1) gives

$$x'=x+a.$$

Similarly we establish the desired transformation law for rotations, dilations, and conformal inversion. Thus we have verified the claim about (C1) and the Bruhat decomposition.

It is straightforward to verify

$$\hat{r}\circ g=\hat{g}\circ\hat{r}, \tag{C10}$$

where $\hat{r}$ is the stereographic projection mapping given by (3.7). Using this we demonstrate the cocycle property (2.12′)

of $D^s(m(\hat{g}, u))$ as a consequence of the cocycle property of the $D^s(m(g, x))$. Since (3.5) is an induced representation of $\overline{\mathrm{SO}_0(4,1)}$ we have

$$D^s(m(g_1 g_2, x)) = D^s(m(g_1, x))\, D^s(m(g_2, g_1^{-1}x)), \tag{C11}$$

for $g_1, g_2 \in \overline{G}, x \in \mathbb{R}^3$. By definition we have

$$D^s(m(\widehat{g_1 g_2}, u)) = D^s(m(g_1 g_2, x)), \tag{C12}$$

$$D^s(m(\hat{g}_1, u)) = D^s(m(g_1, x)). \tag{C13}$$

From (C10) it follows that

$$\tau(g_1^{-1}x) = \hat{g}_1^{-1}u. \tag{C14}$$

Therefore,

$$D^s(m(\hat{g}_2, \hat{g}_1^{-1}u)) = D^s(m(g_2, g_1^{-1}x)). \tag{C15}$$

Substituting (C12), (C13), and (C15) into (C11) gives the desired cocycle property.

[1]P. Moylan, J. Math. Phys. **24**, 2706 (1983).

[2]V. Bargmann, Ann. Math. **48**, 568 (1947).

[3]K. C. Hannabuss, Proc Cambridge Philos. Soc. **70**, 283 (1971).

[4]V. K. Dobrev, G. Mack, V. B. Petkova, S. G. Petrova, and I. T. Todorov, *Harmonic Analysis on the n-Dimensional Lorentz Group and Its Application to Conformal Quantum Field Theory* (Springer, New York, 1977).

[5]Reference 4, p. 14.

[6]C. Chevalley, *Theory of Lie Group* (Princeton U.P., Princeton, NJ, 1946), p. 67, Proposition 3.

[7](2.16), together with (2.14), and (2.10), imply $\nu = ((-n+1)/2) + i\rho$.

[8]C. P. Boyer, J. Math. Phys. **14**, 612 (1973), Theorem 1. He shows that the spin-zero complementary series representations are obtainable by analytic continuation in $\sigma = i\rho$ of the spin-zero principal series representations. It is possible to extend his results to the arbitrary spin case. [His $\mu(g, g^{-1}u)$ equals our $\mu(g^{-1}, u)$ in Sec. III.] (Cf. also Ref. 3.)

[9]F. Schwartz, J. Math. Phys. **12**, 13 (1971).

[10]L. H. Thomas, Ann. Math. **42**, 113 (1941); T. D. Newton, Ann. Math. **51**, 730 (1950); J. Dixmier, Bull. Soc. Math. France **89**, 9 (1961); A. Böhm, in *Studies in Mathematical Physics*, edited by A. O. Barut (Reidel, New York, 1973), cf. also Ref. 4, p. 135 for a summary of their $\mathrm{SO}_0(n,1)$ representations.

[11]Reference 4, p. 32.

[12]Reference 4, p. 33.

[13]Reference 4, p. 38, Eq. (2.27).

[14]Reference 4, p. 16.

[15]Reference 4, p. 56, Eq. (3.27).

[16]Reference 4, p. 38.

[17]P. Moylan, Fortschr. Phys. **28**, 273, (1980), Eq. (8a) with $p_0 = \frac{1}{2}$. [The S_3 measure used there equals $\frac{1}{8}$ times the measure on S_3 in Eq. (3.8).]

[18]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 and 346 (1966).

[19]G. Grensing, J. Math. Phys. **16**, 312 (1975). [See also V. K. Dobrev and V. B. Petkova, Rep. Math. Phys. **13**, 233 (1978). They analyze the $\overline{\mathrm{SO}_0(5,1)}$ case but the treatment of $\mathrm{SO}_0\,\overline{(4,1)}$ is analogous.]

[20]Reference 19; p. 315, Eq. (5.11) or Ref. 4, p. 66, Eq. (4.19).

[21]A. W. Knapp and E. M. Stein, Ann. Math. (2) 93, 489 (1971), Propositions 50 and 55; Raymond C. Fabec, J. Analyse Math. **35**, 150 (1979).

[22]Reference 4, they treat only the integer spin case.

[23]N. J. Villenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968), Chap. X.

[24]G. W. Mackey, in *Lie Groups and their Representations*, Summer School of the Bolyai Janos Mathematical Society, Proceedings, Budapest, 1971, edited by I. M. Gel'fand (Wiley, London, 1975), p. 361.

[25]Reference 1.

[26]K. B. Wolf, Nucl. Phys. B **11**, 159 (1969).

[27]For the operations with tensor products of operators needed here see Chap. X of the book by A Böhm, *Quantum Mechanics* (Springer, New York, 1979). For a more mathematical treatment, but still on an elementary level, see M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972).

[28]Reference 1, p. 2710.

[29]$\hat{A}$ is related to A through a correspondence of the spin groups defined by the Gl(5,$\mathbb{R}$) matrix $\hat{Q}_0$ of $M^{5,1}$ induced by stereographic projection (see Ref. 18, p. 355 and Appendix of Ref. 1). In these references the $\hat{Q}_0$ for projection of S_3 onto T_3 is discussed [Eqs. (3.14) and (3.15)]. (See also Appendix C.)

[30]Reference 1.

[31]See Ref. 29 for the correspondence between g and $D(\hat{g})$.

[32]Compare V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. USA **34**, (5), 211 (1946), and H. Joos, Fortschr. Phys. **10**, 65 (1962) for the case of the Poincaré group. See the Appendix of Ref. 1 for the case at hand, i.e., for $\overline{E(4)}$.

[33]A. Böhm and G. B. Mainland, Nuovo Cimento A **18**, 308 (1973).

[34]See Ref. 1, p. 2710, Eqs. (45), (46), and (47) and the discussion dealing with these equations. See also H. Joos, Ref. 32. The SL(2,$\mathbb{C}$) case is treated in these works but the reader can easily convince himself of the validity of their arguments for $\overline{\mathrm{SO}_0(4)}$. Our $|f^c_{s_3}(u)\rangle$'s correspond, up to an orthogonal transformation of the $|j_3 j;(\pm j_0 c)\rangle$ to the $\overline{\mathrm{SO}_0(4)}$ analogs of the

$$\left(\begin{matrix} |\,pA;\epsilon\} \\ |\,p\dot{B};\epsilon\} \end{matrix}\right)$$

consider in these papers for $s = \frac{1}{2}$. (For s arbitrary see Ref. 33.)

[35]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1960).

[36]A. Chakrabarti, J. Math. Phys. **4**, 1216 (1963), and Ref. 1, p. 2710. [Again the analogous SL(2,$\mathbb{C}$) case is treated.]

[37]A. Böhm and G. B. Mainland, Fortschr. Phys. **18**, 290 (1970).

[38]Compare Ref. 1 and see also A. Böhm and P. Moylan, "On the integrability of certain symmetric representations of the Lie algebra of $\overline{\mathrm{SO}_0(4,1)}$," J. Math. Phys. (to be published, 1985), for the corresponding representation on UIR space of $\mathscr{P}$. In the paper by Böhm and Moylan a very careful analysis of the Lie algebra representation associated with an induced multiplier representation realized on UIR spaces of $\mathscr{P}$ is presented.

[39] $\overline{\mathrm{SO}_0(4)}$ analog of Ref. 36, Eqs. (1.4) and (1.9).

[40]Reference 1, p. 2209 or Ref. 18.

[41]Reference 23, p. 508.

[42]J. M. Lévy-Leblond, Commun. Math. Phys. **6**, 286 (1967).

[43]P. Moylan, MPI preprint, October, 1984.

[44]Reference 23, Chap. X and Ref. 18 ($\mathbb{K}^4_+$ is the four-dimensional upper half cone.)

[45]Reference 18, p. 356.

[46]Reference 23, p. 542 and R. Strichartz, J. Funct. Anal. **12**, 4 and 341 (1973).

[47]Specialize the results of Ref. 19 to the $\overline{\mathrm{SO}_0(4,1)}$ subgroup of $\overline{\mathrm{SO}_0(5,1)}$.

[48]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), Chap. X.

[49]For the analogous treatment of SL(2,$\mathbb{C}$) see E. P. Wigner, *Istanbul Lectures*, edited by F. Gürsey (Gordon and Breach, New York, 1964).

[50]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), p. 367. In his notation $(1/\gamma - 1)(1/v^2) = (1/(1+u_4)$ so (B16) follows immediately from his equation (11.53). See also W. Drechsler, "Linearly and Nonlinearly Transforming Fields on Homogeneous Spaces of the (4 + 1) de Sitter Group, Appendix A," Max Planck Institut preprint, MPI-PAE/PTh 17/84, (1984, and L. Fonda and G. Ghirardi, *Symmetry Principles in Quantum Mechanics* (Dekker, New York, 1970), p. 300, Eq. (5.111a).

Linearly and nonlinearly transforming fields on homogeneous spaces of the (4,1)-de Sitter group

W. Drechsler
Max-Planck-Institut für Physik und Astrophysik, Werner-Heisenberg-Institut für Physik, Munich, Federal Republic of Germany

(Received 29 March 1984; accepted for publication 28 June 1984)

Scalar functions on the homogeneous spaces $\mathscr{H}^R$ of the de Sitter group $G = \mathrm{SO}(4,1)$ are studied, where the spaces $\mathscr{H}^R$ are of the form G/K with K being a subgroup of the Lorentz group $H = \mathrm{SO}(3,1)$ contained in SO(4,1). The spaces $\mathscr{H}^R$ can be regarded as fiber bundles $E^R = E^R(G/H, H/K)$, with the base $V'_4 = G/H$ being a space of constant negative curvature characterized by a fundamental length parameter R [(4,1)-de Sitter space], and the fiber $S = H/K$ being a homogeneous space of the Lorentz group. The action of G on the spaces E^R is a linear action on V'_4 and a nonlinear action on S, where the latter action is defined by a generalized Wigner rotation. A gauge theory based on the (4,1)-de Sitter group is investigated with matter represented in terms of a generalized wave function $\Phi(x;\xi,\tilde{y})$ [with $x \in U_4$ (Riemann–Cartan space-time), $\xi \in V'_4$, and $\tilde{y} \in S$] which is defined as a map from a cross section on the bundle $E = E(U_4, F = E^R, G = \mathrm{SO}(4,1))$ over space-time U_4 with fiber $F = E^R = G/K$ and structural group $G = \mathrm{SO}(4,1)$ into the complex numbers. The introduction of purely nonlinearly transforming fields ${}_{(N)}\Phi(x;\tilde{y})$ is discussed as well as the nonlinear realization of the SO(4,1) symmetry in terms of transformations of the Lorentz subgroup H (generalized Wigner rotations). The geometric implications of symmetry breaking are pointed out.

I. INTRODUCTION

Several years ago it was suggested to represent matter in particle theory in terms of fields defined on the homogeneous space of a group. Since the Poincaré group plays a prominent role in particle physics, the original proposal was to use a homogeneous space of the Poincaré group[1] and introduce a generalized wave function[2] as a scalar-valued function defined on a homogeneous space of the Poincaré group replacing thereby the ordinary quantum mechanical wave function defined over Minkowski space-time by a more general object.[3,4] The idea in this connection was to allow the spin of the material objects described in terms of these generalized wave functions to play a dynamical role in the theory. Instead of introducing multicomponent fields defined over Minkowski space-time transforming as a particular linear representation of the Poincaré group characterized by a fixed number of components, one enlarges the space over which the wave functions are to be defined and considers scalar functions on a homogeneous space of the Poincaré group. (In Ref. 1 the use of the ten-parameter group space of the Poincaré group was proposed; in Ref. 3 lower-dimensional homogeneous spaces of the Poincaré group capable of carrying half-integer spin representations were suggested. Compare also Ref. 5 in this context.) A field associated with a particular fixed value for mass and spin is obtained by posing additional constraint relations, i.e., by requiring that the Casimir operators of the Poincaré group take particular fixed values when applied to the generalized wave function.

It is interesting to combine such a generalized wave function formalism for objects possessing internal degrees of freedom associated with some group G (originally assumed to be the Poincaré group, as mentioned) with the gauge concept for interactions in particle physics and introduce as a representation of matter in the theory a scalar field defined on a (soldered) fiber bundle over space-time with a homogeneous space of the group G as fiber. In previous investigations based on the Poincaré group as a gauge group, a set of bilinear source terms in the generalized matter fields was constructed (generalized current components) which were shown to induce torsion in the underlying space-time base of the bundle geometry yielding thereby a description of a Poincaré gauge theory defined over a Riemann–Cartan base U_4.[6,7] As in general relativity, the metric of space-time is induced in this theory by the energy-momentum tensor of macroscopic matter distributed in classical form, while the nonmetric, i.e., torsion part of the connection,[8] is induced in the base of the underlying bundle geometry through the above-mentioned generalized matter current representing the microscopical, i.e., quantum mechanical or wave function aspect of matter (albeit in a single-particle formulation). The matter fields in the theory giving rise to the generalized matter current were called Poincaré gauge fields being scalar fields transforming as linear[9] representations of the Poincaré group. Mathematically speaking they are cross sections on soldered bundles over a Riemann-Cartan space-time as base possessing a homogeneous space of the Poincaré group as fiber. The soldering is obtained by identifying a Minkowski subspace of the fiber with the local tangent space of the base through an isomorphism (compare Ehresmann[10]).

In this paper we investigate fields defined on the homogeneous spaces of the (4,1)-de Sitter group which we identify with the coset spaces G/G' with $G = \mathrm{SO}(4,1)$ and G' being a subgroup of G leaving a particular point of the homogeneous space invariant (stability group or isotropy group of this

 0022-2488/85/010041-14$02.50

point). In particular, we are interested in homogeneous spaces of the form G/K, where K is a subgroup of the Lorentz group $H = \mathrm{SO}(3,1)$, which itself is a noncompact subgroup of G. K is thus not a maximal subgroup. We shall see in Sec. II that the homogeneous space G/K can itself be viewed as bundles $E^R(G/H, H/K)$ with base $G/H = \mathrm{SO}(4,1)/\mathrm{SO}(3,1) = V'_4$, being a pseudo-Riemannian space of constant (negative) curvature[11] [(4,1)-de Sitter space], and a fiber $H/K = \mathrm{SO}(3,1)/K = S$, being a homogeneous space of the Lorentz group. The action of the group G on the space E^R is a linear action on the base G/H and a nonlinear action on the fiber H/K. We exemplify the situation in the particularly interesting de Sitter case in Sec. II and discuss in Sec. III scalar functions $\Phi(\xi,\tilde{y})$, with $\xi \in G/H$ and $\tilde{y} \in H/K$, transforming linearly under the group G with respect to the first argument and nonlinearly with respect to the second. We call such a field a hybrid field. In the limiting case $K = H$, i.e., for $E^R = V'_4$, the variables $\tilde{y}$ are absent and one obtains a linearly transforming field $\Phi(\xi)$.

Section IV is devoted to a discussion of a gauge theory based on the de Sitter group involving matter fields $\Phi(x;\xi,\tilde{y})$ defined on cross sections on a soldered bundle over space-time possessing the homogeneous space $G/K = E^R(V'_4,S)$ of the (4,1)-de Sitter group as fiber. Soldering is performed by identifying the tangent space of the V'_4-part of E^R at the origin $\mathring{\xi}$ of V'_4 (which is left invariant by H) with the local tangent space T_x of space-time. Having defined the hybrid fields $\Phi(x;\xi,\tilde{y})$ we then go over to purely nonlinearly transforming fields which can be regarded as fields defined on a bundle over space-time with fiber $S = H/K$ together with a nonlinear action of G on S which becomes linear when G is restricted to H (generalized Wigner rotation). This is analogous to the nonlinear realization of a symmetry discussed by Coleman, Wess, and Zumino[12] and by Salam and Strathdee[13] (compare also Cho[14]). Finally, in Sec. V, criteria for a reduction of the gauge symmetry from G to H are given.

The space E^R which is the fiber of the original bundle over space-time can be characterized by a length parameter R being the curvature radius of its base $G/H = V'_4$. The use of the de Sitter group in conjunction with a gauge description of strong interactions has been discussed before.[15,16] In this context R plays the role of a fundamental length parameter of geometric origin which is regarded to be of the order of $R \approx 10^{-13}$ cm. While previously[15] spinor-valued matter fields defined on bundles with fiber V'_4 were used as a carrier for the de Sitter gauge symmetry, we now consider scalar functions on a higher-dimensional coset G/K, containing V'_4, with K being a subgroup of the Lorentz group.

We shall, in the present paper, not discuss field equations relating certain source expressions in terms of the matter fields to the quantities defining the geometry in the underlying bundle over a curved space-time (compare in this context Refs. 6, 7, 15, and 16). We shall come back to this question elsewhere. Here we focus the attention on the group theoretical aspects connected with the representation of matter in a gauge theory with gauge group G in the form of scalar functions[17] defined on a bundle with fiber $G/K = E^R(G/H,H/K)$, where $G \supset H \supset K$ and G and H both noncompact. It is for physical reasons connected with a gauge formulation of gravity that the group G is required to contain a subgroup H isomorphic to the Lorentz group (compare Ref. 7 for a formulation of general relativity as a gauge theory of the Lorentz group). For definiteness and later convenience the discussion is carried through for the specific case $G = \mathrm{SO}(4,1)$, $H = \mathrm{SO}(3,1)$ and an arbitrary subgroup K of $\mathrm{SO}(3,1)$. The method, however, applies more generally.

As space-time base we shall consider a Riemann-Cartan space-time U_4 although this is nowhere essential in the present context since we do not discuss field equations for metric and/or torsion. From the experience gathered in the study of the Poincaré gauge theory it is, however, to be expected that the spin degrees of freedom (contained here in the homogeneous space description related to the de Sitter group) do couple to the torsion of the space-time base.

Finally, Sec. VI is devoted to some concluding remarks. The Appendixes A and B contain various formulas for the de Sitter group used throughout the text, while Appendix C gives a classification of all the homogeneous spaces of the (4,1)-de Sitter group based on a corresponding classification of subgroups K of the Lorentz group. This part is mathematically interesting as a topic on its own right without any reference to gauge theories.

II. THE HOMOGENEOUS SPACE G/K

It is shown in Appendix B that the de Sitter group G can be viewed as a principal fiber bundle $P(G/H,H)$, over the homogeneous space $V'_4 = G/H$ with fiber and structural group $H = \mathrm{SO}(3,1)$. The group G acts on itself by left translation implying a linear action of G on the base of P according to $\xi' = A(\tilde{b},\tilde{\Lambda})\xi$, and a nonlinear action on H by left translation with the generalized Wigner rotation $\Lambda(\Lambda(b',b)) = \Lambda(\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda})$ [compare Eqs. (A37) and (A38) of Appendix A and (B6) of Appendix B]. In going over from ξ to ξ' by means of a transformation $A(\tilde{b},\tilde{\Lambda})$ of SO(4,1) the bases of the Lie algebra of H at the origin $\mathring{\xi}$ of G/H suffer an automorphism

$$\begin{aligned} R_{ik} \to R'_{ik} &= A(\Lambda(b',b))R_{ik}A^{-1}(\Lambda(b',b)) \\ &= R_{jl}[\Lambda(b',b)]^j{}_i[\Lambda(b',b)]^l{}_k. \end{aligned} \tag{2.1}$$

This is a consequence of Eq. (B2) and the metric preserving property of the transformations $A(\tilde{b},\tilde{\Lambda})$ [compare Eqs. (A3) and (A13)]. Note that $A(\Lambda(b',b)) = A(\Lambda(\tilde{b},0)\tilde{\Lambda})$, at the origin $\mathring{\xi}$, is $A(\tilde{\Lambda})$ by (A31). The corresponding automorphism for the de Sitter boost generators spanning the tangent space of V'_4 at $\mathring{\xi}$, is

$$\begin{aligned} R_{5i} \to R'_{5i} &= A(\Lambda(b',b))R_{5i}A^{-1}(\Lambda(b',b)) \\ &= R_{5j}[\Lambda(b',b)]^j{}_i. \end{aligned} \tag{2.2}$$

We now want to consider a homogeneous space of the (4,1)-de Sitter group which is characterized by a stability subgroup K being itself a subgroup of the Lorentz group, i.e., we want to investigate the homogeneous space $\mathrm{SO}(4,1)/K$, with $K \subset H = \mathrm{SO}(3,1)$, which thus has a dimension bigger than four.

Similarly as above for the group G, the homogeneous space G/K can be viewed as a fiber bundle associated to $P(G/H,H)$ with base G/H and fiber H/K, i.e.,[18]

$$G/K = E^R(G/H,H/K). \tag{2.3}$$

We denote the bundle space by E^R, where R is the elementary length parameter characterizing the base space $G/H = V_4'$ of E^R as mentioned in the Introduction (compare also Appendix A). The fiber H/K is a homogeneous space of the Lorentz group which we call S. In Appendix C a classification of the spaces G/K in terms of spaces $S = H/K$ based on a classification of the subgroups K of $H = \mathrm{SO}(3,1)$ is presented. In the subsequent discussions in this paper we shall, however, not choose a particular one from the given list of these spaces. We shall refer to any one of them as to the space E^R possessing as fiber the homogeneous space $S = H/K$ which is parametrized in terms of coordinates $\tilde{y}^n$; $n = 1,...,d$, with $d = \dim H - \dim K = 6 - \dim K$.

The bundle E^R can, for almost all points, be written as a Cartesian product of two homogeneous spaces

$$E^R(G/H,H/K) \overset{\text{a.a.}}{=} V_4' \times \mathrm{S}, \tag{2.4}$$

where the notation a.a. is explained in Appendix C. In the classification given in Appendix C we make use of this fact, while in the text we adhere to the notation (2.3) since, for arbitrary groups $G \supset H \supset K$, $E^R(G/H,H/K)$ need not generally be a product space.

Let us next determine the action of the group G on the space E^R. To this end we think of E^R as parametrized in terms of coordinates ξ labeling the points of the base V_4', and of coordinates $\tilde{y}$ labeling the points of the fiber S, i.e.,

$$E^R = \{\xi,\tilde{y}\}. \tag{2.5}$$

Associated with the structure of the spaces V_4' and S as homogeneous spaces of the groups SO(4,1) and SO(3,1), respectively, we introduce two projections (compare Appendix B)

$$\pi:\ G \to G/H, \tag{2.6}$$

and

$$\pi':\ H \to H/K, \tag{2.7}$$

and write the coordinates $(\xi,\tilde{y})$ by labeling the points of G/K in terms of group actions choosing as origin in G/K the point with the coordinates $(\mathring{\xi},\tilde{y}_0)$,

$$\xi = \pi g(b,\Lambda) = \pi g(b), \tag{2.8a}$$

$$\tilde{y} = \pi' g(\beta,\lambda) = \pi' g(\beta), \tag{2.8b}$$

where $g(b,\Lambda) = g(b)g(\Lambda)$ parametrizes $G = \mathrm{SO}(4,1)$ [see (A19) and (A9)], and

$$g(\Lambda) = g(\beta,\lambda) = g(\beta)g(\lambda) \tag{2.9}$$

parametrizes $H = \mathrm{SO}(3,1)$ in an analogous manner [$g(\beta)$ is a parametrization of H/K (with origin $\tilde{y}_0$), and $g(\lambda)$ is a parametrization of the subgroup K of H leaving $\tilde{y}_0$ invariant].

The action of the de Sitter group G on the homogeneous space $G/K = E^R$ can, in view of Eqs. (2.8) and (B6), be written in the following way [$g = g(\tilde{b},\tilde{\Lambda})$]:

$$E^R \xrightarrow{g} E'^R = \{\pi g(\tilde{b},\tilde{\Lambda})g(b),\pi'(g(\tilde{b},\tilde{\Lambda})g(b))_H g(\beta)\}. \tag{2.10}$$

The first entry on the right-hand side of (2.10) is the action of G on G/H determined in Appendix A and B yielding $\pi g(b')g(\Lambda(b',b)) = \pi g(b') = \xi'$ with $b' = \Lambda(\tilde{b})\tilde{\Lambda}b$. In the second entry the notation $(g(\tilde{b},\tilde{\Lambda})g(b))_H$ means that the part of the transformation of the product $g(\tilde{b},\tilde{\Lambda})g(b)$ lying in the subgroup H in the decomposition (A35) [here written for $g(\tilde{b},\tilde{\Lambda})$ and $g(b,1)$] is to be taken and decomposed according to Eq. (2.9) with only the part $g(\beta')$ surviving the projection π' (see below). Thus exactly the Λ-part of the product $g(\tilde{b},\tilde{\Lambda})g(b)$ which can be dropped under the projection π enters under the projection π' as the transformation $(g(\tilde{b},\tilde{\Lambda})g(b))_H$. According to Eqs. (A36) and (A37) this is just the generalized Wigner rotation $g(\Lambda(b',b))$. Thus Eq. (2.10) takes the form

$$E^R \xrightarrow{g} E'^R = \{\pi g(b'),\pi' g(\Lambda(b',b))g(\beta)\} \tag{2.11a}$$

with

$$b' = \Lambda(\tilde{b})\tilde{\Lambda}b, \tag{2.11b}$$

and

$$\Lambda(b',b) = \Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda}. \tag{2.11c}$$

Hence, under the action of the group G one obtains a linear transformation of the base space of E^R according to

$$\xi' = \Lambda(\tilde{b},\tilde{\Lambda})\xi, \tag{2.12}$$

and a nonlinear transformation of the fiber S of E^R according to

$$\tilde{y}' = \tilde{h}(\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda})\tilde{y}, \tag{2.13}$$

where $\tilde{h}$ denotes a representation of H on S. To obtain this result we use the same arguments for the group H and its subgroup K as was used for G and its subgroup H in Appendixes A and B, i.e., [compare Eqs. (2.8), (2.9), (B7), (B8), and (A36)]

$$\begin{aligned}\pi' g(\Lambda(b',b))g(\beta) &= \pi' g_{b',b}(\beta')g(\lambda_{b',b}(\beta',\beta))\\ &= \pi' g_{b',b}(\beta') = \tilde{y}'.\end{aligned} \tag{2.14}$$

We thus have established the result that the action of the group $G = \mathrm{SO}(4,1)$ on the homogeneous space $G/K = E^R(G/H,H/K)$ is the usual linear action (2.12) on the base $G/H = V_4'$ (being a pseudo-Riemannian space of constant negative curvature), and a nonlinear action (2.13) on the fiber $H/K = S$ (being a homogeneous space of the Lorentz group). The action of G in the second case is just the one given by the SO(4,1) Wigner rotation $\Lambda(b',b) = \Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda}$ depending on the parameters $\tilde{b}$ and $\tilde{\Lambda}$ of the transformation g of G and on the point $\xi \in G/H$ with $b = \xi/R$.

III. SCALAR FUNCTIONS ON HOMOGENEOUS SPACES OF THE de SITTER GROUP

In this section we discuss scalar fields defined on the homogeneous spaces G/H and G/K, with $K \subset H$.

Let us first treat the case of functions $\Phi(\xi)$ defined on $G/H = V_4'$ which we call linearly transforming fields. They form a representation of the (4,1)-de Sitter group according to the relation

$$U_{g(\tilde{b},\tilde{\Lambda})}\Phi(\xi) = \Phi(\Lambda^{-1}(\tilde{b},\tilde{\Lambda})\xi) \tag{3.1}$$

with the operators U_g obeying $U_{g_1}U_{g_2} = U_{g_1g_2}$. Here we denote, as usual, an element of G by g and in the 5×5 matrix representation acting on G/H by Λ.

Using the parametrization (A14) for the one-parameter subgroups, Eq. (3.1) yields the following representation for the generators of the de Sitter group in terms of differential operators to be applied to scalar functions $\Phi(\xi)$:

$$\tilde{L}_{ab}\Phi(\xi) = i \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[U_{g(\epsilon\hat{\omega}^{ab})}\Phi(\xi) - \Phi(\xi)\right]$$
$$= i\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\Phi(A^{-1}(\epsilon\hat{\omega}^{ab})\xi) - \Phi(\xi)\right], \qquad (3.2)$$

where $\hat{\omega}^{ab}$ is regarded to be fixed (equal to unity) and

$$\tilde{L}_{ab} = -(R_{ab})^{c}{}_{d}\,\xi^{d}\frac{\partial}{\partial\xi^{c}} = i\left(\xi_{a}\frac{\partial}{\partial\xi^{b}} - \xi_{b}\frac{\partial}{\partial\xi^{a}}\right). \qquad (3.3)$$

Here the last equality follows from (A13). The operators $\tilde{L}_{ab} = -\tilde{L}_{ba}$ are a set of ten differential operators in the variables ξ^{a} satisfying the Lie algebra (A12) of the (4,1)-de Sitter group. Although formulated in terms of the coordinates of the embedding space $R_{4,1}$ in which V_4' represents a hypersurface the operators (3.3) do not lead out of the de Sitter space V_4' when applied to a function $\Phi(\xi)$ defined on the hypersurface. Contracting $\tilde{L}_{ab}$ with ξ^{a} yields, with (A1),

$$\xi^{a}\tilde{L}_{ab} = -R^{2}i\left(\frac{\partial}{\partial\xi^{b}} + \frac{\xi_{b}\xi^{a}}{R^{2}}\frac{\partial}{\partial\xi^{a}}\right) = -R^{2}i\overset{(i)}{\partial}_{b}, \qquad (3.4)$$

where the "internal" differential operators $\overset{(i)}{\partial}_{a}$; $a = 0,1,2,3,5$, constrained by $\xi^{a}\overset{(i)}{\partial}_{a} = 0$, span the tangent space T_{ξ} of V_4' at ξ.

Let us now go over to scalar functions $\Phi(\xi,\tilde{y})$ defined on $G/K = E^{R}(G/H = V_4', H/K = S)$ with $\xi\in V_4'$ and $\tilde{y}\in S$. The discussion in Sec. II showed that their first argument transforms linearly while their second argument transforms nonlinearly under transformations $g(\tilde{b},\tilde{\Lambda})$ of the de Sitter group

$$U_{g(\tilde{b},\tilde{\Lambda})}\Phi(\xi,\tilde{y}) = \Phi(A^{-1}(\tilde{b},\tilde{\Lambda})\xi, \tilde{h}^{-1}(\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda})\tilde{y}), \qquad (3.5)$$

with Λ (Tb, $\infty\Lambda$b) $\infty\Lambda = \Lambda$ (b',b). Due to the cocycle property of the SO(4,1) Wigner rotations, Λ (b″,b′)Λ (b′,b) $= \Lambda$ (b″,b), Eq. (3.5) defines again a representation of SO(4,1) obeying $U_{g_1}U_{g_2} = U_{g_1g_2}$.[19] The action of the group G is, however, a nonlinear one with respect to the transformation of the second argument of $\Phi(\xi,\mathrm{ty})$. It becomes a linear action if g(Tb, $\infty\Lambda$) is restricted to the subgroup H, i.e., for Tb = 0 [compare Eq. (A31)]. In this case Λ (b′,b) = $\infty\Lambda$. We introduce the notation "hybrid field" to denote a function $\Phi(\xi,\mathrm{ty})$ with a transformation law as given by Eq. (3.5).

The definition of the infinitesimal generators of SO(4,1) for functions $\Phi(\xi,\tilde{y})$ is given by a formula analogous to (3.2) with $\tilde{L}_{ab}$ replaced by $\tilde{M}_{ab}$, where

$$\tilde{M}_{ab} = \tilde{L}_{ab} + \tilde{S}_{ab}. \qquad (3.6)$$

Here the operators $\tilde{L}_{ab}$, given by Eq. (3.3), affect the first argument of $\Phi(\xi,\tilde{y})$ while the $\tilde{S}_{ab}$ are differential operators affecting the second argument of $\Phi(\xi,\tilde{y})$. From the properties (A39) and (A43) of the generalized Wigner rotation we find for the operators $\tilde{S}_{ab}$

$$\tilde{S}_{ik} = -(\tilde{R}_{ik})^{n}{}_{m}\tilde{y}^{m}\frac{\partial}{\partial\tilde{y}^{n}}, \qquad (3.7)$$

and

$$\tilde{S}_{5i} = [1/(1+\tilde{\gamma})]b^{k}\tilde{S}_{ki}, \qquad (3.8)$$

where $\tilde{R}_{ik}$ is a representation of the Lorentz group acting on the homogeneous space $S = H/K$. The explicit form of the operators $\tilde{S}_{ik}$ as differential operators in the variables $\tilde{y}^{n}$ can be constructed, for the various homogeneous spaces H/K, by the method described in Ref. 3. Choosing for K the maximal compact subgroup SO(3) one again obtains a linear action of $\tilde{S}_{ik}$ on H/K while in general, for higher-dimensional cosets, the action of the Lorentz group on H/K is a hybrid one as in the analogous de Sitter case treated above. Equation (3.8) shows that the operators $\tilde{M}_{5i}$ generating the de Sitter boosts for scalar functions $\Phi(\xi,\tilde{y})$ defined on G/K possess an $\tilde{S}$-part given in terms of the operators $\tilde{S}_{ik}$ satisfying the Lie algebra of SO(3,1). The resulting expression, however, depends on the point ξ on V_4' since $b^{k}/(1+\tilde{\gamma}) = \xi^{k}/(R-\xi^{5})$. It is easy to check that the operators $\tilde{M}_{ik}$ and $\tilde{M}_{5i}$ satisfy the algebra (A18a)–(A18c) of the (4,1)-de Sitter group.

IV. GAUGE THEORY ON A BUNDLE WITH FIBER *G/K*

After the discussion given in the preceding sections concerning various homogeneous spaces of the (4,1)-de Sitter group we now turn to a gauge formulation based on the group SO(4,1). To this end we introduce a soldered fiber bundle over a Riemann–Cartan space-time U_4 as base possessing as fiber F the homogeneous space $G/K = E^{R}(V_4',S)$ and as structural or gauge group G the (4,1)-de Sitter group with an action on the fiber as determined in Sec. II. We thus introduce the bundle

$$E = E(U_4, F = E^{R}(V_4',S), G = \mathrm{SO}(4,1)). \qquad (4.1)$$

This bundle is associated to the principal bundle

$$P(U_4, F = G = \mathrm{SO}(4,1)), \qquad (4.2)$$

where G can be viewed, as noted before, as the bundle $P(G/H,H)$. The soldering of E is obtained by identifying the tangent space of the subspace V_4' of F at the point $\xi = \mathring{\xi}$ with the local tangent space T_x of U_4 at x through an isomorphism[10] (compare also Refs. 20 and 21). Thus the point $\mathring{\xi}$—the origin of V_4'— is the point of contact between base space (space-time) and fiber at $x\in U_4$. Due to the soldering there are two different Lorentz groups to be considered here acting in a "parallel" manner: On the one hand, there is the Lorentz subgroup of SO(4,1) associated with the point $\mathring{\xi}$ (its elements are the generalized Wigner rotations), on the other hand, there is a Lorentz group acting as gauge group on the usual Lorentz frame bundle $L(U_4) = P(U_4, F = G = \mathrm{SO}(3,1))$ over a Riemann–Cartan base. The local sections on $L(U_4)$ define a connection with metric and torsion components (see Refs. 6 and 7). We shall come back to a relation between these two Lorentz group actions in a different context and focus here the attention on the de Sitter gauge symmetry and its subsymmetry leaving the Lorentz gauge symmetry associated with a vierbein formulation of general relativity aside. We thus treat the de Sitter group as an internal gauge group.

Also the bundle (4.2), which is the de Sitter frame bundle over space-time studied before,[15,16] can be regarded as a soldered bundle. We denote the connection on (4.2) by the set of ten one-forms $\omega^{R}_{ab} = -\omega^{R}_{ba}$ with coefficients $\Gamma^{R}_{\mu ab}$, i.e.,

$$\omega^R_{ab} = dx^\mu \Gamma^R_{\mu ab}, \tag{4.3}$$

where dx^μ is a natural basis in T^*_x, the dual tangent space of U_4 at the point x. The $\Gamma^R_{\mu ab}$ are 40 de Sitter gauge potentials which were called de Sitter rotation coefficients in Refs. 15 and 16. Let us for later use introduce the 5×5 matrix quantity

$$\omega^R = -(i/2)\omega^R_{ab} R^{ab} \tag{4.4}$$

having matrix elements [compare (A13)]

$$(\omega^R)^a{}_b = \begin{pmatrix} \omega^{Ri}{}_k & \omega^{Ri}{}_5 \\ \omega^{R_{5i}}{}_k & 0 \end{pmatrix} \tag{4.5}$$

with the row index $a = (i,5)$ and the column index $b = (k,5)$. Introducing the one-forms

$$\hat\theta^i = \omega^{Ri}{}_5 = \omega^{R_{5i}}, \tag{4.6}$$

Eq. (4.4) can be written as

$$\omega^R = \begin{pmatrix} \hat\omega^R & \hat\theta \\ \hat\theta^T & 0 \end{pmatrix}, \tag{4.7}$$

where $\hat\omega^R$ is a 4×4 Lorentz matrix with matrix elements $(\omega^R)^i{}_k$, and $\hat\theta$ is a column vector of one-forms and $\hat\theta^T$ is transpose. The form (4.7) for the connection corresponds to the decomposition $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{p}$ of the Lie algebra of the de Sitter group with $\mathfrak{m}$ generating the Lorentz subgroup of SO(4,1) leaving the point $\mathring{\xi}$ fixed, and with $\mathfrak{p}$ being a vector subspace of $\mathfrak{g}$ spanning the tangent space to V'_4 at $\mathring{\xi}$ (see Appendix B). The forms $\hat\theta^i$, $i = 0,1,2,3$, associated with the latter, are called the soldering forms.[20]

Under a de Sitter gauge transformation, the matrix ω^R transforms according to

$$\omega'^R = A\omega^R A^{-1} - A\, dA^{-1}, \tag{4.8}$$

where $A = A(b(x),\Lambda(x))$ and, in a natural local coordinate basis on U_4, $d = dx^\mu \partial_\mu$. It is apparent that the matrix ω^R will only keep its form (4.7) if A is restricted to the Lorentz subgroup, i.e., for $b(x) = 0$. We shall come back to this below.

The curvature Ω^R is given by Cartan's structural equation for the space (4.2)

$$\Omega^R = d\omega^R - \omega^R \wedge \omega^R, \tag{4.9}$$

where d denotes the exterior derivative and $\wedge$ the exterior product of forms. In analogy to Eq. (4.7) Ω^R decomposes into a matrix of two-forms:

$$\Omega^R = \begin{pmatrix} \hat\Omega^R & \hat\Theta \\ \hat\Theta^T & 0 \end{pmatrix}, \tag{4.10}$$

where $\hat\Omega^R$ is the 4×4 matrix

$$\hat\Omega^R = d\hat\omega^R - \hat\omega^R \wedge \hat\omega^R - \hat\theta \wedge \hat\theta^T \tag{4.11}$$

and

$$\hat\Theta = \hat\nabla\hat\theta = d\hat\theta - \hat\omega^R \wedge \hat\theta. \tag{4.12}$$

In the last equation we have denoted the SO(3,1) covariant exterior derivative with respect to $\hat\omega^R$ by $\hat\nabla$.

The Bianchi identities following from Eq. (4.9) as integrability conditions read

$$\nabla\Omega^R = d\Omega^R - \omega^R \wedge \Omega^R + \Omega^R \wedge \omega^R = 0, \tag{4.13}$$

where ∇ denotes the covariant derivative with respect to the de Sitter connection (4.4).

We mention in passing that the bundle (4.2) goes over into the affine frame bundle over space-time in the limit $R\to\infty$ in which the de Sitter group contracts to the Poincaré group ISO(3,1) (Inönü–Wigner contraction), and the de Sitter space V'_4 with group of motion SO(4,1) goes over into Minkowski space M_4 with group of motion ISO(3,1).[21] This relation is worth remembering in connection with the extensive research done on the Poincaré group as a gauge group and its relation to extended theories of gravitation (compare Refs. 6 and 7 and the literature quoted there). The bundle (4.1) goes over for $R\to\infty$ into the bundles discussed in Refs. 6 and 7 possessing a homogeneous space of the Poincaré group as fiber. In the contraction limit $\hat\omega^R$ has to be identified with the connection on the Lorentz frame bundle over space-time, and the $\hat\theta^i$ go over into the forms $\theta^i + \nabla\tilde{x}^i$, with the θ^i denoting the fundamental one-forms (vierbein forms) $\theta^i = \lambda^i_\mu(x)dx^\mu$, and $\tilde{x}^i$ being an affine vector field (see Refs. 6 and 7). The $\hat\Theta^i$ play the role of the torsion in this limit. Except for some remarks in Sec. VI below we shall, however, not discuss in detail the Inönü–Wigner contraction of the structural group in the bundles (4.1) and (4.2) and treat instead the de Sitter symmetry as a gauge symmetry on its own right.

We now consider as a representative of matter in a gauge theory of the de Sitter group a generalized scalar wave function $\Phi(x;\xi,\tilde{y})$ which is defined as a map from a cross section σ: $U_4 \to E$ on $E(U_4, G/K, G = \mathrm{SO}(4,1))$ into the complex numbers[17] with $\Phi(x;\xi,\tilde{y})$ transforming as in Eq. (3.5) under de Sitter gauge transformations, $g(x) = g(\tilde{b}(x),\tilde{\Lambda}(x))$, i.e., behaving under changes of the cross section according to

$$\Phi'(x;\xi,\tilde{y}) = U_{g(x)}\Phi(x;\xi,\tilde{y}) = \Phi(x;A^{-1}\xi,\tilde{h}^{-1}(A(b',b))\tilde{y}). \tag{4.14}$$

$\Phi(x;\xi,\tilde{y})$ is a hybrid scalar de Sitter gauge field. Along with the field $\Phi(x;\xi,\tilde{y})$ there is a connection defined on E induced from a corresponding connection on (4.2). The form Γ^R defining a connection on the bundle (4.1)—and thereby defining a covariant derivative for $\Phi(x;\xi,\tilde{y})$—is the SO(4,1)-Lie algebra valued one-form acting on $\Phi(x;\xi,\tilde{y})$

$$\Gamma^R = dx^\mu \Gamma^R_\mu = dx^\mu [\tfrac{1}{2}\Gamma^R_{\mu ik}\widetilde{M}^{ik} + \Gamma^R_{\mu 5i}\widetilde{M}^{5i}]\ . \tag{4.15}$$

Here the $\widetilde{M}_{ab}$ are given by Eqs. (3.6), (3.3), (3.7), and (3.8), and the $\Gamma^R_{\mu ab} = -\Gamma^R_{\mu ba}$ are the connection coefficients defined above. We denote by $\overset{G/H}{\Gamma}{}^R$ the part in (4.15) associated with the soldering forms $\hat\theta^i$:

$$\overset{G/H}{\Gamma}{}^R = -dx^\mu \Gamma^R_{\mu 5}{}^i \widetilde{M}_{5i} = \hat\theta^i \widetilde{M}_{5i}. \tag{4.16}$$

The covariant derivative of the function $\Phi(x;\xi,\tilde{y})$ is given by

$$D_\mu \Phi(x;\xi,\tilde{y}) = (\partial_\mu + i\Gamma^R_\mu)\Phi(x;\xi,\tilde{y}), \tag{4.17}$$

with $D_\mu\Phi(x;\xi,\tilde{y})$ transforming in the same way as $\Phi(x;\xi,\tilde{y})$ implying the usual transformation rule for the connection Γ^R [compare Eq. (4.8)]

$$\Gamma'^R = U_{g(x)}\Gamma^R U^{-1}_{g(x)} - iU_{g(x)}\, dU^{-1}_{g(x)}. \tag{4.18}$$

Let us now go over to purely nonlinearly transforming fields, called ${}_{(N)}\Phi(x;\tilde{y})$, which are identical to the fields $\Phi(x;\xi,\tilde{y})$ boosted back to the origin $\xi=\overset{\circ}{\xi}$ in G/H with $\overset{\circ}{\xi}$ being the point of contact of fiber and space-time base in the soldered bundle $E(U_4,G/K,\mathrm{SO}(4,1))$ (see above), i.e., with (A9),

$$ {}_{(N)}\Phi(x;\tilde{y}) = U_{g(b)}\Phi(x;\xi,\tilde{y}). \tag{4.19}$$

Here we made use of the fact that a pure boost $g(b)$ does not affect the last argument of $\Phi(x;\xi,\tilde{y})$. Similarly one defines

$$ {}_{(N)}\Phi(x;\tilde{y}') = U_{g(b')}\Phi(x;\xi',\tilde{y}'). \tag{4.20}$$

It is easy to show from Eqs. (4.14), (4.19), and (4.20) together with the formula [compare Eqs. (A33) and (A34)]

$$ U_{g(b)}U_{g(b')}U_{g(\Lambda(b',b))}U^{-1}_{g(b)}U^{-1}_{g(b')} = U^{-1}_{g(\Lambda(b',b))} \tag{4.21}$$

that

$$ {}_{(N)}\Phi'(x;\tilde{y}) = U_{g(\Lambda(b',b))}\,{}_{(N)}\Phi(x;\tilde{y}) = {}_{(N)}\Phi(x;\tilde{y}'), \tag{4.22}$$

with

$$ \tilde{y}' = \tilde{h}(\Lambda(b',b))\tilde{y}. \tag{4.23}$$

This shows that the nonlinear fields ${}_{(N)}\Phi(x;\tilde{y})$ transform under the generalized Wigner rotation $\Lambda(b',b)$, i.e., transform nonlinearly under transformations of SO(4,1). The fields ${}_{(N)}\Phi(x;\tilde{y})$ could be viewed as defined on a cross section of a bundle

$$ E' = E'(U_4, F=S=H/K, G=\mathrm{SO}(4,1)), \tag{4.24}$$

possessing as fiber the homogeneous space S of the Lorentz group together with a nonlinear action of SO(4,1) on S given by $g(\Lambda(b',b)) = g(\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda})$. A gauge transformation on E' is expressed for scalar functions defined on E' by Eq. (4.22).

Now, what is the corresponding connection on E' defining a covariant derivative for the nonlinearly transforming fields (4.19)? In view of Eqs. (4.18) and (4.19) the connection associated with the fields ${}_{(N)}\Phi(x;\tilde{y})$ is given by

$$ {}_{(N)}\Gamma^R = U_{g(b)}\Gamma^R U^{-1}_{g(b)} - iU_{g(b)}\,dU^{-1}_{g(b)}. \tag{4.25}$$

Changing the cross section on E' according to Eqs. (4.22) and (4.23) implies that ${}_{(N)}\Gamma^R$ transforms as

$$ \begin{aligned} {}_{(N)}\Gamma'^R = {} & U_{g(\Lambda(b',b))}\,{}_{(N)}\Gamma^R U^{-1}_{g(\Lambda(b',b))} \\ & - iU_{g(\Lambda(b',b))}\,dU^{-1}_{g(\Lambda(b',b))}. \end{aligned} \tag{4.26}$$

In order to interpret Eq. (4.25) let us first observe that $U_{g(b)}$ is generated by the $\tilde{L}_{5i}$ only (a pure boost does not act on S, as mentioned above), i.e., it can be written as

$$ U_{g(b)} = l^{\,i\hat{\omega}_5{}^i\tilde{L}_{5i}}, \tag{4.27}$$

with $\hat{\omega}_5{}^i$ and b^i related by (A17) [both being x-dependent here]. The $\hat{\omega}_5{}^i(x)$ are a set of fields parametrizing the boost $U_{g(b)}$ for generalized wave functions defined as cross sections on E. In second quantized theories they are usually called the "Goldstone fields."

Any transformation $\Lambda\in\mathrm{SO}(4,1)$ induces an automorphism of the Lie algebra $\mathfrak{g}$ which in the defining 5×5 matrix representation reads

$$ R_{ab}\overset{\Lambda}{\rightarrow}R'_{ab} = \Lambda R_{ab}\Lambda^{-1} = R_{a'b'}[\Lambda]^{a'}{}_a[\Lambda]^{b'}{}_b. \tag{4.28}$$

A similar relation holds for the operators $\tilde{L}_{ab}$ as follows from Eq. (4.28) together with (3.3). Specializing to a de Sitter boost $\Lambda(b)$ one obtains

$$ \tilde{L}_{ab}\overset{\Lambda(b)\circ}{\longrightarrow}\overset{\circ}{\tilde{L}}_{ab} = U_{g(b)}\tilde{L}_{ab}U^{-1}_{g(b)} = [\Lambda(b)]^{a'}{}_a[\Lambda(b)]^{b'}{}_b\tilde{L}_{a'b'}, \tag{4.29}$$

where $\overset{\circ}{\tilde{L}}_{ab}$ is a set of operators taking particular values at the origin $\xi=\overset{\circ}{\xi}$ of V'_4 with $\overset{\circ}{\tilde{L}}_{ik}$ being zero when applied to ${}_{(N)}\Phi(x;\tilde{y})$, and with $\overset{\circ}{\tilde{L}}_{5i}$ being equivalent, in application to ${}_{(N)}\Phi(x;\tilde{y})$, i.e., at $\xi=\overset{\circ}{\xi}$, to

$$ \overset{\circ}{\tilde{L}}_{5i} = iR\overset{\circ}{\partial}_i = iR\left.\frac{\partial}{\partial\xi^i}\right|_{\xi=\overset{\circ}{\xi}}. \tag{4.30}$$

Furthermore, one finds

$$ U_{g(b)}\xi U^{-1}_{g(b)} = \overset{\circ}{\xi}, \tag{4.31}$$

showing that

$$ U_{g(b)}\tilde{M}_{ab}U^{-1}_{g(b)} = \begin{cases} \overset{\circ}{\tilde{L}}_{ik} + \tilde{S}_{ik} = \tilde{M}_{ik} & \text{for } a,b=i,k \\ \overset{\circ}{\tilde{L}}_{5i} = \overset{\circ}{\tilde{M}}_{5i} & \text{for } a,b=5,i \end{cases}. \tag{4.32}$$

The result for $a,b=i,k$ is true since the $\tilde{S}_{ik}$ and the $\tilde{L}_{5i}$ commute; the result for $a,b=5,i$ follows from (4.31) and (3.8). With (4.32) Eq. (4.25) explicitly decomposes into the following contributions taking values in the Lie algebra of H and in the vector space generating G/H, respectively,

$$ {}_{(N)}\Gamma^R = {}_{(N)}\overset{H}{\Gamma}{}^R + {}_{(N)}\overset{G/H}{\Gamma}{}^R, \tag{4.33}$$

where

$$ {}_{(N)}\overset{H}{\Gamma}{}^R = \tfrac{1}{2}\omega^R_{ik}(\overset{\circ}{\tilde{L}}{}^{ik} + \tilde{S}^{ik}), \tag{4.34}$$

which depends only on the operators $\tilde{S}^{ik}$ in applications to ${}_{(N)}\Phi(x;\tilde{y})$, and where

$$ {}_{(N)}\overset{G/H}{\Gamma}{}^R = (\hat{\theta}^i + \hat{\nabla}\hat{\omega}^{5i})\overset{\circ}{\tilde{L}}_{5i} \tag{4.35}$$

depends only on $\overset{\circ}{\tilde{L}}_{5i}$, with the last term on the right-hand side of (4.35) originating from the right-most term in (4.25) using, moreover,

$$ d(\hat{\omega}_5{}^i\tilde{L}_{5i}) = d(\hat{\omega}_5{}^i\overset{\circ}{\tilde{L}}_{5i}) = (\hat{\nabla}\hat{\omega}_5{}^i)\overset{\circ}{\tilde{L}}_{5i}, \tag{4.36}$$

which is true since $\hat{\omega}_5{}^i\tilde{L}_{5i}$ commutes with $U_{g(b)}$ as defined by (4.27). It is now immediately apparent that Eq. (4.26) can be split into an equation for ${}_{(N)}\overset{H}{\Gamma}{}^R$ possessing an inhomogeneous transformation character typical for a connection on (4.24), and into an equation for ${}_{(N)}\overset{G/H}{\Gamma}{}^R$ being, in fact, invariant under gauge transformations $g(\Lambda(b',b))$ (with $\hat{\theta}^i$ and $\hat{\nabla}\hat{\omega}_5{}^i$ transforming as Lorentz vector fields), i.e.,

$$ \begin{aligned} {}_{(N)}\overset{H}{\Gamma}{}'^R = {} & U_{g(\Lambda(b',b))}\,{}_{(N)}\overset{H}{\Gamma}{}^R U^{-1}_{g(\Lambda(b',b))} \\ & - iU_{g(\Lambda(b',b))}\,dU^{-1}_{g(\Lambda(b',b))}, \end{aligned} \tag{4.37}$$

$$ {}_{(N)}\overset{G/H}{\Gamma}{}'^R = {}_{(N)}\overset{G/H}{\Gamma}{}^R. \tag{4.38}$$

Here $U_{g(\Lambda(b',b))}$ can be viewed as being generated by the $\tilde{S}_{ik}$ only, and ${}_{(N)}\overset{G/H}{\Gamma}{}^R$ being an $\tilde{S}_{ik}$- valued one-form acting on ${}_{(N)}\Phi(x;\tilde{y})$.

V. SYMMETRY BREAKING

If $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$ is a field which is dynamically of no importance, i.e., if, for example, $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$ does not appear in the Lagrangian, one could think of eliminating it from further considerations by putting both sides of Eq. (4.38) equal to zero implying, with regard to Eq. (4.35) that the soldering forms are covariant derivatives with respect to the Lorentz part of the connection ω^R:

$$\hat{\theta}^i = \hat{\nabla}\hat{\omega}_5{}^i = d\hat{\omega}_5{}^i + \omega^R{}_k{}^i\hat{\omega}_5{}^k. \tag{5.1}$$

This means that the "Goldstone fields" $\hat{\omega}_5{}^i$ appearing in Eq. (4.27), in fact, determine the soldering forms. The $\hat{\Theta}^i$, given by Eq. (4.12), now are

$$\hat{\Theta}^i = \hat{\nabla}\hat{\theta}^i = \hat{\nabla}\hat{\nabla}\hat{\omega}_5{}^i = R^R{}_j{}^i\hat{\omega}_5^j, \tag{5.2}$$

with

$$R^R{}_{ij} = d\omega^R{}_{ij} + \omega^R{}_{ik} \wedge \omega^R{}_j{}^k. \tag{5.3}$$

These equations imply that the $\hat{\omega}_5{}^i$ together with the pure Lorentz part of the matrix of curvature forms given by the first two terms on the right-hand side of Eq. (4.11) determine the $\hat{\Theta}^i$.

We mentioned above that the form (4.7) for the connection on (4.2) remains invariant under transformations Λ (Λ) of the stability subgroup H of the point $\overset{\circ}{\xi}$ with the entries in Eq.(4.7) transforming as

$$\hat{\omega}'^R = \Lambda\hat{\omega}^R\Lambda^{-1} - \Lambda\, d\Lambda^{-1}, \tag{5.4a}$$

$$\hat{\theta}' = \Lambda\hat{\theta}. \tag{5.4b}$$

With $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$ assumed to be zero everywhere the gauge symmetry has been reduced to the subgroup H, and the original de Sitter gauge symmetry is broken down to a linear Lorentz gauge symmetry with connection ω^R_{ij} and curvature R^R_{ij}. Only the relation (5.1) is a remainder of the original SO(4,1) gauge symmetry. Assuming further that the $\hat{\omega}^{5i}$ are covariant constant would eliminate the $\hat{\theta}^i$ altogether from the description. It is tempting to relate the Lorentz gauge degrees of freedom obtained above to gravitation which in Cartan's moving frame formulation can be regarded as a gauge theory of the Lorentz group.[7] However, we shall not explore this relation here in any detail since we intend to come back to this topic in a different context.

If, on the other hand, one does not assume (5.1) to hold true, one has, in going over to the nonlinearly transforming gauge and matter fields (4.37), (4.38), and (4.19), broken the symmetry down to the subgroup H yielding now a nonlinear realization of the original gauge symmetry G. This nonlinear realization of a gauge symmetry with group G on the stability subgroup H associated with the origin of the coset space G/H is usually called, in second quantized theories, a "spontaneous symmetry breaking." In these theories G is the symmetry group of some Lagrangian defining a gauge theory, and the subgroup H is the symmetry group of the ground state of the system. While $\hat{\omega}^R$ or $\underset{(N)}{} \overset{H}{\Gamma}{}^R$ remain true gauge fields associated with the subgroup H (generalized Wigner rotations) the remaining fields $\hat{\theta}^i$ or $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$ associated with the boost generators are homogeneously transforming fields {$\hat{\theta}^i$ and $\hat{\nabla}\hat{\omega}_5{}^i$ are Lorentz vector fields [compare Eq. (5.4b)], and $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$ is an invariant [compare Eqs. (4.35) and (4.38)]}. The soldering forms $\hat{\theta}^i$ and the "Goldstone fields" $\hat{\omega}_5{}^i$ are now independent quantities which determine $\underset{(N)}{} \overset{G/H}{\Gamma}{}^R$. Actually, the $\hat{\omega}_5{}^i$ are neither geometric quantities like $\hat{\omega}^R$, $\hat{\theta}$, $\hat{\Omega}^R$, and $\hat{\Theta}$, nor are they matter quantities; they are "parameter fields" parametrizing the boosts $U_{g(b)}$ transforming the hybrid matter fields $\Phi\,(x;\xi,\tilde{y})$ into the nonlinear matter fields $\underset{(N)}{}\Phi\,(x;\tilde{y})$. Finally, the curvature on the bundle (4.2) is given by $\hat{\Omega}^R$ and $\hat{\Theta}$ defined by Eqs. (4.11) and (4.12). They are independent geometric quantities characterizing the manifold (4.2).

VI. DISCUSSION

We studied in this paper functions defined on homogeneous spaces of the (4,1)-de Sitter group G which are of the type G/K, where K is not a maximal subgroup of G but a subgroup of the Lorentz group $H = \mathrm{SO}(3,1)$ which itself is a noncompact (maximal) subgroup of G. It was shown that the spaces G/K can be viewed as fiber bundles $G/K = E^R(G/H, H/K)$, where the base, $V'_4 = G/H$, is a pseudo-Riemannian space of constant negative curvature and the fiber, $S = H/K$, is a homogeneous space of the Lorentz group. The action of G on the space G/K is a linear action on V'_4 and a nonlinear action on S with the latter given by the generalized Wigner rotations studied in Appendix A. After presenting a classification of the spaces G/K based on an analogous classification of the homogeneous spaces of the Lorentz group (see Appendix C) and introducing functions defined on G/K, we went on to discuss a gauge theory based on the (4,1)-de Sitter group with matter represented in terms of generalized wave functions, $\Phi\,(x;\xi,\tilde{y})$, defined as a map into the complex numbers from a cross section on a soldered bundle $E^R(U_4, F = G/K, G = \mathrm{SO}(4,1))$ over a Riemann–Cartan space-time U_4 having as fiber the homogeneous space G/K of the (4,1)-de Sitter group. The physical idea in this context is to represent the spin content of matter in a gauge theory in terms of scalar wave functions defined on a higher-dimensional space, in fact, a homogeneous space of the (4,1)-de Sitter group. This is a gauged version of the proposal made in Refs. 1, 4, and 3 for a field theory based on the Poincaré group, with the group $G = \mathrm{SO}(4,1)$ replacing the Poincaré group. [A discussion of a Poincaré gauge theory in a related context can be found in Refs. 6 and 7 (see also the literature quoted there). For the original motivation to use the de Sitter group see Refs. 22–25 and also 15, 16, and 26.]

Having defined the de Sitter gauge fields $\Phi\,(x;\xi,\tilde{y})$, which we call hybrid gauge fields since the argument ξ suffers a linear, while the argument $\tilde{y}$ suffers a nonlinear substitution under de Sitter gauge transformations, we introduced purely nonlinearly transforming fields $\underset{(N)}{}\Phi\,(x;\tilde{y})$. This was done by boosting the variable ξ of $\Phi\,(x;\xi,\tilde{y})$ back to the origin $\overset{\circ}{\xi}$ of $V'_4 = G/H$. In this way the gauge group G is realized in a nonlinear manner on the stability subgroup H of the point $\mathring{\xi}\in G/H$ in terms of the generalized Wigner rotations.

Finally, the geometric implications of the reduction of the symmetry from the original gauge group $G = SO(4,1)$ to the gauge group $H = SO(3,1)$ were investigated. The relation of the Lorentz subsymmetry H to the Lorentz gauge formulation of classical general relativity in terms of the Lorentz frame bundle over a Riemannian or Riemann–Cartan space-time was not discussed in detail in this paper. We intend to come back to this question elsewhere in an investigation of the Inönü–Wigner contraction of the de Sitter group to the Poincaré group in the bundle formalism presented in this paper. (This is the limit $R\to\infty$.) In a forthcoming article the field equations will be discussed which relate the matter functions to the geometry of the underlying bundle space and clarify the connection between the spin degrees of freedom contained in the generalized hybrid or nonlinear wave fields on E, and the torsion degrees of freedom contained in the soldering forms $\hat{\theta}^i$.

ACKNOWLEDGMENT

Appendix C was written by Dr. Patrick Moylan. It is a pleasure to thank Dr. Moylan for numerous discussions.

APPENDIX A: PROPERTIES OF THE (4,1)-de SITTER GROUP

The (4,1)-de Sitter group is the group of hyperbolic rotations in a five-dimensional Lorentzian space $R_{4,1}$ which leaves the quadratic form

$$\xi^a \xi_a = \xi^a \xi^b \eta_{ab} = -R^2, \tag{A1}$$

with

$$\eta_{ab} = \mathrm{diag}(1,-1,-1,-1,-1) \tag{A2}$$

invariant. (The summation convention for repeated indices over the range $a,b = 0,1,2,3,5$ is assumed throughout.) We denote by SO(4,1) the identity component of the four-fold connected pseudo-orthogonal group G leaving (A1) invariant, and denote its elements in the defining 5×5 matrix representation by A. SO(4,1) is a short-hand notation for $O(4,1)^{++}$ with $++$ standing for sign $A^0{}_0 = +1$ and $\det(A^a{}_b) = +1$. The metric preserving property of the de Sitter transformations implies the pseudo-orthogonality relation

$$A^{-1} = \eta A^T \eta, \tag{A3}$$

where T denotes the transpose and η is the de Sitter metric with components (A2).

The hypersurface defined in $R_{4,1}$ by Eq. (A1) is a one-shell hyperboloid which is a model for a four-dimensional pseudo-Riemannian space V_4' of constant negative curvature with curvature radius R and curvature scalar $-1/R^2$ [(4,1)-de Sitter space] on which the group SO(4,1) acts as a group of motions, i.e., on which SO(4,1) acts transitively. V_4' is thus a homogeneous space (compare Appendix C). In the mathematical literature the radius of the de Sitter hyperboloid V_4' is usually considered to be unity. For physical reasons we shall not make this assumption here; moreover, V_4' will be identified with the noncompact coset space (compare Helgason[11])

$$G/H = SO(4,1)/SO(3,1) = V_4', \tag{A4}$$

where H is the (noncompact) stability subgroup with elements

$$A(\Lambda) = \left(\begin{array}{c|c} \Lambda & \begin{matrix}0\\0\\0\\0\end{matrix} \\ \hline \begin{matrix}0 & 0 & 0 & 0\end{matrix} & 1 \end{array}\right)$$

leaving the particular point $\mathring{\xi}$ of V_4' invariant [being the proper orthochronous Lorentz group $SO(3,1) = O(3,1)^{++}$] and with $\mathring{\xi}$ given by

$$\mathring{\xi} = \begin{pmatrix} \mathring{\xi}^0 \\ \mathring{\xi}^1 \\ \mathring{\xi}^2 \\ \mathring{\xi}^3 \\ \mathring{\xi}^5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -R \end{pmatrix}. \tag{A5}$$

The aspects of the homogeneous space V_4' as a coset space are discussed in Appendix B.

Any point $\xi\in V_4'$ can be obtained from $\mathring{\xi}$ with the help of a de Sitter boost $A(b)$ according to $\xi = A(b)\mathring{\xi}$, which reads in components

$$\xi^a = [A(b)]^a{}_b \mathring{\xi}^b, \tag{A6}$$

with

$$[A(b)]^a{}_b = \begin{pmatrix} \delta^i_k + \dfrac{b^i b_k}{1-\epsilon\bar{\gamma}} & -b^i \\ -b_k & -\epsilon\bar{\gamma} \end{pmatrix}, \tag{A7}$$

where

$$|b^5| = \bar{\gamma} = {}_{+}\sqrt{1 + b^j b_j} \text{ and } \epsilon = \mathrm{sign}\, b^5. \tag{A8}$$

Here $a = (i,5)$ is the row index and $b = (k,5)$ the column index with the indices i,k in the range 0,1,2,3. Indices $a,b,c,d,\ldots$ are raised and lowered with the de Sitter metric η^{ab} and η_{ab}, respectively, while indices $i,j,k,l,\ldots$ are raised and lowered with the Lorentz metric η^{ik} and η_{ik}, respectively; $[\eta_{ik} = \mathrm{diag}(1,-1,-1,-1)]$. The boost $A(b)$ depends on b^k; $k = 0,1,2,3$, and ϵ. However, for simplicity we write $A(b)$ instead of $A(b^k,\epsilon)$ [with the exception of Eqs. (10a) and (10b) below]. Equation (A6) shows that

$$b^k = \xi^k/R \text{ and } \epsilon\bar{\gamma} = b^5 = \xi^5/R. \tag{A9}$$

From (A3) it follows that the inverse boost and the identity are given by

$$A^{-1}(b) = A^{-1}(b^k,\epsilon) = A(-b^k,\epsilon), \tag{A10a}$$

$$A(b^k = 0, \epsilon = -1) = 1. \tag{A10b}$$

For $\epsilon = +1$ one has for

$$b^k\to 0:\ A(b^k = 0, \epsilon = 1) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix},$$

which is an element of $O(4,1)^{-+}$ carrying $\mathring{\xi}$ over into $\mathring{\xi}' = (0,0,0,0,R)$. We mention in passing that frequently in the literature antipodal points ξ^a and $-\xi^a$ are identified on the de Sitter hyperboloid (compare Gelfand, Graev, and Vilenkin[27]) thus allowing one to restrict the boost to the case $\epsilon = -1$ [compare the corresponding choice (A5) for $\mathring{\xi}$]. In this paper we shall follow this convention taking thus essentially half the hyperboloid (A1) as a model for the (4,1)-de Sitter space and assuming $\epsilon = -1$ in Eq. (A7). (Compare also Hannabus[28] in this context.)

A transformation $A \in SO(4,1)$ carrying a point $\xi \in V_4'$ over into $\xi' \in V_4'$, i.e., $\xi' = A\xi$, can be broken down into two boosts and a Lorentz transformation $\overline{\Lambda}(b',b)$ according to

$$A = A(b')A(\overline{\Lambda}(b',b))A^{-1}(b), \tag{A11}$$

with b^k given by (A9) and similarly for b'^k. The form and properties of the transformation $A(\overline{\Lambda}(b',b))$, being the analog for SO(4,1) of the usual Wigner rotation, will be discussed below.

A 5×5 matrix representation of the Lie algebra of the (4,1)-de Sitter group is given by the matrices $R_{ab} = -R_{ba}$ obeying the commutation relations

$$i[R_{ab},R_{cd}] = \eta_{ac}R_{bd} + \eta_{bd}R_{ac} - \eta_{ad}R_{bc} - \eta_{bc}R_{ad}. \tag{A12}$$

The matrix elements of the R_{ab} are

$$(R_{ab})^c{}_d = -i[\eta_{ad}\delta_b{}^c - \eta_{bd}\delta_a{}^c]. \tag{A13}$$

For the ten one-parameter subgroups of SO(4,1) with parameters $\hat{\omega}^{ab} = -\hat{\omega}^{ba}$ generated by R_{ab} we write

$$A(\hat{\omega}^{ab}) = e^{-i\hat{\omega}^{ab}R_{ab}}, \tag{A14}$$

where in the exponent no summation is performed. The parameters $\hat{\omega}_5{}^i = -\hat{\omega}^{5i}$ can be used to parametrize the transformations (A7) for boosts in a particular direction characterized by b^i for $i = 0,1,2$ or 3, according to (remember that $\epsilon = -1$)

$$b^0 = \sinh \hat{\omega}_5{}^0$$
$$b^r = \sin \hat{\omega}_5{}^r, \quad r = 1,2,3. \tag{A15}$$

The parameters $\hat{\omega}^{ik}$ parametrize the Lorentz subgroup of SO(4,1).

The space V_4' can be parametrized by the exponential

$$[\exp(i\hat{\omega}_5{}^j R_{5j})]^a{}_b = \begin{pmatrix} \delta^i_k + \dfrac{b^i b_k}{1+\bar{\gamma}} & -b^i \\ -b_k & \bar{\gamma} \end{pmatrix}, \tag{A16}$$

with

$$b^i = \hat{\omega}_5{}^i \frac{\sinh\sqrt{\hat{\omega}_5{}^j\hat{\omega}_{5j}}}{\sqrt{\hat{\omega}_5{}^j\hat{\omega}_{5j}}}, \tag{A17}$$

where here and in Eq. (A16) a summation over $j = 0,1,2,3$ is implied (compare Gilmore[20]). It is apparent that Eqs. (A16) and (A17) reduce to the one-parameter subgroups generated by a single R_{5i}, for $i = 0,1,2,3$, with the b^i given by (A15).

Decomposing the Lie algebra $\mathfrak{g}$ of SO(4,1) into $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{p}$, where the elements of $\mathfrak{m}$ (represented here in terms of six 5×5 matrices R_{ik}) generate the subgroup H, and those of $\mathfrak{p}$ (represented here by the four boost generators R_{5i}) generate the space V_4', i.e., the coset space G/H (see Appendix B), the commutation relations (A12) take the form

$$i[R_{ij},R_{kl}] = \eta_{ik}R_{jl} + \eta_{jl}R_{ik} - \eta_{il}R_{jk} - \eta_{jk}R_{il}, \tag{A18a}$$

$$i[R_{5i},R_{jk}] = \eta_{ik}R_{5j} - \eta_{ij}R_{5k}, \tag{A18b}$$

$$i[R_{5i},R_{5j}] = -R_{ij}. \tag{A18c}$$

Any element A of SO(4,1) can be parametrized in the following way (noncompact version of the Cartan decomposition) (compare Ref. 28):

$$A = A(b,\Lambda) = A(b)A(\Lambda) \tag{A19}$$

in terms of boost parameters b^k parametrizing the coset space G/H, and in terms of parameters $\hat{\omega}^{ik}$ parametrizing the Lorentz subgroup H of SO(4,1). We always write the boost to the left of the Λ-transformation. [The particular parametrization of $A(\Lambda)$ in terms of one-parameter subgroups of the Lorentz group will be discussed in Appendix C in connection with the discussion of the homogeneous spaces of the Lorentz group.] It is easy to show from (A7) that

$$A(\Lambda)A(b)A^{-1}(\Lambda) = A(\Lambda b). \tag{A20}$$

Next we consider the product of two boosts with parameters b'^k and b^k and write the result in the form (A19). Explicit calculation yields

$$\begin{aligned} A(b')A(b) &= A(B)A(\Lambda(b',b)) \\ &= A(\Lambda(b',b))A(\overline{B}), \end{aligned} \tag{A21}$$

where

$$B = A(b')b \quad\text{and}\quad \overline{B} = A(b)b', \tag{A22}$$

with

$$B = \begin{pmatrix} B^i \\ B^5 = -\Gamma \end{pmatrix} \quad\text{and}\quad \overline{B} = \begin{pmatrix} \overline{B}^i \\ \overline{B}^5 = -\Gamma \end{pmatrix}, \tag{A23}$$

where

$$B^i = b^i + b'^i(\bar{\gamma} + b^j b'_j/(1+\bar{\gamma}')), \tag{A24}$$

$$\overline{B}^i = b'^i + b^i(\bar{\gamma}' + b^j b'_j/(1+\bar{\gamma})), \tag{A25}$$

and

$$B^5 = \overline{B}^5 = -\Gamma = -(b^k b'_k + \bar{\gamma}\bar{\gamma}'), \tag{A26}$$

with $\bar{\gamma}$ given by (A8) and analogously for $\bar{\gamma}'$. It is apparent that $\overline{B}^a$ and B^a are related by a Lorentz transformation $\Lambda^{-1}(\Lambda(b',b))$. Written down for the first four components (the fifth component remains unaltered) one has

$$\overline{B}^k = [\Lambda^{-1}(b',b)]^k{}_j B^j, \tag{A27}$$

where the matrix elements of $\Lambda(b',b)$ are given by

$$\begin{aligned} [\Lambda(b',b)]^i_s &= \left(\delta^i_k + \frac{b'^i b'_k}{1+\bar{\gamma}'}\right)\left(\delta^k_s + \frac{b^k b_s}{1+\bar{\gamma}}\right) \\ &\quad + b'^i b_s - \frac{B^i \overline{B}_s}{1+\Gamma}. \end{aligned} \tag{A28}$$

From this expression one immediately derives the following properties for the Lorentz transformation $\Lambda(b',b)$:

$$\Lambda^{-1}(b',b) = \Lambda(b,b'), \tag{A29}$$

$$\Lambda(b,b) = \Lambda(b,-b) = 1, \tag{A30}$$

$$\Lambda(b,0)=\Lambda(0,b)=1, \tag{A31}$$

$$\Lambda(\tilde{\Lambda}b,b)=\tilde{\Lambda}. \tag{A32}$$

The last equation will be established below in a more direct way.

It is now easy to see that the generalized Wigner rotation $\overline{\Lambda}(b',b)$ appearing in (A11) is identical to the Lorentz transformation appearing in the boost product (A21). Using (A27)—written for the five-component vectors with the help of (A22) in the form $A(b)b'=A(\Lambda^{-1}(b',b))A(b')b$ —and comparing with (A11) one finds

$$A(\overline{\Lambda}(b',b))=A^{-1}(b')A^{-1}(b)A(\Lambda^{-1}(b',b))A(b')A(b), \tag{A33}$$

which results, with the help of (A21) and (A29), in

$$A(\overline{\Lambda}(b',b))=A(\Lambda^{-1}(b,b'))=A(\Lambda(b',b)). \tag{A34}$$

Henceforth, we thus shall drop the bar and refer to the Lorentz transformation $\Lambda(b',b)$ as to the generalized Wigner rotation [or SO(4,1) Wigner rotation] associated with the points $\xi'=Rb'$ and $\xi=Rb$ of V_4'.

Using Eq. (A21) the multiplication rule for the elements of the de Sitter group in the parametrization (A19) takes the form

$$A(b_1,\Lambda_1)A(b_2,\Lambda_2)$$
$$=A(B=A(b_1)\Lambda_1 b_2)A(\Lambda(b_1,\Lambda_1 b_2)\Lambda_1\Lambda_2). \tag{A35}$$

Let us now regard the element A of SO(4,1), with the property $\xi'=A\xi$ in the parametrization (A19) as given by $A=A(\tilde{b},\tilde{\Lambda})$. After multiplication with $A(b)$ from the right Eq. (A11) then takes the form

$$A(\tilde{b},\tilde{\Lambda})A(b)=A(b')A(\Lambda(b',b)). \tag{A36}$$

This is the typical equation discussed in theories based on nonlinear realizations of groups.[12,13] In the present case the de Sitter group is nonlinearly realized on the stability subgroup of the origin $\mathring{\xi}$ of the coset space $G/H=\mathrm{SO}(4,1)/\mathrm{SO}(3,1)=V_4'$ through what we called the "Wigner rotations." The realization becomes linear if $A(\tilde{b},\tilde{\Lambda})$ is restricted to the subgroup H, i.e., for $A(\tilde{b}=0,\tilde{\Lambda})$. This can be seen immediately in the following way: Considering the left-hand side of (A36) as a product according to (A35) one finds

$$A(\tilde{B}=A(\tilde{b})\tilde{\Lambda}b)A(\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda})=A(b')A(\Lambda(b',b)), \tag{A37}$$

implying that

$$b'=A(\tilde{b})\tilde{\Lambda}b \tag{A38a}$$

and

$$\Lambda(b',b)=\Lambda(\tilde{b},\tilde{\Lambda}b)\tilde{\Lambda}. \tag{A38b}$$

These equations express the nonlinear character of the realization of the de Sitter group G on the Lorentz subgroup H. For $\tilde{b}^k=0$, i.e., $A(\tilde{b})=1$ implying $b'=\Lambda b$, Eq. (A38b) yields, using (A31),

$$\Lambda(\tilde{\Lambda}b,b)=\Lambda(0,\tilde{\Lambda}b)\tilde{\Lambda}=\tilde{\Lambda}. \tag{A39}$$

This is identical to (A32) showing, as stated above, that one obtains a linear realization of the group G on H, the stability subgroup of the origin $\mathring{\xi}$, in restricting the transformations of G to H. Taking finally $\tilde{\Lambda}=1$ in Eqs. (A38) leads to the formula

$$A(\Lambda(\tilde{b})b,b)=\Lambda(\tilde{b},b)=\Lambda(\tilde{b},A(b)\tilde{b}), \tag{A40}$$

where the last equality follows from (A29) interchanging the roles of b and $\tilde{b}$. This shows that, in particular, the transformation $A(\Lambda(b',b))$ can be expressed as a product of three boosts:

$$A^{-1}(b')A(\tilde{b})A(b)=A(\Lambda(\tilde{b},b)) \tag{A41}$$

with $b'=A(\tilde{b})b$.

We finally derive the expression for the transformation $A(\Lambda(\tilde{b},b))$ for infinitesimal $\tilde{b}$ and arbitrary b. Computing to first order in the $\tilde{b}^k$ using

$$A(\tilde{b})=1+i\tilde{b}^i R_{5i} \tag{A42}$$

in (A41), or starting directly from (A28), one finds

$$A(\Lambda(\tilde{b},b))=1-(i/(1+\overline{\gamma}))\tilde{b}^i b^k R_{ik}. \tag{A43}$$

This relation shows explicitly the dependence of the SO(4,1) Wigner rotation on the boost parameters $\tilde{b}^i$ and on the parameters of an arbitrary point ξ of the de Sitter space with $b^k/(1+\overline{\gamma})=\xi^k/(R-\xi^5)$.

APPENDIX B: COSET SPACES OF SO(4,1)

In Appendix A the Lie algebra $\mathfrak{g}$ of the de Sitter group G was decomposed according to $\mathfrak{g}=\mathfrak{m}\oplus\mathfrak{p}$, where $\mathfrak{m}$ is a subalgebra generating the stability subgroup H of the point $\mathring{\xi}$ of V_4' (the Lorentz group), and where $\mathfrak{p}$ is a vector subspace of $\mathfrak{g}$ spanning the tangent space to V_4' at $\mathring{\xi}$ with exp $\mathfrak{p}$ parametrizing the homogeneous space V_4' which can be identified with the coset space G/H (see below). The elements of exp $\mathfrak{p}$ were called the de Sitter boosts in Appendix A which were denoted, in the defining 5×5 matrix representation, by $A(b)$. The basis of the subalgebra $\mathfrak{m}$ was in this representation denoted by R_{ik}, $i<k$; $i,k=0,1,2,3$, and the basis of the vector space $\mathfrak{p}$ was denoted by R_{5i}; $i=0,1,2,3$. There is an involutive automorphism τ defined on $\mathfrak{g}$ with eigenvalue $+1$ on $\mathfrak{m}$ and eigenvalue -1 on $\mathfrak{p}$. This automorphism is defined by $R_{ab}\to R'_{ab}=I_5 R_{ab} I_5^{-1}$, where $I_5=\mathrm{diag}(1,1,1,1,-1)$.

Since we need in Sec. II the action of $G=\mathrm{SO}(4,1)$ on homogeneous spaces different from V_4' we generalize the notation slightly so that it applies to any representation of the group G. Denoting now by g the elements of G and by h those of H we have according to (A19) the parametrization

$$g=g(b,\Lambda)=g(b)g(\Lambda)=g(b)h, \tag{B1}$$

where $g(b)$ is a boost associated with a transformation on V_4' carrying $\mathring{\xi}$ over into ξ, and where $g(\Lambda)=h$ is a Lorentz transformation associated with a corresponding transformation on V_4' leaving the origin $\mathring{\xi}$ fixed. Whenever the action of G on V_4' is meant we can write again $A(b,\Lambda)$ instead of $g(b,\Lambda)$.

Since H is the stability subgroup associated with the point $\mathring{\xi}$ the subgroup H_ξ leaving the point $\xi\in V_4'$ invariant is given by conjugation of H with g:

$$H_\xi=gHg^{-1}. \tag{B2}$$

Equation (B1) shows that this is equal to $g(b)Hg^{-1}(b)$ since $hHh^{-1}=H$.

Clearly, with every boost $g(b)$ also the transformation $g(b)\tilde{h}$, with any $\tilde{h}\in H$, corresponds to a transformation on V_4'

carrying $\overset{\circ}{\xi}$ over into ξ. Let us, therefore, introduce the following equivalence relation: Two elements g_1 and g_2 of G are equivalent (i.e., determine the same point ξ on V'_4 when applied to $\overset{\circ}{\xi}$) if $g_1^{-1}g_2 \in H$. This allows one to identify the point $\xi \in V'_4$ with the left coset gH and thus to identify V'_4 with G/H, the space of left cosets of G with respect to H. One can view G/H as the quotient space of G by the above-mentioned equivalence relation and define a projection π:

$$\pi: G \to G/H \tag{B3}$$

associating with every $g \in G$ the element

$$\pi g = gH \tag{B4}$$

of G/H. The group G acts transitively on $G/H = V'_4$ according to (compare Lichnerowicz[18])

$$A(\tilde{b},\tilde{A})\pi = \pi L_{g(\tilde{b},\tilde{A})}, \tag{B5}$$

where L_g denotes the left translation in the group G by g. Applying this to the group element $g(b,A)$ yields

$$A(\tilde{b},\tilde{A})\pi g(b,A) = \pi L_{g(\tilde{b},\tilde{A})} g(b,A) = \pi g(\tilde{b},\tilde{A}) g(b,A), \tag{B6}$$

with

$$\pi g(b,A) = \pi g(b) = \xi. \tag{B7}$$

Thus, Eq. (B6) is the transformation $\xi' = A(\tilde{b},\tilde{A})\xi$ in its relation to the action of the group on itself by left translation. Using Eq. (A35) for the multiplication rule of the de Sitter group the right-hand side of (B6) can be written as

$$\pi g(\tilde{b},\tilde{A}) g(b,A) = \pi g(b') = \xi', \tag{B8}$$

with

$$b' = A(\tilde{b})\tilde{A}b.$$

The computational rule in connection with Eqs. (B7) and (B8) is that under the projection operator π any subgroup transformation h can be dropped after having moved it to the right in using $g(A)g(b) = g(Ab)g(A)$ [compare (A20)].

It is seen from Eqs. (B3)–(B5) that under the projection π the group G can be viewed as a principal fiber bundle over G/H with fiber and structural group H,

$$G = P(G/H,H), \tag{B9}$$

with the group H acting on itself by left translation [see the right-hand side of (B5) for $\tilde{b} = 0$]. The action of the group G on the base is provided, as we have seen, by Eq. (B6).

We note that $P(G/H,H)$, which is a Lorentz frame bundle over de Sitter space V'_4, can be viewed for almost all points as a Cartesian product (compare Appendix C). Moreover, it is shown in Refs. 18 and 30 that a canonical invariant connection is defined on $P(G/H,H)$.

APPENDIX C: A CLASSIFICATION OF CERTAIN HOMOGENEOUS SPACES OF SO(4,1) (BY P. MOYLAN)

In this appendix we present a classification of the homogeneous spaces of the group SO(4,1) which are based on a corresponding classification of the homogeneous spaces of the Poincaré group presented in Refs. 3, 4, and 31. First we recall the definition of a homogeneous space E (compare Appendix B).

Definition: A homogeneous space E of a group G is a manifold E, on which an action of G is defined, and which satisfies: (a) the action $y \xrightarrow{g \in G} gy \in E$ is a C^∞ differentiable map on E; (b) the action is transitive.

There is a one-to-one correspondence between homogeneous spaces of G and subgroups of G which differ by conjugation with elements of G from one another. This is seen as follows: write every element of G in the form $g = g_x g'$, where $g' \in G' \subset G$ is an element of the stability subgroup G' of the point $y_0 \in E$, and $g_x \in G/G'$. By transitivity every element $y \in E$ can be written as

$$y = g_x g' y_0 = g_x y_0. \tag{C1}$$

Thus the space G/G' can be identified with E as shown in Appendix B. As stated there the stability subgroups of two different points y and y_0 are conjugate to one another [Eq. (B2)] and therefore the classification of all homogeneous spaces is equivalent to the determination of all subgroups of G up to conjugation. The mathematical problem of determining all subgroups of a given Lie group is a very difficult one and is certainly not explicitly solved.[32] However, the classification of all subgroups of $G = \mathrm{SO}(4,1)$ that are obtained by exponentiation from the corresponding classification of the Lie subalgebras of the Lie algebra of SO(4,1) has been given in Ref. 31. In our classification we will require that the subgroup $K \subset \mathrm{SO}(4,1)$ satisfies the additional condition $K \subset H = \mathrm{SO}(3,1)$, and we will only need the results presented in Refs. 3 and 4.

In Sec. II it was stated that the homogeneous space $G/K = \mathrm{SO}(4,1)/K$ is a fiber bundle $E^R(G/H,H/K)$ with fiber H/K and base G/H associated to the principal fiber bundle $G = P(G/H,H)$. Explicitly it is given by

$$E^R = E^R(G/H,H/K) = (P(G/H,H) \times H/K)/H \tag{C2}$$

with the projection π_E being the mapping of E^R onto G/H induced by the mapping π_P of $P(G/H,H)$ onto G/H. It is a fact that "almost all" elements of SO(4,1) admit a decomposition as[28]:

$$\mathrm{SO}(4,1) = P(G/H,H) \overset{\text{a.a}}{=} G/H \times H. \tag{C3}$$

Here $\overset{\text{a.a}}{=}$ means equality except for the points of a set of measure zero in $P(G/H,H)$ and G/H (see Ref. 28). Since we wish to investigate the properties of functions defined on the spaces E^R, sets of measure zero play a relatively unimportant role in most of the analysis and we consider the principal bundle (C3) as essentially a Cartesian product. An analogous result for E^R (almost every element has a unique decomposition into an element of V'_4 and an element of S) follows from (C3) together with the construction of E^R in (C2). Thus the homogeneous spaces which we consider have for "almost all" elements a decomposition of the following form:

$$E^R \overset{\text{a.a}}{=} V'_4 \times S, \tag{C4}$$

with $S = H/K$, where $K \subseteq H = \mathrm{SO}(3,1) \subset \mathrm{SO}(4,1)$. K is a stabilizer subgroup associated with a corresponding Lie subalgebra of the Lie algebra of H through exponentiation. The elements $(\xi,\tilde{y})$ of E^R are given by Eq. (2.8), and the action of

$g \in G$ on $(\xi,\tilde{y})$ is given from Eqs. (2.11a), (2.11b), and (2.11c) by

$$(\xi,\tilde{y}) \overset{g=g(\tilde{b},\tilde{A})}{\longrightarrow} (\xi',\tilde{y}'), \tag{C5}$$

with

$$\xi' = A_g \xi \quad [\text{Eq. (2.12)}] \tag{C6}$$

and

$$\tilde{y}' = \tilde{h}(A(\tilde{b},\tilde{A}b)\tilde{A})\tilde{y} \quad [\text{Eq. (2.13)}], \tag{C7}$$

where $\tilde{h}$ denotes an action of H on S and $A(\tilde{b},\tilde{A}b)\tilde{A}$ is the SO(4,1) Wigner rotation associated with the points ξ and ξ' of the base V'_4. We note that for certain g's and certain ξ's and $\tilde{y}$'s the transformed points may not have such a decomposition as in (C5). However, these points do not affect the functional analysis, since they form, for fixed g, always a set of measure zero.[33,34]

The infinitesimal generators of the representation of SO(4,1) defined by (3.5) are defined for $g(\epsilon\hat{\omega}^{ab}) = \exp(-i\epsilon\hat{\omega}^{ab}R_{ab})$ (no summation over a, b, and $\hat{\omega}^{ab}=1$ fixed) by

$$\widetilde{M}_{ab}\Phi(\xi,\tilde{y}) = i\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[U_{g(\epsilon\hat{\omega}^{ab})}\Phi(\xi,\tilde{y}) - \Phi(\xi,\tilde{y})\right], \tag{C8}$$

where $U_g\Phi(\xi,\tilde{y})$ is given by Eq. (3.5). Here the functions

$$\Phi:\ E^R \to \mathbb{C}$$

are required to be C^∞. The $\widetilde{M}_{ab}$ (see Sec. III) are

$$\widetilde{M}_{ab} = \widetilde{L}_{ab} + \widetilde{S}_{ab}, \tag{C9}$$

with

$$\widetilde{L}_{ab} = i\left(\xi_a\frac{\partial}{\partial\xi^b} - \xi_b\frac{\partial}{\partial\xi^a}\right), \tag{C10}$$

and

$$\widetilde{S}_{ab} = \begin{cases}\widetilde{S}_{ik}; & i,k = 0,1,2,3,\\ \widetilde{S}_{5i} = [1/(1+\gamma)]b^k\widetilde{S}_{ki}.\end{cases} \tag{C11}$$

The $\widetilde{S}_{ik}$ for the various homogeneous spaces $S = H/K$ are determined by the equation

$$\widetilde{S}_{ik}f(\tilde{y}) = i\frac{d}{d\epsilon}f\{\tilde{h}(e^{-i\epsilon R_{ik}})\tilde{y}\}, \tag{C12}$$

where R_{ik} is the generator of the Lorentz transformation in the i,k plane.

The classification of the various possible homogeneous spaces subject to the conditions given above reduces to the problem of finding all Lie subalgebras k of the Lie algebra of SL(2,$\mathbb{C}$) [the covering group of SO(3,1)]. The stabilizers K are then obtained through exponentiation from k. This problem was solved by Finkelstein[4] and later treated in more detail by Bacry and Kihlberg.[3] Using their results we give a list of the various possible homogeneous spaces in Table I. We have listed explicitly in column 1 those homogeneous spaces E^R for which the stability subgroup G' is particularly simple. In column 2 we give a convenient parametrization of these homogeneous spaces which arise from various decompositions of H as discussed below (see also Ref. 3). In column 3 the dimensions of the spaces E^R are listed, and in column 4 the bases of the Lie algebras associated with the stabilizer subgroups are given. In column 5 the SO(4,1) invariant measures of the homogeneous spaces are listed, and in column 6 the Finkelstein classification is given.

In order to understand the parametrization listed in column 2 of the table we consider the Iwasawa decomposition of H

$$H = KAN, \tag{C13}$$

where K is the maximal compact subgroup SO(3) of H and A

TABLE I. Homogeneous spaces of SO(4,1) associated with stabilizer subgroups of $H = \mathrm{SO}(3,1)$.

$\overline{E}^R$	Parameters of covering space[a]	Dimension	Generators of stabilizer	SO(4,1) invariant measure	Notation of Finkelstein
$\overline{\mathrm{SO}(4,1)}$	$\xi^\mu,\varphi,\theta,\psi,s,t,u$	10	0	$d\xi\, e^{2s}\,dt\,du\,d\varphi\,d\cos\theta\,d\psi$	[6]
	$\xi^\mu,\varphi,\theta,\psi,s,t$	9	$L_{02}-L_{23}$	$d\xi\, e^{2s}\,ds\,dt\,d\varphi\,d\cos\theta\,d\psi$	[5]
$\overline{\mathrm{SO}(4,1)}/\overline{\mathrm{SO}(2)}$	$\xi^\mu,\varphi,\theta,s,\tilde{t},\tilde{u}$	9	L_{12}	$d\xi\,ds\,d\tilde{t}\,d\tilde{u}\,d\varphi\,d\cos\theta$	[5_0]
	$\xi^\mu,\varphi,\theta,\tilde{\psi},\tilde{t},\tilde{u}$	9	$\cos\frac{f}{2}L_{12}+\sin\frac{f}{2}L_{03}$ $0<f\leqslant\pi$	$d\xi\,d\tilde{t}\,d\tilde{u}\,d\varphi\,d\cos\theta\,d\tilde{\psi}$	[5_f]
$\overline{\mathrm{SO}(4,1)}/\overline{N}$	$\xi^\mu,\varphi,\theta,\psi,s$	8	$L_{02}-L_{23}$, $L_{01}+L_{31}$	$d\xi\, e^{2s}\,ds\,d\varphi\,d\cos\theta\,d\psi$	[4]
	$\xi^\mu,\varphi,\theta,\psi,\tilde{t}$	8	$L_{02}-L_{23}$ L_{03}	No	[4′]
	$\xi^\mu,\varphi,\theta,\tilde{t},\tilde{u}$	8	L_{12},L_{03}	$d\xi\,d\tilde{t}\,d\tilde{u}\,d\varphi\,d\cos\theta$	[4″]
$\overline{\mathrm{SO}(4,1)}/\overline{\mathrm{SO}(3)}$	$\xi^\mu,q^\mu;q^\mu q_\mu=1$	7	L_{12},L_{23},L_{31}	$d\xi\,\frac{d^3q}{q_0}$	[3]
$\overline{\mathrm{SO}(4,1)}/\overline{\mathrm{SO}(2,1)}$	$\xi^\mu,q^\mu;q^\mu q_\mu=-1$	7	L_{12},L_{01},L_{02}	$d\xi\,\frac{dq^0\,dq^1\,dq^2}{q^3}$	[3′]
	ξ^μ,φ,θ,s	7	$L_{12},L_{02}-L_{23}$, $L_{01}+L_{31}$ $0<f\leqslant\pi$	$d\xi\, e^{2s}\,ds\,d\varphi\,d\cos\theta$	[3_0]
	$\xi^\mu,\varphi,\theta,\tilde{\psi}$	7	$\cos\frac{f}{2}L_{12}+\sin\frac{f}{2}L_{03};$ $L_{02}-L_{23};L_{01}+L_{31}$ $0<f\leqslant\pi$	No	[3_f]
	ξ^μ,φ,θ	6	$L_{12},L_{03},L_{02}-L_{23}$, $L_{01}+L_{31}$	No	[2]
V'_4	ξ^μ	4	$L_{\mu\nu}$	$d\xi$	[0]

[a] The covering space is the corresponding homogeneous space, $\overline{E}$, associated with the universal covering group of SO(4,1), i.e., $\overline{E} = \overline{\mathrm{SO}(4,1)}/\overline{K}$, where a bar denotes the covering group.

and N are abelian subgroups of H. For the covering group SL(2,$\mathbb{C}$) this decomposition leads to the following parametrization:

$$\mathrm{SL}(2,\mathbb{C}):\ \Lambda = e^{-i\varphi L_{12}} e^{-i\theta L_{31}} e^{-i\psi L_{12}} \times e^{isL_{03}} e^{-it[L_{01}+L_{31}]} e^{-iu[L_{02}-L_{23}]}, \tag{C14}$$

where L_{12}, L_{23} and L_{31} generate SU(2), the covering group of K, L_{03} generates A, and $L_{01}+L_{31}$,$L_{02}-L_{23}$ generate N. Thus φ, θ, and ψ are the parameters of K, s is the parameter of A and t,u are the parameters of N. The ranges of the parameters $(\varphi,\theta,\psi,s,t,u)$ are[3]

$$0\leqslant\varphi,\psi\leqslant 2\pi,\quad 0\leqslant\theta\leqslant\pi,\quad -\infty<s,t,u<\infty. \tag{C15}$$

We have also the following parametrization of SL(2,$\mathbb{C}$):

$$\Lambda = e^{-i\varphi L_{12}} e^{-i\theta L_{31}} e^{-i\tilde t[L_{01}+L_{31}]} \times e^{-i\tilde u[L_{02}-L_{23}]} e^{-i\psi L_{12}} e^{isL_{03}}, \tag{C16}$$

where[3]

$$\tilde t = e^s[t\cos\psi - u\sin\psi],\qquad \tilde u = e^s[u\cos\psi + t\sin\psi]. \tag{C17}$$

The case [3] is treated using the Cartan decomposition for SL(2,$\mathbb{C}$)

$$\mathrm{SL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\mathrm{SU}(2)\times\mathrm{SU}(2)\approx T_3\times S_3. \tag{C18}$$

This decomposition expresses SL(2,$\mathbb{C}$) as the product of a hyperboloid of two sheets in four dimensions

$$T_3 = \{q^\mu | q^\mu q_\mu = q_0^2 - q_1^2 - q_2^2 - q_3^2 = +1\}, \tag{C19}$$

and the three-sphere

$$S_3 = \left\{u^\mu \middle| \sum_{\mu=1}^{4}(u^\mu)^2 = 1\right\}. \tag{C20}$$

Notice that the generators of SU(2) are L_{12}, L_{23}, and L_{31}. The case [3'] is perhaps not so well understood in terms of global decompositions (see Refs. 3 and 28).

The calculation of SO(4,1) invariant measures on the coset spaces G/K in terms of the various parametrizations listed in Table I is a consequence of the following result.

Lemma: Let G be SO(4,1) and let $G/K \overset{\text{a.a}}{=} V_4'\times H/K$ be the decomposition of "almost all" elements of G/K into $H/K=\mathrm{SO}(3,1)/K$ and $V_4'=G/H$ considered above ($K\subset H$). Furthermore, suppose there exists an H invariant measure $d\tilde y$ on H/K, and let

$$d\xi = \frac{1}{|\xi^5|}\, d\xi^0\, d\xi^1\, d\xi^2\, d\xi^3 \tag{C21}$$

be the G left-invariant measure on V_4'. Then there exists a positive G invariant measure dX on G/K so that

$$\int_{G/K} f(X)dX = \int_{V_4'\times H/K} f(\xi,\tilde y)d\xi\, d\tilde y \tag{C22}$$

for all $f\in C^\infty(G)$. [Here $(\xi,\tilde y)=X\in G/K$ is the decomposition of an element $X\in G/K$, considered above, which is not in the set of measure zero.]

Proof: Let $X^{g(\bar b,\bar\Lambda)}\to X'=[\xi'(\xi,g),\tilde y'(\tilde y;\xi,g)]$ and define dX as $dX=d\xi\, d\tilde y$.

Now from Eq. (C6)

$$d\xi' = \det(A_g)d\xi = d\xi$$

because $g\in\mathrm{SO}(4,1)$, and from Eq. (C1)

$$d\tilde y' = \det(\tilde h\,[\Lambda\,(\bar b,\bar\Lambda b)\bar\Lambda\,])d\tilde y = d\tilde y$$

because $\tilde h\,[\Lambda\,(\bar b,\bar\Lambda b)\bar\Lambda\,]\in\mathrm{SO}(3,1)$ which is unity since $d\tilde y$ is H left invariant on H/K.

Thus, for $X'=(gX)$

$$\int_{G/K} f(X')dX = \int_{V_4'\times H/K} f(\xi',\tilde y')d\xi\, d\tilde y = \int_{(V_4'\times H/K)'} f(\xi',\tilde y')d\xi'\, d\tilde y' = \int_{(G/K)'} f(X')dX'. \tag{C23}$$

This proves the lemma. [Since $\mathscr{L}^2$-functions are defined only up to sets of measure zero the "almost all" equality is actually an equality in (C22) and (C23).]

From this result it follows that the invariant measure on G/K is determined by Eq. (C21) and the H invariant measure $d\tilde y$ on H/K. The $d\tilde y$'s, when they exist, have been explicitly determined for the various homogeneous spaces in Ref. 3. Combining their results with Eq. (C21) we have obtained column 5 of the table. Note that when no H invariant measure on H/K exists we have listed No for these cases.

[1]F. Lurçat, Physics **1**, 95 (1964).

[2]Since, at this level of the presentation, we do not want to enter into a discussion on the many-particle, i.e., Fock space aspect of these fields we treat them as "generalized wave functions" like in a one-particle quantum theory. Which aspects of the description presented in this paper require actually a "second quantization" yielding eventually an operator formalism is a difficult question to answer at the beginning and is left to a later investigation. (The general opinion of the author is that only part of the degrees of freedom described here in using homogeneous space techniques require a second quantization, others remain quasiclassical albeit coupled through nonlinear interactions to the geometry of the embedding space.) Thus we shall concentrate on the semiclassical aspects of these fields, i.e., treat them as c-number quantities.

[3]H. Bacry and A. Kihlberg, J. Math. Phys. **10**, 2132 (1969).

[4]D. Finkelstein, Phys. Rev. **100**, 924 (1955).

[5]C. P. Boyer and G. N. Fleming, J. Math. Phys. **15**, 1007 (1974).

[6]W. Drechsler, Ann. Inst. Henri Poincaré **37**, 155 (1982).

[7]W. Drechsler, "Poincaré Gauge Theory, Gravitation, and Transformation of Matter Fields," talk presented at the XII International Conference on Differential Geometric Methods in Physics, Clausthal 1983 (to be published in Fortschritte der Physik).

[8]Actually, only axial vector torsion can be induced in this way.

[9]The transformation character is linear only for the simplest case (lowest fiber dimension, $N=4$): The fiber being affine Minkowski space-time. Otherwise the transformation rule is a hybrid one (compare the discussion below and in Sec. III).

[10]Ch. Ehresmann, Colloque de Topologie, Bruxelles, p. 29, 1950.

[11]S. Helgason, *The Radon Transform* (Birkhäuser-Verlag, Basel, 1980), Chap. IV.

[12]S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239, 2247 (1969).

[13]A. Salam and J. Strathdee, Phys. Rev. **184**, 1750 (1969).

[14]Y. M. Cho, Phys. Rev. D **18**, 2810 (1978).

[15]W. Drechsler, Fortschr. Phys. **23**, 607 (1975); Found. Phys. **7**, 629 (1976).

[16]W. Drechsler and M. E. Mayer, "Fiber Bundle Techniques in Gauge Theories," *Lecture Notes in Physics* (Springer-Verlag, Heidelberg, 1977), Vol. 67.

[17]Neglecting electromagnetism we consider functions possessing a phase determined by the representation chosen for the de Sitter group. In later applications we shall add a U(1) fiber at each space-time point to include the electromagnetic interaction, i.e., we go over to a gauge theory with gauge group $G \otimes U(1)$ and a wave function with an additional U(1) phase factor.
[18]A. Lichnerowicz, *Géometrie des Groupes de Transformations* (Dunod, Paris, 1958).
[19]It is convenient to call in computations the argument of Φ in equations like (3.5) and similarly in those of Sec. IV [See Eq. (4.3) below] on both sides of these equations $(\xi'', \tilde{y}'')$ or $(\xi', \tilde{y}')$ and use Eqs. (2.12) and (2.13).
[20]S. Kobayashi, Annali di Mat. **43**, 119 (1957).
[21]W. Drechsler, J. Math. Phys. **18**, 1358 (1977).
[22]A. O. Barut and A. Bohm, Phys. Rev. **139**, 1107 (1965).
[23]A. Bohm, Phys. Rev. **145**, 1212 (1966).
[24]A. Bohm, Phys. Rev. **175**, 1767 (1968).
[25]A. Bohm, "The Dynamical Group of a Simple Particle Model," *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1967), Vol. IX B, p. 327.
[26]R. R. Aldinger, A. Bohm, W. Drechsler, S. R. Komy, M. Loewe, P. Magnollay, and N. Mukunda, Phys. Rev. D **28**, 3020 (1983).
[27]I. M. Gelfand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, London, 1966), Vol. 5, Chap. V.
[28]K. C. Hannabus, Proc. Cambridge Philos. Soc. **70**, 283 (1971).
[29]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[30]K. Nomizu, Amer. J. Math. **76**, 33 (1954).
[31]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 717 (1976).
[32]G. D. Mostow, in *Lie Theories and Their Applications, Queen's Papers in Pure and Applied Mathematics*, No. 48, edited by A. J. Coleman and P. Ribenboim.
[33]P. Moylan, J. Math. Phys. **24**, 2706 (1983).
[34]P. Moylan, MPI Preprint, MPI-PAE/PTh 27/84 (1984).

Classical and quantum symmetry groups of a free-fall particle

Toshihiro Iwai and See-Gew Rew
Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Kyoto 606, Japan

(Received 29 December 1983; accepted for publication 28 June 1984)

Symmetry of a free-fall particle is studied in quantum as well as classical mechanics. The quantum symmetry group is shown to be a central extension of the classical one. In the case of two degrees of freedom, the action of the quantum symmetry group is expressed in the form of integral transform as a unitary operator on the space of wave functions.

I. INTRODUCTION

Symmetry of certain mechanical systems has been highly studied in classical and quantum mechanics.[1] The Kepler problem (or the hydrogen atom in quantum mechanics) and the harmonic oscillator are celebrated examples.[2] In particular, the harmonic oscillator has been investigated thoroughly in quantum mechanics as well as in classical mechanics because of the simplicity of the Hamiltonian. The Hamiltonian is, of course, quadratic in x and p, a position vector and a momentum vector, respectively.

The dynamical system to be dealt with in this article has the Hamiltonian quadratic in p and linear in x, that is, a free-fall particle. This system seems not to have been noted in symmetry theory. In spite of its simplicity, the free-fall particle exhibits symmetry of high interest. In fact, the classical and quantum symmetry groups of this system are not isomorphic to each other. As will be shown in the succeeding sections, the quantum symmetry group is a central extension of the classical one. This is in marked contrast with the harmonic oscillator whose symmetry groups of classical and quantum systems are the same, both being the special unitary group $SU(n)$. It is worth mentioning here that the Galilei group is known to give rise to a central extension when represented in the space of wave functions for the free-particle Schrödinger equation.[3,4]

Section II deals with the classical symmetry group of the free-fall particle. The Lie algebra of the symmetry group and that of the associated first integrals under a Poisson bracket are shown not to be isomorphic to each other. Through quantization this fact makes the quantum symmetry group not isomorphic to the classical one.

Section III is concerned with the quantum symmetry group of the free-fall particle. First, integral transforms are treated to describe the symmetry of the quantum system by unitary operators on the space of wave functions. The integral kernels, however, are determined up to arbitrary phase factors. To fix the phase factors is the point of this article. After that, the quantum symmetry group manifests itself. For the sake of simplicity the system is assumed to be of dimension 2. The classical symmetry group is then broken up into the product of one-parameter subgroups. The corresponding one-parameter groups for the quantum system are fully represented in the space of wave functions as one-parameter unitary operators in terms of integral transform with definite integral kernel. Put together, these one-parameter groups give rise to the quantum symmetry group the action of which is expressed in the form of an integral transform.

Section IV contains remarks on the one-parameter unitary operator generated by the Hamiltonian and on an implication of the fact that the quantum symmetry group is a central extension of the classical symmetry group.

II. CLASSICAL SYMMETRY GROUP

A free-fall particle is a dynamical system defined on $\mathbb{R}^n \times \mathbb{R}^n$, the phase space, with the Hamiltonian

$$H = \tfrac{1}{2}\langle p,p\rangle + g\langle k,x\rangle, \tag{2.1}$$

where $(x,p)\in\mathbb{R}^n \times \mathbb{R}^n$, $\langle\ ,\ \rangle$ denotes the standard inner product, g is a positive constant, and k is a constant vector. The symplectic two-form ω is written in the form

$$\omega = \sum dp_j \wedge dx_j, \tag{2.2}$$

where (x_j,p_j) are the Cartesian coordinates.

Consider inhomogeneous linear transformations

$$\begin{pmatrix} x \\ p \end{pmatrix} \to \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + g\begin{pmatrix} u \\ v \end{pmatrix}, \tag{2.3}$$

where $A,B,C,$ and D are $n\times n$ real constant matrices, and u and v are constant vectors in $\mathbb{R}^n$. The transformations (2.3) are said to be symmetries of the free-fall particle when they leave the Hamiltonian (2.1) and the symplectic form (2.2) invariant. It is a matter of calculation to find a necessary and sufficient condition for (2.3) to be symmetries. We have, in fact,

$$\begin{aligned} &AA^T = I, \quad B^TA = A^TB, \quad C=0, \quad D=A, \\ &Ak=k, \quad v=-Bk, \quad \langle k,u\rangle = -\tfrac{1}{2}\langle Bk,Bk\rangle, \end{aligned} \tag{2.4}$$

where A^T denotes the transpose of A, and so on.

From (2.4), the matrix M defined to be BA^{-1} is symmetric. Since $B=MA$, $C=0$, $D=A$, and $v=-Mk$, we can describe the symmetries in terms of A, M, and u. Here, from (2.4), A must be an orthogonal matrix fixing the vector k, and u is subject to $\langle k,u\rangle = -\tfrac{1}{2}\langle Mk,Mk\rangle$. Thus the symmetries obtained have

$$\tfrac{1}{2}(n-1)(n-2) + \tfrac{1}{2}n(n+1) + (n-1) = n^2$$

parameters. It is easy to show that the symmetries form a group and can be expressed in the $(n+1)\times(n+1)$ matrix form

 0022-2488/85/010055-07$02.50

$$\begin{pmatrix} x \\ p \\ g \end{pmatrix} \mapsto \begin{pmatrix} A & MA & u \\ & A & -Mk \\ & & 1 \end{pmatrix} \begin{pmatrix} x \\ p \\ g \end{pmatrix} \tag{2.5}$$

(see Ref. 5).

Theorem 2.1: A free-fall particle admits a symmetry group of dimension n^2 [to be denoted by FF(n)], the action of which is expressed in the form (2.5) subject to the conditions

$$AA^T = I, \quad Ak = k, \quad M^T = M, \quad \langle k,u \rangle = -\tfrac{1}{2}\langle Mk,Mk \rangle. \tag{2.6}$$

We remark that if $k = 0$ then the Hamiltonian (2.1) becomes that for a free particle, and the second and the last conditions in (2.6) vanish, so that the matrices in (2.5) form a symmetry group for the free-particle which is larger than the Euclidean group E(n) viewed usually as a symmetry group for the free-particle. We point out further that for $A = I$, $M = tI$, and $u = -\frac{1}{2}t^2k$ the symmetry group FF(n) restricts to a one-parameter subgroup whose orbits are just the Hamiltonian flows of the equation of motion. This one-parameter subgroup is commutative with all the elements of FF(n). We can then get rid of this subgroup by imposing an additional condition tr $M = 0$ to (2.6). The identity component of the restricted symmetry group obtained will be referred to as $\mathrm{FF}^r(n)$, which is subject to the condition (2.6) plus tr $M = 0$ and det $A = 1$, and is of dimension $n^2 - 1$.

We now proceed to generating functions of the infinitesimal transformations by the symmetry group FF(n). We first note the matrix in (2.5) has the decomposition

$$\begin{pmatrix} A & MA & u \\ & A & -Mk \\ & & 1 \end{pmatrix} = \begin{pmatrix} I & M & u \\ & I & -Mk \\ & & 1 \end{pmatrix} \begin{pmatrix} A & & \\ & A & \\ & & 1 \end{pmatrix}. \tag{2.7}$$

The first and the second factors in the right-hand side of (2.7) form respective subgroups of FF(n). We start with the subgroup determined by M and u, the first factor in the right-hand side of (2.7). As is easily shown, the Lie algebra of this subgroup is formed by

$$\begin{pmatrix} 0 & N & w \\ & 0 & -Nk \\ & & 0 \end{pmatrix}, \tag{2.8}$$

with $N^T = N$ and $\langle k,w \rangle = 0$, which will be referred to as $X(N,w)$. On infinitesimalizing the action of the subgroup under consideration, we obtain an infinitesimal transformation $U(N,w)$, the action of $X(N,w)$, on $\mathbb{R}^n \times \mathbb{R}^n$

$$U(N,w) = \left\langle (Np + gw), \frac{\partial}{\partial x} \right\rangle - g \left\langle Nk, \frac{\partial}{\partial p} \right\rangle, \tag{2.9}$$

where $\partial/\partial x$ and $\partial/\partial p$ stand for the gradient operators. The $U(N,w)$ is an infinitesimal canonical transformation whose generating function is then to be obtained by

$$\frac{\partial F}{\partial x} = gNk, \quad \frac{\partial F}{\partial p} = Np + gw. \tag{2.10}$$

Calculation results in

$$F(N,w) = \tfrac{1}{2}\langle p,Np \rangle + g\langle p,w \rangle + g\langle x,Nk \rangle. \tag{2.11}$$

This is, of course, a first integral.

The generating functions associated with the subgroup determined by A, the second factor in the right-hand side of (2.7), are clearly angular momentums. Let R be an antisymmetric $n \times n$ matrix with $Rk = 0$. By $Y(R)$ and $V(R)$ we mean an element of the Lie algebra of the subgroup under consideration and the induced infinitesimal transformation on $\mathbb{R}^n \times \mathbb{R}^n$, respectively. Here $Y(R)$ and $V(R)$ are, respectively, of the form

$$Y(R) = \begin{pmatrix} R & & \\ & R & \\ & & 0 \end{pmatrix},$$

$$V(R) = \left\langle Rx, \frac{\partial}{\partial x} \right\rangle + \left\langle Rp, \frac{\partial}{\partial p} \right\rangle. \tag{2.12}$$

The angular momentums generating $V(R)$ then have the form

$$L(R) = \langle Rx,p \rangle, \quad \text{with } R^T = -R \text{ and } Rk = 0. \tag{2.13}$$

We are going into details of the Lie algebras relevant to the infinitesimal symmetry of the free-fall particle. The Lie algebra, denoted by ff(n), of the symmetry group FF(n) has the commutation relations

$$\begin{aligned} &[X(N,w),X(N',w')] = X(0, -[N,N']k), \\ &[Y(R),X(N,w)] = X([R,N],Rw), \\ &[Y(R),Y(R')] = Y([R,R']). \end{aligned} \tag{2.14}$$

To show this is an easy matter. Further, a straightforward calculation shows that the correspondences

$$X(N,w) \mapsto -U(N,w), \quad Y(R) \mapsto -V(R) \tag{2.15}$$

make up a Lie algebra isomorphism of the Lie algebra ff(n) to the Lie algebra of the infinitesimal canonical transformations on $\mathbb{R}^n \times \mathbb{R}^n$. We next focus our attention on the Lie algebra consisting of the generating functions F and L under a Poisson bracket $\{\ ,\ \}$. The Poisson brackets between F's and L's are calculated to give

$$\begin{aligned} &\{F(N,w),F(N,w')\} = F(0, -[N,N']k) \\ &\qquad + g^2(\langle Nk,w' \rangle - \langle N'k,w \rangle), \\ &\{L(R),F(N,w)\} = F([R,N],Rw), \\ &\{L(R),L(R')\} = L([R,R']). \end{aligned} \tag{2.16}$$

We see from (2.16) that the generating functions form a Lie algebra together with a constant function 1. Of course, the 1 commutes with all $F(N,w)$ and $L(R)$. The Lie algebra obtained is then a central extension,[6] denoted by eff(n), of the Lie algebra ff(n) given in (2.14). Thus we have the following theorem.

Theorem 2.2: Two Lie algebras are associated with the infinitesimal symmetry of a free-fall particle. One is the Lie algebra ff(n) of the symmetry group FF(n) given in Theorem 2.1, the other is the Lie algebra eff(n) formed by the generating functions and a constant. The latter is a central extension of the former.

In the next section we will concentrate on a study of the symmetry group for a quantum free-fall particle of two degrees of freedom ($n = 2$). We make in advance a detailed review of the classical restricted symmetry group $\mathrm{FF}^r(2)$ for a comparison between the classical and the quantum symmetry groups. We fix the vector k as $e_2 = (0,1)^T$. We will also set $e_1 = (0,1)^T$ in what follows. The condition $Ak = k$ in (2.6) with $\det A = 1$ then implies that $A = I$. Accordingly, the symmetry group FF^r(2) and its Lie algebra ff^r(2) have the relation

$$\exp\begin{pmatrix} 0 & 0 & t & s & r \\ & 0 & s & -t & 0 \\ & & 0 & 0 & -s \\ & & & 0 & t \\ & & & & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t & s & r \\ & 1 & s & -t & -\frac{1}{2}(t^2+s^2) \\ & & 1 & 0 & -s \\ & & & 1 & t \\ & & & & 1 \end{pmatrix}, \tag{2.17}$$

where r,s, and t are real numbers. We note here that $\det M = -(t^2+s^2) \neq 0$ unless $t = s = 0$. This fact will be a reason for concentration on the case $n = 2$.

Let X_λ, $\lambda = 1,2,3$, be a basis of the symmetry algebra ff^r(2) such that the matrix (2.17) takes the form $\exp(tX_1 + sX_2 + rX_3)$. A calculation then shows that (2.17) is broken up into the product

$$\exp(tX_1 + sX_2 + rX_3) = \exp tX_1 \exp sX_2 \exp(r + st)X_3. \tag{2.18}$$

The commutation relations among X_λ's come from (2.14) to be

$$[X_1,X_2] = -2X_3, \quad [X_1,X_3] = 0, \quad [X_2,X_3] = 0. \tag{2.19}$$

By F_λ, $\lambda = 1,2,3$, we mean the generating functions corresponding to X_λ. Equation (2.11) then reads

$$F_1 = \tfrac{1}{2}(p_1^2 - p_2^2) - gx_2, \quad F_2 = p_1p_2 + gx_1, \quad F_3 = gp_1. \tag{2.20}$$

Their commutation relations are known from (2.16) to be

$$\{F_1,F_2\} = -2F_3, \ \{F_1,F_3\} = 0, \ \{F_2,F_3\} = g^2, \tag{2.21}$$

which should be compared with (2.19). By eff^r(2) we denote the Lie algebra defined through (2.21), a central extension of ff^r(2).

III. QUANTUM SYMMETRY GROUP

To discuss symmetry for a quantum free-fall particle, we start with infinitesimal symmetry. According to the Schrödinger quantization procedure, $p_j = -i\partial/\partial x_j$, the classical observables $F(N,w)$ and $L(R)$ are quantized to be denoted by $\hat{F}(N,w)$ and $\hat{L}(R)$, respectively. The commutation relations are calculated by the use of $[x_j,p_k] = i\delta_{jk}$, etc., to give

$$\begin{aligned} [\hat{F}(N,w),\hat{F}(N',w')] &= i\hat{F}(0, -[N,N']k) \\ &\quad + ig^2(\langle Nk,w'\rangle - \langle N'k,w\rangle), \\ [\hat{L}(R),\hat{F}(N,w)] &= i\hat{F}([R,N],Rw), \\ [\hat{L}(R),\hat{L}(R')] &= i\hat{L}([R,R']). \end{aligned} \tag{3.1}$$

Thus we have obtained a Lie algebra describing infinitesimal symmetry of the quantum free-fall particle, which is isomorphic with the Lie algebra eff(n) given by (2.16) for the classical system.

In order to get a symmetry group for the quantum system, we follow Moshinsky and Quesne,[7] and Wolf.[8] Suppose we are given the transformation (2.5). If x and p are regarded as operators in the quantum system, Eq. (2.5) will be thought of as an inhomogeneous linear transformation of the operators x and p. However, from the viewpoint of group representation, we should consider that Eq. (2.5) is the action of the inverse of the matrix in (2.5) on the operators x and p. We assume that this transformation is induced by a unitary operator W in the space of wave functions, so that we have

$$W\begin{pmatrix} x \\ p \end{pmatrix}W^{-1} = \begin{pmatrix} A & MA \\ & A \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + g\begin{pmatrix} u \\ -Mk \end{pmatrix}. \tag{3.2}$$

It should be noted that Eq. (3.2) is invariant if W is replaced by αW, α being a complex number with $|\alpha| = 1$, so that Eq. (3.2) determines W up to a factor α. We now assume that W is given by the integral transform

$$Wf(x) = \int C(x,\xi)f(\xi)d\xi, \tag{3.3}$$

where $d\xi = d\xi_1 \cdots d\xi_n$. Writing out the identities $Wxf = WxW^{-1}Wf$ and $Wpf = WpW^{-1}Wf$ by using (3.2) and (3.3) under a suitable boundary condition at infinity, we obtain sufficient conditions for the kernel $C(x,\xi)$ to define the integral transform desired,

$$\begin{aligned} \xi C(x,\xi) &= \left(Ax + \frac{1}{i}MA\frac{\partial}{\partial x} + gu\right)C(x,\xi), \\ -\frac{1}{i}\frac{\partial}{\partial \xi}C(x,\xi) &= \left(\frac{1}{i}A\frac{\partial}{\partial x} - gMk\right)C(x,\xi). \end{aligned} \tag{3.4}$$

Equations (3.4) allow us to express $C(x,\xi)$ in the form

$$\begin{aligned} C(x,\xi) = c\exp[(i/2)(&\langle x,Qx\rangle + 2\langle x,S\xi\rangle \\ &+ \langle \xi,P\xi\rangle + 2\langle a,x\rangle + 2\langle b,\xi\rangle)], \end{aligned} \tag{3.5}$$

where c is a complex constant, P, Q, and S are real constant matrices, and a and b are real constant vectors. Integral kernels of this form were discussed in Refs. 9–11. Substitution of (3.5) into (3.4) yields

$$\begin{aligned} &(A + MAQ)x + (MAS - I)\xi + MAa + gu = 0, \\ &(AQ + S^T)x + (AS + P)\xi + Aa + b - gMk = 0. \end{aligned} \tag{3.6}$$

With the assumption that $\det M \neq 0$, we get

$$\begin{aligned} &Q = -A^{-1}M^{-1}A, \quad S = A^{-1}M^{-1}, \quad P = -M^{-1}, \\ &a = -gA^{-1}M^{-1}u, \quad b = gMk + gM^{-1}u. \end{aligned} \tag{3.7}$$

Then, after calculation, $C(x,\xi)$ takes the form

$$\begin{aligned} c\exp[-(i/2)&\langle(\xi - Ax, M^{-1}(\xi - Ax)\rangle \\ &+ ig\langle \xi - Ax, M^{-1}u\rangle + ig\langle \xi, Mk\rangle]. \end{aligned} \tag{3.8}$$

The constant c is determined, up to a constant factor of absolute value 1, by the unitary condition

$$\int C(\eta,x)^* C(\eta,\xi)d\eta = \delta(x - \xi), \tag{3.9}$$

where the superscript asterisk (*) indicates the complex conjugate. If we introduce $\bar{\eta} = A\eta$ and $\bar{C}(\bar{\eta},\xi) = C(\eta,\xi)$, Eq.

(3.9) is put, on account of $\det A = 1$, into

$$\int \bar{C}(\bar{\eta},x)^* \bar{C}(\bar{\eta},\xi)d\bar{\eta} = \delta(x-\xi). \tag{3.10}$$

It follows from (3.8) and (3.10) that

$$c = \alpha(2\pi)^{-n/2}|\det M|^{-1/2}, \quad \text{with } |\alpha| = 1. \tag{3.11}$$

Thus we have found the kernel $C(x,\xi)$ within a constant factor α of absolute value 1. In the literature the constant α is undetermined or set equal to 1, and hence the transform (3.3) is a ray representation. Wolf[8] gained insight into the constant α in the case of $n = 1$. We will soon consider how to determine it.

However, we notice that the integral transform with α undetermined allows of the inversion formula on the space of rapidly decreasing functions,

$$f(x) = \int d\eta\, C(\eta,x)^* \int C(\eta,\xi)f(\xi)d\xi. \tag{3.12}$$

This can be verified by the use of the Fourier integral theorem.

We next consider the case where $M = 0$ and $u = 0$. Then Eq. (3.2) turns into

$$T\begin{pmatrix} x \\ p \end{pmatrix}T^{-1} = \begin{pmatrix} A & \\ & A \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix}. \tag{3.13}$$

Here we have used T in place of W. Clearly, Eq. (3.13) is satisfied by the unitary operator T defined by

$$Tf(x) = f(Ax). \tag{3.14}$$

Let W_0 be the integral transform (3.3) with $A = I$, so that the kernel is denoted by $\bar{C}(x,\xi)$. Then we have a decomposition of W, $W = TW_0$, or

$$Wf(x) = \int \bar{C}(Ax,\xi)f(\xi)d\xi. \tag{3.15}$$

The decomposition $W = TW_0$ corresponds to the inverse matrix relation of (2.7).

To study a symmetry group for the quantum free-fall particle, we have to determine the constant α in (3.11) in terms of the parameters A, M, and u. To simplify computation, we restrict ourselves to the case of $n = 2$. Then, for $\mathrm{FF}^r(2)$, $\det M$ does not vanish unless $s = t = 0$, as was already pointed out, so that we can get the integral kernel $C(x,\xi)$ up to α through the calculations (3.4)–(3.11) for $n = 2$.

Since Eq. (3.2) can determine the unitary operator W only within a complex number α, we need another condition for α. A reasonable condition is that W should tend to the identity as the transformation in the right-hand side of (3.2) tends to the identity, though in what way we bring the transformation to the identity is still an open problem. However, for a one-parameter group of transformations we have no problem, and therefore we can actually determine α under the condition just stated. Because of this fact, we see that if one can decompose any element of a group G into a product of one-parameter subgroups, then one can obtain a unitary representation of G or of its extension by putting together unitary representations of the one-parameter subgroups of G. For $\mathrm{FF}^r(2)$ we can carry out this recipe, as $\mathrm{FF}^r(2)$ has the decomposition (2.18).

Let $\hat{F}_\lambda$, $\lambda = 1,2,3$, be quantum observables formed from (2.20). Their commutation relations come from (3.1) to be

$$[\hat{F}_1,\hat{F}_2] = -2i\hat{F}_3, \quad [\hat{F}_1,\hat{F}_3] = 0, \quad [\hat{F}_2,\hat{F}_3] = ig^2, \tag{3.16}$$

which describe the infinitesimal symmetry $\mathrm{eff}^r(2)$ of the quantum free-fall particle for $n = 2$. Exponentiating this Lie algebra will yield the symmetry group $\mathrm{EFF}^r(2)$ for the quantum system.

We denote by $W_\lambda(t)$ the one-parameter group of unitary operators corresponding to $\exp tX_\lambda$. It is plausible that the infinitesimal operator of $W_\lambda(t)$ is $i\hat{F}_\lambda$, since $\exp it\hat{F}_\lambda$ (if it exists) induces the same action on the operators x and p as that of $W_\lambda(t)$ which is given by (3.2) with entries corresponding to $\exp tX_\lambda$. In fact, we can prove the relation

$$\exp it\hat{F}_\lambda\begin{pmatrix} x \\ p \end{pmatrix}\exp(-it\hat{F}_\lambda) = \begin{pmatrix} I & tM_\lambda \\ & I \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + g\begin{pmatrix} u_\lambda(t) \\ -M_\lambda e_2 \end{pmatrix} \tag{3.17}$$

by showing that the both sides of (3.17) satisfy the same differential equation in t, where

$$M_1 = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \quad M_3 = 0,$$

$$u_1(t) = u_2(t) = -\tfrac{1}{2}t^2 e_2, \quad u_3(t) = te_1. \tag{3.18}$$

We start with $W_1(t)$. The one-parameter subgroup $\exp tX_1$ of the classical symmetry group $\mathrm{FF}^r(2)$ is given by setting $A = I$, $M = tM_1$, and $u = -\frac{1}{2}t^2 e_2$ [see (2.17) and (2.18) with $r = s = 0$]. Since $\det tM_1 \neq 0$ for $t \neq 0$, Eq. (3.8) gives

$$C_1(t;x,\xi) = c_1(t)\exp[-(i/2)\langle \xi - x,(1/t)M_1(\xi - x)\rangle - i(tg/2)\langle \xi + x,e_2\rangle], \tag{3.19}$$

where we have denoted the integral kernel and the undetermined factor by $C_1(t;x,\xi)$ and $c_1(t)$, respectively, in order to indicate explicitly their dependence on the parameter t.

We turn to the factor $c_1(t)$. Since the infinitesimal generator of $W_1(t)$ is supposed to be $i\hat{F}_1$, the kernel $C_1(t;x,\xi)$ is required, as a function of t and x, to satisfy

$$\frac{\partial}{\partial t}C_1(t;x,\xi) = i\hat{F}_1 C_1(t;x,\xi). \tag{3.20}$$

Working out the both sides of (3.20) yields the differential equation for $c_1(t)$

$$\frac{1}{c_1(t)}\frac{dc_1(t)}{dt} = -\frac{ig^2}{8}t^2 - \frac{1}{t}. \tag{3.21}$$

An easy calculation gives with an integration constant γ

$$c_1(t) = \gamma|t|^{-1}\exp(-ig^2t^3/24). \tag{3.22}$$

The constant γ is to be determined under the condition that

$$C_1(t;x,\xi) \to \delta(x-\xi) \quad \text{as } t \to 0, \tag{3.23}$$

because $W_1(0)$ must be the identity. For this purpose it suffices that

$$\int C_1(t;x,\xi)d\xi \to 1 \quad \text{as } t \to 0. \tag{3.24}$$

After a calculation we obtain

$$\int C_1(t;x,\xi)d\xi = \gamma 2\pi \exp\left[\frac{-ig^2t^3}{6} - igt\langle x,e_2\rangle\right]. \tag{3.25}$$

Thus we have $\gamma = (2\pi)^{-1}$, therefore

$$c_1(t) = (2\pi|t|)^{-1} \exp(-ig^2 t^3/24). \tag{3.26}$$

As a result, from (3.26) and (3.11) with $n = 2$ and $M = tM_1$, we know that α is determined to be $\exp(-ig^2 t^3/24)$.

The integral transform with the kernel $C_1(t;x,\xi)$ given by (3.19) and (3.26) is an isometry on the space of rapidly decreasing functions and can be extended to a unitary transformation on $L^2(\mathbb{R}^2)$. This is quite analogous to the Plancherel theorem on the Fourier transform.

Theorem 3.1: The one-parameter group of unitary operators $W_1(t) = \exp it\hat{F}_1$ is expressed in the form of the integral transform whose kernel is given by (3.19) and (3.26).

We proceed to $W_2(s)$. As in the case of $W_1(t)$, the integral kernel for $W_2(s)$ is obtained, up to a factor $c_2(s)$, from (3.8) by setting $A = I$, $M = sM_2$, and $u = -\frac{1}{2}s^2 e_2$,

$$C_2(s;x,\xi) = c_2(s)\exp[-(i/2)\langle \xi - x,(1/s)M_2(\xi - x)\rangle + i(gs/2)\langle \xi + x,e_1\rangle]. \tag{3.27}$$

The factor $c_2(s)$ is determined under the conditions

$$\frac{\partial}{\partial s} C_2(s;x,\xi) = i\hat{F}_2\, C_2(s;x,\xi), \tag{3.28}$$

$$C_2(s;x,\xi) \to \delta(x - \xi) \quad \text{as } s \to 0. \tag{3.29}$$

We here give the result without writing down the calculation for $c_2(s)$,

$$c_2(s) = (2\pi|s|)^{-1}. \tag{3.30}$$

Thus we have, like Theorem 3.1, the following theorem.

Theorem 3.2: The one-parameter group of unitary operators $W_2(s) = \exp is\hat{F}_2$ is given by the integral transform that has the kernel (3.27) with (3.30).

Now we are left with $W_3(r)$. Since the one-parameter subgroup $\exp rX_3$ acts as a translation, we have the following theorem.

Theorem 3.3: The operator $W_3(r) = \exp ir\hat{F}_3$ is manifestly expressed in the form

$$W_3(r) f(x) = f(x + rge_1). \tag{3.31}$$

So far we have obtained the one-parameter groups of unitary operators $W_\lambda(t_\lambda) = \exp it_\lambda \hat{F}_\lambda$, $\lambda = 1,2,3$, which correspond to the classical one-parameter symmetry groups $\exp t_\lambda X_\lambda$, where $(t_\lambda) = (t,s,r)$. We now make the attempt to get a quantum symmetry group by putting $\exp it_\lambda \hat{F}_\lambda$ together. We recall here that the quantum symmetry algebra eff$^r(2)$ formed by $\hat{F}_\lambda$, $\lambda = 1,2,3$, and a constant is a central extension of the classical one ff$^r(2)$ formed by X_λ, $\lambda = 1,2,3$ [compare (2.19) and (3.16)]. We may therefore deduce that the quantum symmetry group EFF$^r(2)$ is a central extension of the classical one FF$^r(2)$. To think of the extension, we refer to the Baker–Campbell–Hausdorff formula,[12]

$$\exp F \exp H = \exp(F + H + \tfrac{1}{2}[F,H] + \tfrac{1}{12}[F,[F,H]] + \tfrac{1}{12}[[F,H],H] + \cdots). \tag{3.32}$$

In contrast with the decomposition (2.18) for the classical symmetry group, we have by using (3.32) the decomposition

$$\exp it\hat{F}_1 \exp is\hat{F}_2 \exp i(r - st)\hat{F}_3 = \exp ig^2(-\tfrac{1}{2}sr + \tfrac{2}{3}s^2 t)\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3). \tag{3.33}$$

This is to be compared with (2.18) in which t, s, and r are replaced by $-t$, $-s$, and $-r$, respectively. The factor $\exp ig^2(-\frac{1}{2}sr + \frac{2}{3}s^2 t)$ is in the group U(1) and commutes with any unitary operators. Thus the quantum symmetry group will lead up to a central extension of the classical symmetry group by U(1).

Equation (3.33) will define $\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$, since the left-hand side of (3.33) is defined by successive applications of $\exp i(r - st)\hat{F}_3$, $\exp is\hat{F}_2$, and $\exp it\hat{F}_1$. We start by composing two of $\exp it_\lambda \hat{F}_\lambda$, $\lambda = 1,2,3$. The following lemma is applied for obtaining the integral kernel for the composition of two integral transforms.

Lemma 3.4: For a nonsingular real symmetric matrix B one has

$$\int \exp\Big(\frac{i}{2}\langle \xi, B\xi\rangle + i\langle \xi,\eta\rangle\Big)d\xi = (2\pi)^{n/2}\,|\det B|^{-1/2} \exp(i\pi \operatorname{sgn} B/4) \times \exp(-(i/2)\langle \eta, B^{-1}\eta\rangle), \tag{3.34}$$

where sgn B is the number of positive eigenvalues minus the number of negative eigenvalues of B.

By successive applications of $\exp is\hat{F}_2$ and $\exp it\hat{F}_1$ to a rapidly decreasing function f and by exchanging the order of integration, we see that the integral kernel for $\exp it\hat{F}_1 \times \exp is\hat{F}_2$ is given by

$$C(s,t;x,\zeta) = \int C_1(t;x,\xi) C_2(s;\xi,\zeta)d\xi. \tag{3.35}$$

By writing out the integrand in the right-hand side of (3.35) by the use of (3.34) with $B = -(M_1/t + M_2/s)$ and $\eta = M_1 x/t + M_2\zeta/s + \frac{1}{2}g(se_1 - te_2)$, we obtain

$$C(s,t;x,\zeta) = c(s,t)\exp\Big[-\frac{i}{2}\Big\langle \zeta - x,\Big(\frac{t}{s^2+t^2}M_1 + \frac{s}{s^2+t^2}M_2\Big)(\zeta - x)\Big\rangle - ig\Big\langle x,\Big(\frac{st^2}{s^2+t^2} - \frac{1}{2}s\Big)e_1 + \Big(\frac{s^2 t}{s^2+t^2} + \frac{1}{2}t\Big)e_2\Big\rangle + ig\Big\langle \zeta,\Big(\frac{st^2}{s^2+t^2} + \frac{1}{2}s\Big)e_1 + \Big(\frac{s^2 t}{s^2+t^2} - \frac{1}{2}t\Big)e_2\Big\rangle\Big], \tag{3.36}$$

with

$$c(s,t) = (2\pi)^{-1}(s^2+t^2)^{-1/2} \times \exp\Big(i\frac{g^2}{24}\Big(-t^3 + \frac{3s^2 t(s^2 - 3t^2)}{s^2+t^2}\Big)\Big).$$

According to Eq. (3.33) with $r = st$, we can define $\exp i(t\hat{F}_1 + s\hat{F}_2 + st\hat{F}_3)$ to be

$$\exp i(t\hat{F}_1 + s\hat{F}_2 + st\hat{F}_3) f(x) = \exp(-ig^2 s^2 t/6)\exp it\hat{F}_1 \exp is\hat{F}_2 = \exp\Big(\frac{-ig^2 s^2 t}{6}\Big)\int C(s,t;x,\xi) f(\xi)d\xi, \tag{3.37}$$

where the kernel is given by (3.36). On the other hand, Eqs. (3.8) and (3.11) with $A = I$, $M = tM_1 + sM_2$, and $u = ste_1 - \frac{1}{2}(t^2 + s^2)e_2$ also give (3.36) without the exponential factor in $c(s,t)$. Thus we have checked the consistency that $\exp i(t\hat{F}_1 + s\hat{F}_2 + st\hat{F}_3)$ is expressed in the form of the integral transform which corresponds to $\exp(tX_1 + sX_2 + stX_3)$.

The remaining operators $\exp i(t\hat{F}_1 + r\hat{F}_3)$ and $\exp i(s\hat{F}_2 + r\hat{F}_3)$ are easy to define. For example, we pick up $\exp i(s\hat{F}_2 + r\hat{F}_3)$. By Theorems 3.2 and 3.3 and by introducing new variables $\bar{\xi} = \xi + gre_1$, we have

$$\exp is\hat{F}_2 \exp ir\hat{F}_3 f(x) = \int C_2(s;x,\bar{\xi} - gre_1) f(\bar{\xi}) d\bar{\xi}. \tag{3.38}$$

Therefore we can define $\exp i(s\hat{F}_2 + r\hat{F}_3)$ to be

$$\begin{aligned}\exp i(s\hat{F}_2 + r\hat{F}_3) f(x) &= \exp(ig^2 sr/2) \exp is\hat{F}_2 \exp ir\hat{F}_3 f(x)\\ &= \exp\left(\frac{ig^2 sr}{2}\right) \int C_2(s;x,\bar{\xi} - gre_1) f(\bar{\xi}) d\bar{\xi}.\end{aligned} \tag{3.39}$$

Written out, the kernel $C_2(s;x,\bar{\xi} - gre_1)$ is shown to be also obtained from (3.8) and (3.11) with $A = I$, $M = sM_2$, and $u = re_1 - \frac{1}{2}gs^2 e_2$ within a factor of absolute value 1. Therefore, the $\exp i(s\hat{F}_2 + r\hat{F}_3)$ defined corresponds to $\exp(sX_2 + rX_3)$ actually. The same method is applicable for defining $\exp i(t\hat{F}_1 + r\hat{F}_3)$.

We are now in a position to define $\exp i(\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$ according to (3.33). Since the operator $\exp i(r - st)\hat{F}_3$ means a translation of the independent variables and since the operator $\exp it\hat{F}_1 \exp is\hat{F}_2$ is expressed in the form of the integral transform with the kernel (3.36), we can define $\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$ from (3.33) to be

$$\begin{aligned}\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3) f(x) &= \exp(ig^2(\tfrac{1}{2}sr - \tfrac{2}{3}s^2 t))\\ &\quad\times \int C(s,t;x,\bar{\xi} - g(r - st)e_1) f(\bar{\xi}) d\bar{\xi}.\end{aligned} \tag{3.40}$$

Writing out the right-hand side of (3.40) yields the integral kernel for $\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$,

$$\begin{aligned}&c(r,s,t)\exp\Big[-\frac{i}{2}\Big\langle \bar{\xi} - x, \frac{1}{s^2 + t^2}(tM_1 + sM_2)(\bar{\xi} - x)\Big\rangle\\ &\quad - ig\Big\langle x, \Big(\frac{rt}{s^2 + t^2} - \frac{1}{2}s\Big)e_1 + \Big(\frac{rs}{s^2 + t^2} + \frac{1}{2}t\Big)e_2\Big\rangle\\ &\quad + ig\Big\langle \bar{\xi}, \Big(\frac{rt}{s^2 + t^2} + \frac{1}{2}s\Big)e_1 + \Big(\frac{rs}{s^2 + t^2} - \frac{1}{2}t\Big)e\Big\rangle\Big],\end{aligned} \tag{3.41}$$

with

$$\begin{aligned}c(r,s,t) &= \frac{1}{2\pi(s^2 + t^2)^{1/2}} \exp\Big(ig^2\Big(\frac{1}{2}sr - \frac{2}{3}s^2 t - \frac{1}{24}t^3\\ &\quad + \frac{s^2 t(s^2 - 3t^2)}{8(s^2 + t^2)} - \frac{(r - st)(rt + 2st^2 + s^3)}{2(s^2 + t^2)}\Big)\Big).\end{aligned}$$

We can easily verify that Eq. (3.41) is derived, up to a factor of absolute value 1, also from (3.8) and (3.11) with $A = I$, $M = tM_1 + sM_2$, and $u = re_1 - \frac{1}{2}(s^2 + t^2)e_2$, so that Eq. (3.40) with (3.41) is shown to define $\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$ corresponding to $\exp(tX_1 + sX_2 + rX_3)$. We remark here that the exponential factor appearing in $c(r,s,t)$ is what we wanted to know, which cannot be determined by Eqs. (3.2) and (3.9) only.

Along with the commutation relations (3.16) the Baker–Campbell–Hausdorff formula gives rise to the composition law for $\exp i\Sigma t_\lambda \hat{F}_\lambda$, from which we can see that the quantum symmetry group is a central extension of the classical symmetry group. Thus we have the following theorem.

Theorem 3.5: For a free-fall particle of dimension 2, the quantum symmetry group corresponding to the classical symmetry group $\exp(tX_1 + sX_2 + rX_3)$ given by (2.17) is a central extension of the latter. Its action on the space of wave functions $L^2(\mathbf{R}^2)$ is described by the unitary operator $e^{i\theta} \exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$, where $e^{i\theta}$ is a multiplication operator and $\exp i(t\hat{F}_1 + s\hat{F}_2 + r\hat{F}_3)$ is expressed in the form of the integral transform with the kernel given by (3.41).

IV. CONCLUDING REMARKS

In the preceding section we have not discussed the unitary operator generated by the Hamiltonian

$$\hat{H} = \tfrac{1}{2}(p_1^2 + p_2^2) + gx_2. \tag{4.1}$$

However, to accomplish the quantum symmetry theory for a free-fall particle, we are to discuss the unitary operator $\exp it\hat{H}$. Since the classical Hamiltonian flows are given by (2.5) with $A = I$, $M = tI$, and $u = -\frac{1}{2}t^2 e_2$ $(k = e_2)$, the integral kernel $K(t,x,\xi)$ for $\exp it\hat{H}$ is obtained, up to a factor $k(t)$, from (3.8),

$$\begin{aligned}K(t;x,\xi) = k(t)\exp[&-(i/2)\langle \xi - x, (1/t)(\xi - x)\rangle\\ &+ i(gt/2)\langle \xi + x, e_2\rangle].\end{aligned} \tag{4.2}$$

The factor $k(t)$ is determined under the conditions

$$\frac{\partial}{\partial t}K(t;x,\xi) = i\hat{H}K(t;x,\xi), \tag{4.3}$$

$$K(t;x,\xi) \to \delta(x - \xi) \quad \text{as } t \to 0. \tag{4.4}$$

Equation (4.3) gives a differential equation for $k(t)$ and Eq. (4.4) is used to evaluate the integration constant for $k(t)$. After a straightforward calculation we obtain

$$k(t) = (2\pi|t|)^{-1} \exp(i\pi \operatorname{sgn} t/2) \exp(ig^2 t^3/24), \tag{4.5}$$

where $\operatorname{sgn} t$ denotes the signum of t. Thus we have obtained $\exp it\hat{H}$ in the form of integral transform. The operator $\exp(-it\hat{H})$ gives the solutions to the Schrödinger equation for the free-fall particle.

In conclusion we make a mention of an implication of $[X_2, X_3] = 0$ and $[\hat{F}_2, \hat{F}_2] = ig^2$. In contrast with the commutator

$$\exp(-sX_2)\exp(-rX_3)\exp sX_2 \exp rX_3 = id \tag{4.6}$$

for the classical system, we have for the quantum system the commutator

$$\exp(-is\hat{F}_2)\exp(-ir\hat{F}_3)\exp s\hat{F}_2 \exp r\hat{F}_3 = e^{-ig^2 sr}. \tag{4.7}$$

Hence, if a particle in the classical system is subject to the transformation (4.6), it comes back to the original state. However, its quantum state becomes, according to (4.7),

$$\psi \to e^{-ig^2 sr}\psi. \tag{4.8}$$

The gauge transformation (4.8), of course, does not affect the quantum system. We conclude the remark by saying that for the Galilei group the analogous equations to (4.6) and (4.7) lead to a superselection rule for mass.

ACKNOWLEDGMENTS

The authors would like to offer their cordial thanks to the late Professor Mineo Ikeda at Kyoto University for his

having guided them to the symmetry theory, and to dedicate this paper to his memory.

[1]R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.
[2]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966); T. Iwai, J. Math. Phys. **23**, 1088 (1982); **23**, 1093 (1982); J. Milnor, Am. Math. Mon. **90**, 353 (1983), and references therein.
[3]V. Bargmann, Ann. Math. **59**, 1 (1954).
[4]J-M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963); *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. 2, pp. 221–229.
[5]Missing matrix entries are all zero.
[6]D. J. Simms and N.M.J. Woodhouse, "Lectures on Geometric Quantization," in *Lecture Notes in Physics*, Vol. 53 (Springer, New York, 1976), p. 154.
[7]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[8]K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979).
[9]M. H. Boon and T. H. Seligman, J. Math. Phys. **14**, 1224 (1973).
[10]V. V. Dodonov, I. A. Malkin, and V. I. Man'ko, Int. J. Theor. Phys. **14**, 37 (1975).
[11]G. Burdet, M. Perrin, and M. Perroud, Commun. Math. Phys. **58**, 241 (1978).
[12]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 170.

Factors of the Fock functional

Gerhard C. Hegerfeldt
Institut für Theoretische Physik, Universität Göttingen, Göttingen, Federal Republic of Germany

(Received 9 April 1984; accepted for publication 25 May 1984)

Can the fields ϕ and π of a representation of the CCR's be written as $\phi = 1/\sqrt{2}\{\phi_1 \times \mathbf{1} + \mathbf{1} \otimes \phi_2\}$ and similarly for π, such that ϕ_i and π_i satisfy the CCR's? What are the possible ϕ_i's and π_i's? This is equivalent to a factorization of the corresponding generating functionals (scaled by $1/\sqrt{2}$). Generalizing this question somewhat we show a noncommutative analog of Cramér's theorem of probability theory. If ϕ and π are Fock fields then so are ϕ_i, π_i, $i = 1,2$; similarly for quasifree representations of the CCR's. As an application we show that the fields of a representation of the CCR's whose generating functional differs from a Fock functional by a phase factor only are just shifted Fock fields.

I. INTRODUCTION AND MAIN RESULT

A representation of the canonical commutation relations[1] (CCR's) with cyclic vector Ω can be characterized by its generating or expectation functional

$$\begin{aligned} E(f,g) &= \langle \Omega, \exp i\{\phi(f) + \pi(g)\}\Omega \rangle \\ &\equiv \langle \Omega, U(f,g)\Omega \rangle, \quad f,g \in \mathscr{V}, \end{aligned} \tag{1.1}$$

where, without loss of generality,[2] $\mathscr{V}$ can be taken as a real pre-Hilbert space with scalar product $(\cdot,\cdot)$. In ordinary quantum mechanics $\mathscr{V} = R^n$, while in quantum field theory one usually has $\mathscr{V} \subset L^2(R^3)$. The fields ϕ and π satisfy on a suitable domain[3]

$$[\phi(f), \pi(g)] = i(f,g) \tag{1.2}$$

and zero otherwise. Here Ω need not be in the domain of the field operators.

The generating functional for a general Fock representation can be derived from the standard functional

$$E_F(f,g) = \exp\{-\|f\|^2/4 - \|g\|^2/4\} \tag{1.3}$$

by a Bogoliubov transformation.[4] This is an invertible map on $\mathscr{V} \oplus \mathscr{V}$ giving rise to

$$\begin{aligned} \phi'(f) &= \phi(\alpha f) + \pi(\beta f), \\ \pi'(g) &= \phi(\gamma g) + \pi(\delta g), \end{aligned} \tag{1.4}$$

which leaves the CCR's invariant. See also Eqs. (2.5) and (2.6) below.

The Fock representation belonging to E_F in Eq. (1.3) is characterized by the existence of annihilation operators:

$$\{\phi_F(f) + i\pi_F(f)\}\Omega = 0, \quad f \in \mathscr{V}. \tag{1.5}$$

Recently, Ruijsenaars[5] raised the following question. What is the form of a generating functional $E(f,g)$ which differs from a Fock functional by a phase factor $\exp\{i\lambda_{f,g}\}$ only? He conjectured that the corresponding fields ϕ and π were just Fock fields, possibly *shifted* by c-numbers (λ_1, f) and (λ_2, g), respectively, where λ_1, λ_2 are linear functionals on $\mathscr{V}$. This would result in a "shifted Fock functional" with phase factor $\exp i\{(\lambda_1, f) + (\lambda_2, g)\}$.

This question, which will be answered in the affirmative, is a special case of the following. It is well known that the tensor product of two representations of a group is again a representation, where the corresponding generating functionals multiply. This can be carried over to the CCR's; however, in order to retain the commutation relations one needs some form of renormalization. We note the following simple fact; here ^{t} denotes transpose, $(f,\alpha g) = (\alpha^t f, g)$.

Lemma: Let E_1 and E_2 be generating functionals for representations of the CCR's with fields ϕ_i, π_i and cyclic vectors Ω_1, Ω_2. Let $\alpha_i, \beta_i, \gamma_i, \delta_i$, $i = 1,2$, be operators on $\mathscr{V}$. Then

$$\begin{aligned} \phi(f) :&= \{\phi_1(\alpha_1 f) + \pi_1(\beta_1 f)\} \otimes \mathbf{1} \\ &\quad + \mathbf{1} \otimes \{\phi_2(\alpha_2 f) + \pi_2(\beta_2 f)\}, \\ \pi(g) :&= \{\phi_1(\gamma_1 g) + \pi_1(\delta_1 g)\} \otimes \mathbf{1} \\ &\quad + \mathbf{1} \otimes \{\phi_2(\gamma_2 g) + \pi_2(\delta_2 g)\} \end{aligned} \tag{1.6}$$

is a representation of the CCR's and the curly brackets do not vanish for $f, g \neq 0$ iff

$$\sum_{i=1}^{2} \begin{pmatrix} \alpha_i^t & \gamma_i^t \\ \beta_i^t & \delta_i^t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{1.7a}$$

and

$$\begin{pmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{pmatrix}^{-1} \quad \text{exists}, \quad i = 1,2. \tag{1.7b}$$

The generating functional for the new representation—restricted to the subspace generated by the Weyl operators $U(f,g)$ applied to $\Omega_1 \otimes \Omega_2$—is given by

$$E(f,g) = \prod_{i=1}^{2} E_i(\alpha_i f + \gamma_i g, \beta_i f + \delta_i g). \tag{1.8}$$

Conversely, let $E, E_1, E_2, \alpha_i, \ldots, \delta_i$ be given such that Eqs. (1.7) and (1.8) hold. Then ϕ, π can be written in the form of Eq. (1.6).

The proof is evident. Equation (1.7a) is most easily seen in the notation of Eq. (2.8) below. Similarly for Eq. (1.7b), where the CCR's imply that $\phi_i(f) + \pi_i(g) \neq 0$ unless $f = g = 0$.

Definition: A generating functional $E_i(f,g)$ appearing in a decomposition of the above form is called a factor of $E(f,g)$.

Definition: A cyclic representation of the CCR's is quasifree[6] if the generating functional is an exponential of bilinear and linear forms in f and g ("quasifree functional").

Theorem: Any factor of a Fock functional is a (shifted) Fock functional. Any factor of a quasifree generating functional is quasifree.

As a corollary we can answer the question of Ruijsenaars.

Corollary: If a generating functional $E(f,g)$ differs only by a phase factor from a Fock functional, it is just a shift of this Fock functional.

Proof: By Eq. (1.1), $\overline{E(f,g)}$ is also a generating functional, being obtained by $f \to -f$, $g \to -g$. Then

$$E(f/\sqrt{2},g/\sqrt{2})\,\overline{E(f/\sqrt{2},g/\sqrt{2})}$$

is the Fock functional, and the Theorem applies. Q. E. D.

The above theorem will be proved in full generality in Sec. II. But since the generality obscures the extraordinary simplicity of the underlying idea we will consider here a special Fock case, namely

$$E_F(f,g) = E_1(f/\sqrt{2},g/\sqrt{2})E_2(f/\sqrt{2},g/\sqrt{2}).$$

Then ϕ_F, π_F can be decomposed as in Eq. (1.6). It is easy to see that Ω_1 and Ω_2 are in the domain of the corresponding field operators (see Lemma 2.1 below). Hence, by a shift, one can assume their one-point functions to vanish. The annihilation property, Eq. (1.5), now reads

$$\{\phi_1(f) + i\pi_1(f)\}\Omega_1 \otimes \Omega_2 + \Omega_1 \otimes \{\phi_2(f) + \pi_2(f)\}\Omega_2 = 0.$$

Squaring this, the mixed terms vanish and we get

$$\|\{\phi_j(f) + i\pi_j(f)\}\Omega_j\|^2 = 0, \quad j = 1,2.$$

So E_1, E_2 have annihilation operators and are thus Fock functionals!

Remark 1: The theorem can be viewed as a noncommutative analog of Cramér's Theorem[7] in probability theory. This states that if a normal random variable ξ is the sum of independent random variables ξ_1 and ξ_2 then ξ_1 and ξ_2 are also normal (or degenerate). Equivalently, if a Gaussian positive-definite function is a product of two positive-definite functions, then the latter are also Gaussian.

Remark 2: The theorem can be carried over to completely general fields, without CCR's and without space-time symmetries. We call a field *Gaussian* if all its truncated n-point functions vanish for $n > 2$. It will be shown elsewhere[8] that if A is Gaussian and

$$A = A_1 \otimes \mathbb{1} + \mathbb{1} \otimes A_2,$$

then both A_1 and A_2 are Gaussian, too.

Remark 3: For non-Gaussian fields a general decomposition into prime fields and infinitely divisible fields was obtained by the author some time ago.[9]

II. PROOF OF THE THEOREM

We first show that Ω_1 and Ω_2 are in the domain of any power of a field operator.

Lemma 2.1: Let A_1 and A_2 be self-adjoint operators in Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$. Let

$$A := A_1 \otimes \mathbb{1} + \mathbb{1} \otimes A_2.$$

If $0 \neq \Omega_1 \otimes \Omega_2 \in \mathscr{D}_{\exp\{\pm tA\}}$ for some $t > 0$, then $\Omega_i \in \mathscr{D}_{\exp\{t|A_i|\}}$, $i = 1, 2$.

Proof: With the spectral measures of A_1, A_2 we can write in self-evident notation

$$\|e^{\pm tA}\Omega_1 \otimes \Omega_2\|^2 = \int d\mu_1(x_1)d\mu_2(x_2)e^{\pm 2t(x_1+x_2)} < \infty.$$

By Fubini's theorem the integrals over x_1, x_2 exist separately. Since

$$\int d\mu_i(x)e^{2t|x|} = \int_{x>0} d\mu_i\, e^{2tx} + \int_{x<0} d\mu_i\, e^{-2tx},$$

one obtains the statement. Q. E. D.

We now turn to the Fock part of the theorem and first formalize our basic observation on annihilation operators.

Lemma 2.2: Let a_1 and a_2 be operators in Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$. Let $0 \neq \Omega_i \in \mathscr{D}_{a_i}$ and let

$$\langle \phi_i, a_i\Omega_i\rangle = 0, \quad i = 1,2. \tag{2.1}$$

If

$$\{a_1 \otimes \mathbb{1} + \mathbb{1} \otimes a_2\}\Omega_1 \otimes \Omega_2 = 0, \tag{2.2}$$

then

$$a_i\Omega_i = 0, \quad i = 1,2. \tag{2.3}$$

Proof: We square Eq. (2.2). The mixed terms vanish by Eq. (2.1), and we get $\|a_i\Omega_i\|^2 = 0$. Q. E. D.

In analogy to Segal's formulation[10] of the CCR's we introduce the following notation:

$$\begin{aligned}
&\mathbf{h} = \begin{pmatrix} f \\ g \end{pmatrix}, \quad \mathbf{h} \in \mathscr{V} \oplus \mathscr{V}, \\
&(\mathbf{h}_1,\mathbf{h}_2) = (f_1,f_2) + (g_1,g_2), \\
&\Phi(\mathbf{h}) = \phi(f) + \pi(g), \\
&J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \\
&K_i = \begin{pmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{pmatrix}, \quad i = 1,2.
\end{aligned} \tag{2.4}$$

By Eq. (1.7b), each K_i is invertible.

The CCR's now read

$$[\Phi(\mathbf{h}_1),\Phi(\mathbf{h}_2)] = i(\mathbf{h}_1,J\mathbf{h}_2). \tag{2.5}$$

If T is an invertible operator on $\mathscr{V} \oplus \mathscr{V}$, the field

$$\Phi_T(\mathbf{h}) := \Phi(T\mathbf{h})$$

satisfies the CCR's iff T is symplectic, i.e., iff

$$T^tJT = J, \tag{2.6}$$

which characterizes a Bogoliubov transformation.

Equation (1.6) now reads

$$\Phi(\mathbf{h}) = \Phi_1(K_1\mathbf{h}) \otimes \mathbb{1} + \mathbb{1} \otimes \Phi_2(K_2\mathbf{h}). \tag{2.7}$$

The CCR's imply Eq. (1.7a), which reads

$$K_1^tJK_1 + K_2^tJK_2 = J. \tag{2.8}$$

Without loss of generality we can assume that we have a Fock functional as in Eq. (1.3), since a Bogoliubov transformation does not alter the factorization properties. In view of Lemma 2.1 we can also assume

$$\langle \Omega_i,\Phi_i(\mathbf{h})\Omega_i\rangle = 0.$$

The annihilation property, Eq. (1.5), reads

$$\{\Phi(\mathbf{h}) - i\Phi(J\mathbf{h})\}\Omega = 0. \tag{2.9}$$

Equation (2.7) and Lemma 2.2 thus imply, for $j = 1,2$,

$$\{\Phi_j(K_j\mathbf{h}) - i\Phi_j(K_jJ\mathbf{h})\}\Omega_j = 0. \tag{2.10}$$

Lemma 2.3: For $j = 1,2$,

$$S_j := K_j^t J^t K_j J > 0 \tag{2.11}$$

and

$$JS_j = S_j J. \tag{2.12}$$

The operator

$$T_j := K_j S_j^{-1/2} \tag{2.13}$$

is symplectic.

Proof: Squaring Eq. (2.10) for $\mathbf{h} \neq 0$ and using the CCR's (in their weak form on matrix elements) we obtain

$$\|\Phi_j(K_j\mathbf{h})\Omega_j\|^2 + \|\Phi_j(K_jJ\mathbf{h})\Omega_j\|^2 = -(K_j\mathbf{h}, JK_jJ\mathbf{h}).$$

For $\mathbf{h} \neq 0$, each term on the lhs is nonzero by the CCR's, since $K_j\mathbf{h} \neq 0$. Since $J^t = -J$, Eq. (2.11) follows.

This implies $S_j^t = S_j$ and thus

$$S_jJ = S_j^tJ = J^tK_j^tJK_jJ = JS_j,$$

proving Eq. (2.12). Using $\mathbf{1} = J^tJ$, we obtain

$$\begin{aligned} T_j^tJT_j &= S_j^{-1/2}S_jJS_j^{-1/2} \\ &= S_j^{-1/2}S_jJS_j^{-1/2} \\ &= J \end{aligned}$$

by commutativity. Q. E. D.

In Eq. (2.10) we now replace $\mathbf{h}$ by $S_j^{-1/2}\mathbf{h}$. With Eq. (2.12) we then have

$$\{\Phi_j(T_j\mathbf{h}) - i\Phi_j(T_jJ\mathbf{h})\}\Omega_j = 0, \tag{2.14}$$

with T_j symplectic. Hence Φ_j is a Bogoliubov transform of

$$\widehat{\Phi}_j(\mathbf{h}) := \Phi_j(T_j^{-1}\mathbf{h}),$$

which satisfies

$$\{\widehat{\Phi}_j(\mathbf{h}) - i\Phi_j(J\mathbf{h})\}\Omega_j = 0.$$

Thus Φ_j is a Fock representation by the remark before Eq. (1.5). This proves the Fock part of the theorem.

To prove the quasifree part of the theorem we use Cramér's theorem mentioned in Remark 1 of Sec. I. By Lemma 2.1 we can again assume that all one-point functions vanish. A quasifree generating functional is then of the form

$$E(\mathbf{h}) := \langle \Omega, e^{i\Phi(\mathbf{h})}\Omega \rangle = \exp\{-\langle \Omega, \Phi(\mathbf{h})^2\Omega \rangle/2\}, \tag{2.15}$$

as seen by differentiation. Now let $\mathbf{h}$ be fixed and $t \in R$. Equation (1.8) reads for $E(t\mathbf{h})$

$$E(t\mathbf{h}) = E_1(tK_1\mathbf{h})E_2(tK_2\mathbf{h}). \tag{2.16}$$

The factors on the rhs are positive-definite functions of t, and the lhs is a Gaussian by Eq. (2.15). Hence, by Cramér's theorem,

$$E_i(tK_i\mathbf{h}) = e^{-c_it^2/2}, \quad i = 1,2. \tag{2.17}$$

As in Eq. (2.15), the constants c_i are the "moments"

$$c_i = \langle \Omega_i, \Phi_i(K_i\mathbf{h})^2\Omega_i \rangle. \tag{2.18}$$

Replacing $\mathbf{h}$ by $K_i^{-1}\mathbf{h}$ yields the statement. Q. E. D.

III. DISCUSSION

Cramér's theorem is a highly nontrivial result. It was first conjectured by Lévy and then proved by Cramér by Hadamard's factorization theorem for entire function, a rather deep result. The first part of our proof—based on annihilation operators—suggests trying a "quantum mechanical" proof of Cramér's theorem.

The quasifree part of our theorem may in principle be used to prove the Fock part. But it is easier to prove it directly, as we do, without recourse to Cramér's theorem.

We point out that the quasifree part holds also true without condition Eq. (1.7a)—invariance of the CCR's—as is apparent from the proof. For the Fock part, however, this condition is needed. This is seen from the following.

Example:

$$\phi_1 = \phi_F \otimes \mathbf{1} + \mathbf{1} \otimes \sqrt{2}\,\phi_F,$$

$$\pi_1 = \pi_F \otimes \Omega_F,$$

$$\Omega_1 = \Omega_F \otimes \Omega_F.$$

This is a non-Fock reducible representation of the CCR's and

$$E_1(f,g) = E_F(f,g)e^{-\|f\|^2/2}.$$

With $\phi_2(f) = \pi_1(f)$, $\pi_2(g) = -\phi_1(g)$ we have

$$E_2(f,g) = E_1(g,f)$$

and

$$E_F(f,g) = E_1(f/2,g/2)E_2(f/2,g/2).$$

In this case, however, Eq. (1.7a) is violated.

Condition Eq. (1.7a) insures that a factorization of the generating function corresponds to a decomposition of the fields. The invertibility condition Eq. (1.7b) insures that the decomposition is not just a splitting of $\mathscr{V}$ into disjoint subspaces.

ACKNOWLEDGMENTS

I would like to thank J. R. Klauder and S. N. M. Ruijsenaars for stimulating discussions.

[1]There is a vast literature on CCR's. See, e.g., H. Araki, J. Math. Phys. **1**, 492 (1960); J. R. Klauder, J. Math. Phys. **11**, 609 (1970); review by G. C. Hegerfeldt, in *Functional Analysis: Surveys and Results II*, edited by K.-D. Bierstedt and B. Fuchsteiner (North-Holland, Amsterdam, 1979), pp. 149–165.

[2]This follows from the equivalence of the basis-dependent and basis-independent approaches to CCR's as shown in G. C. Hegerfeldt, J. Math. Phys. **12**, 167 (1971).

[3]G. C. Hegerfeldt, J. Math. Phys. **13**, 821 (1972), and *ibid.* **15**, 621 (1974); M. C. Reed, Commun. Math. Phys. **14**, 336 (1969).

[4]See, e.g., M. Reed and B. Simon, *Methods of Modern Mathematical Physics III* (Academic, New York, 1979), p. 316.

[5]S. N. M. Ruijsenaars (private communication).

[6]D. W. Robinson, Commun. Math. Phys. **1**, 159 (1965). See also J. Manuceau and A. Verbeure, Commun. Math. Phys. **9**, 293 (1968); H. Araki and M. Shiraishi, Publ. RIMS Kyoto Univ. **7**, 105 (1971).

[7]H. Cramér, Math. Zeitschr. **41**, 405 (1936). See also, E. Lukacs, *Characteristic Functions* (Griffin, London, 1970), 2nd ed., p. 243.

[8]G. C. Hegerfeldt (to be published).

[9]G. C. Hegerfeldt, Commun. Math. Phys. **45**, 137 (1975).

[10]I. E. Segal, *Mathematical Problems of Relativistic Physics. Lectures in Applied Mathematics, Vol. II* (AMS, Providence, 1963). See also Ref. 4.

Symmetric-tensor eigenspectrum of the Laplacian on n-spheres

Mark A. Rubin
Theory Group, Department of Physics, The University of Texas, Austin, Texas 78712
Carlos R. Ordóñez
Center for Theoretical Physics, Department of Physics, The University of Texas, Austin, Texas 78712

(Received 13 March 1984; accepted for publication 28 June 1984)

The eigenvalues and degeneracies of the covariant Laplacian acting on symmetric tensors of rank $m \leqslant 2$ defined on n-spheres with $n \geqslant 3$ are given.

I. INTRODUCTION

In a previous paper[1] we computed the eigenvalues and degeneracies associated with the action of the Laplacian on symmetric transverse traceless tensors of rank $m \leqslant 2$ defined on n-dimensional spheres S^n, $n \geqslant 3$. Here we make use of these results to obtain the eigenvalues and degeneracies associated with nontransverse tensors, and present a table of the complete eigenspectrum of the S^n Laplacian on symmetric-tensor fields up to rank two. Knowledge of this spectrum is important for quantum-gravitational computations[2,3] in Kaluza–Klein theories.

II. NONTRANSVERSE VECTORS

On a compact, simply connected manifold without boundary, such as the n-sphere, an arbitrary vector field V_a may be written in a unique manner as the sum of two other vector fields[4]:

$$V_a = T_a + L_a, \tag{2.1}$$

where T_a is transverse,

$$\tilde{\nabla}^a T_a = 0, \tag{2.2}$$

the "longitudinal" vector field L_a is the gradient of a scalar field S,

$$L_a = \tilde{\nabla}_a S, \tag{2.3}$$

and T_a is orthogonal to L_a in that

$$\int d^n x\, g^{1/2} g^{ab} T_a L_b = 0. \tag{2.4}$$

($\tilde{\nabla}_a$ is the covariant derivative operator on S^n.)

The spectrum of $\tilde{\nabla}^a\tilde{\nabla}_a$ acting on transverse vectors has been computed in Ref. 1 (see Table I). Keeping in mind the well-known spectrum of $\tilde{\nabla}^a\tilde{\nabla}_a$ acting on scalars on S^n, we can easily obtain the spectrum of $\tilde{\nabla}^a\tilde{\nabla}_a$ acting on longitudinal vectors on S^n using Eq. (2.3).

Define

$$L_a^{(l)} = \tilde{\nabla}_a T^{(l)} = \partial_a T^{(l)}, \quad l = 1,2\ldots, \tag{2.5}$$

where $T^{(l)}$ is a scalar harmonic (see Table I). Since the $T^{(l)}$'s form a complete set of scalar functions, the $L_a^{(l)}$'s defined in (2.5) will span the space of longitudinal vectors (2.3). Furthermore, the $L_a^{(l)}$'s form a complete (rather than overcomplete) set of longitudinal vectors, since $L_a^{(l)}$'s constructed from orthogonal $T^{(l)}$'s will themselves be orthogonal:

$$\begin{aligned}
&\int d^n x\, g^{1/2} g^{ab}\tilde{\nabla}_a T^{(l)}\tilde{\nabla}_b T'^{(l')} \\
&= \int d^n x\, g^{1/2} g^{ab}\, \partial_a T^{(l)}\, \partial_b T'^{(l')} \\
&= -\int d^n x\, T^{(l)}\, \partial_a (g^{1/2} g^{ab}\, \partial_b T'^{(l')}) \\
&= -\int d^n x\, g^{1/2}\, T^{(l)}\tilde{\nabla}^a\tilde{\nabla}_a T'^{(l')} \\
&= \frac{l'(l'+n-1)}{r^2}\int d^n x\, g^{1/2}\, T^{(l)} T'^{(l')}.
\end{aligned} \tag{2.6}$$

Integration by parts shows that every $L_a^{(l)}$ is orthogonal to every $T_a^{(l)}$. In addition, it is clear that none of the $L_a^{(l)}$'s will be identically zero: $L_a^{(l)} = \partial_a T^{(l)} = 0$ would imply $T^{(l)} = \text{const}$. But the only constant $T^{(l)}$ is $T^{(0)}$ [see Eq. (2.2), Ref. 1], and that one has already been excluded.

To show that the $L_a^{(l)}$'s are eigenfunctions of $\tilde{\nabla}^a\tilde{\nabla}_a$, we make use of the commutation relation[5] for covariant derivatives acting on vectors,

$$(\tilde{\nabla}_a\tilde{\nabla}_c - \tilde{\nabla}_c\tilde{\nabla}_a)V_b = V_d R^d{}_{bac} \tag{2.7}$$

and the relation[5] between the metric and Riemann tensors on an n-sphere of radius r,

$$R_{abcd} = (1/r^2)[g_{bc}g_{ad} - g_{ac}g_{bd}]. \tag{2.8}$$

Employing (2.5), (2.7), and (2.8), we obtain

$$\tilde{\nabla}^a\tilde{\nabla}_a L_b^{(l)} = -([l(l+n-1) - (n-1)]/r^2) L_b^{(l)}, \quad l = 1,2\ldots. \tag{2.9}$$

The degeneracy of the lth eigenvalue in (2.9) is the same as the degeneracy of the lth scalar eigenvalue (see Table I).

III. NONTRANSVERSE TENSORS

A unique decomposition,[4] analogous to (2.1) for vectors, exists as well for rank-two symmetric tensor fields H_{ab} on S^n:

$$H_{ab} = T_{ab} + (1/n) g_{ab} H^c_c + L_{ab}. \tag{3.1}$$

T_{ab} is symmetric, transverse, and traceless:

$$T_{[ab]} = \tilde{\nabla}^a T_{ab} = T^a{}_a = 0. \tag{3.2}$$

The longitudinal traceless part L_{ab} can always be written in the following manner:

$$L_{ab} = \tilde{\nabla}_a V_b + \tilde{\nabla}_b V_a - (2/n) g_{ab}\tilde{\nabla}^c V_c \tag{3.3}$$

TABLE I. Spectrum of $\tilde{\nabla}^a\tilde{\nabla}_a$ on S^n of radius r acting on scalars, vectors, and rank-two symmetric tensors.

Eigenfunction	Eigenvalue	Degeneracy	
Scalars, S:			
$T^{(l)}$	$-\frac{l(l+n-1)}{r^2}$	$D_l(n,0)=\frac{(2l+n-1)(l+n-2)!}{l!(n-1)!}$,	$l=0,1,\ldots$
Vectors, V_a			
$T_a^{(l)};\tilde{\nabla}^aT_a^{(l)}=0$	$-\frac{l(l+n-1)-1}{r^2}$	$D_l(n,1)=\frac{l(l+n-1)(2l+n-1)(l+n-3)!}{(n-2)!(l+1)!}$,	$l=1,2,\ldots$
$L_a^{(l)}=\tilde{\nabla}_aT^{(l)}$	$-\frac{l(l+n-1)-(n-1)}{r^2}$	$D_l(n,0)$	$l=1,2,\ldots$
Symmetric Tensors, H_{ab}			
$T_{ab}^{(l)};\tilde{\nabla}^aT_{ab}^{(l)}=g^{ab}T_{ab}^{(l)}=T_{[ab]}^{(l)}=0$	$-\frac{l(l+n-1)-2}{r^2}$	$D_l(n,2)=\frac{(n+1)(n-2)(l+n)(l-1)(2l+n-1)(l+n-3)!}{2(n-1)!(l+1)!}$	$l=2,3,\ldots$
$\underset{T}{L}{}_{ab}^{(l)}=\tilde{\nabla}_aT_b^{(l)}+\tilde{\nabla}_bT_a^{(l)}$	$-\frac{l(l+n-1)-(n+2)}{r^2}$	$D_l(n,1)$	$l=2,3,\ldots$
$\underset{L}{L}{}_{ab}^{(l)}=\tilde{\nabla}_aL_b^{(l)}+\tilde{\nabla}_bL_a^{(l)}-(2/n)g_{ab}\tilde{\nabla}^cL_c^{(l)}$ $=2\tilde{\nabla}_a\tilde{\nabla}_bT^{(l)}-(2/n)g_{ab}\tilde{\nabla}^c\tilde{\nabla}_cT^{(l)}$	$-\frac{l(l+n-1)-2n}{r^2}$	$D_l(n,0)$	$l=2,3,\ldots$
$g_{ab}T^{(l)}$	$-\frac{l(l+n-1)}{r^2}$	$D_l(n,0)$	$l=0\ 1,\ldots$

for some vector V_b. This vector, however, is not unique, since if any V_b satisfies (3.3) for a given L_{ab}, the vector V_b+C_b will satisfy (3.3) with the same L_{ab} provided C_b is a conformal Killing vector field, i.e.,

$$\tilde{\nabla}_aC_b+\tilde{\nabla}_bC_c-(2/n)g_{ab}\tilde{\nabla}^cC_c=0. \tag{3.4}$$

Since we know that any vector can be written as a linear combination of transverse harmonics $T_a^{(l)}$ (see Table I) and the longitudinal vector harmonics (2.5), we see that the space of longitudinal traceless tensor fields L_{ab} is spanned by the following sets of "longitudinal-transverse" and "longitudinal-longitudinal" tensor fields, respectively,

$$\underset{T}{L}{}_{ab}^{(l)}=\tilde{\nabla}_aT_b^{(l)}+\tilde{\nabla}_bT_a^{(l)} \tag{3.5a}$$

and

$$\underset{L}{L}{}_{ab}^{(l)}=\tilde{\nabla}_aL_b^{(l)}+\tilde{\nabla}_bL_a^{(l)}-(2/n)g_{ab}\tilde{\nabla}^cL_c^{(l)}$$

$$=2\tilde{\nabla}_{(a}\tilde{\nabla}_{b)}T^{(l)}-(2/n)g_{ab}\tilde{\nabla}^c\tilde{\nabla}_cT^{(l)}. \tag{3.5b}$$

(We shall determine later those values of l for which these tensors are nonzero.) The commutation relation for covariant derivatives on a rank-two tensor[5] is

$$(\tilde{\nabla}_a\tilde{\nabla}_b-\tilde{\nabla}_b\tilde{\nabla}_a)H_{cd}=H_{ed}R^e{}_{cab}+H_{ce}R^e{}_{dab}. \tag{3.6}$$

After some straightforward manipulations using (3.6), we find that the tensors (3.5) are eigenfunctions of $\tilde{\nabla}^a\tilde{\nabla}_a$,

$$\tilde{\nabla}^a\tilde{\nabla}_a\underset{T}{L}{}_{bc}^{(l)}=-\frac{[l(l+n-1)-(n+2)]}{r^2}\underset{T}{L}{}_{bc}^{(l)} \tag{3.7a}$$

and

$$\tilde{\nabla}^a\tilde{\nabla}_a\underset{L}{L}{}_{bc}^{(l)}=-\frac{[l(l+n-1)-2n]}{r^2}\underset{L}{L}{}_{bc}^{(l)}. \tag{3.7b}$$

We now determine which of the longitudinal tensor harmonics (3.5) are identically zero. Consider first the longitudinal transverse tensors (3.5a). If $\underset{T}{L}{}_{ab}^{(l)}=0$, then the corresponding $T_b^{(l)}$ is a Killing vector field,

$$0=\tilde{\nabla}_aT_b^{(l)}+\tilde{\nabla}_bT_a^{(l)}, \tag{3.8}$$

so

$$0=\tilde{\nabla}^a(\tilde{\nabla}_aT_b^{(l)}+\tilde{\nabla}_bT_a^{(l)}). \tag{3.9}$$

Using (2.7), (3.9) becomes

$$0=-\frac{[l(l+n-1)-n]}{r^2}T_b^{(l)}. \tag{3.10}$$

Since $T_b^{(l)}$ is not identically zero, l must equal 1. Thus, the nonzero $\underset{T}{L}{}_{ab}^{(l)}$'s are those with $l=2,3,\ldots$.

Similarly, if a longitudinal-longitudinal tensor (3.5b) is zero, it arises from a longitudinal vector harmonic satisfying the conformal Killing equation,

$$0=\tilde{\nabla}_aL_b^{(l)}+\tilde{\nabla}_bL_a^{(l)}-(2/n)g_{ab}\tilde{\nabla}^cL_c^{(l)}. \tag{3.11}$$

Taking the divergence of this equation and using (2.5) and (2.7), we find

$$0=-\frac{[l(l+n-1)-n]}{r^2}L_b^{(l)}, \tag{3.12}$$

so the nonzero $\underset{L}{L}{}_{ab}^{(l)}$'s also start at $l=2$.

From Table I we note that the number of $T_a^{(l)}$'s is $n(n+1)/2$, which is equal to the number of linearly independent Killing vector fields on the n-sphere.[5] We also see that the number of $L_a^{(l)}$'s is $n+1$, which is the number of conformal Killing vector fields on S^n which are not also true Killing vector fields.[6,7]

Finally, we show that the longitudinal tensor harmonics (3.5a) and (3.5b) are linearly independent, provided that the $T_a^{(l)}$'s and $T^{(l)}$'s from which they are constructed are linearly independent. Suppose not; then there exists at least one relation of the form

$$\sum_{l,q}\underset{T}{C}(l,q)\underset{T}{L}{}_{ab}^{(l,q)}+\sum_{l,q}\underset{L}{C}(l,q)\underset{L}{L}{}_{ab}^{(l,q)}=0, \tag{3.13}$$

where the $\underset{T}{C}(l,q)$'s and $\underset{L}{C}(l,q)$'s are constants not all of which are zero. The index "q" distinguishes between $\underset{T}{L}{}_{ab}^{(l)}$'s ($\underset{L}{L}{}_{ab}^{(l)}$'s) constructed out of different $T_a^{(l)}$'s ($L_a^{(l)}$'s) corresponding to the same l. Using the definitions (3.5a) and (3.5b), Eq. (3.13) becomes

$$\tilde{\nabla}_a\Big(\sum_{l,q}C_T(l,q)T_b^{(l,q)}+\sum_{l,q}C_L(l,q)L_b^{(l,q)}\Big)+\tilde{\nabla}_b\Big(\sum_{l,q}C_T(l,q)T_a^{(l,q)}+\sum_{l,q}C_L(l,q)L_a^{(l,q)}\Big)-\frac{2}{n}g_{ab}\tilde{\nabla}^c\Big(\sum_{l,q}C_T(l,q)T_c^{(l,q)}+\sum_{l,q}C_L(l,q)L_c^{(l,q)}\Big)=0. \tag{3.14}$$

Since the $T_b^{(l,q)}$'s and $L_b^{(l,q)}$'s are linearly independent, the quantity in parentheses in (3.14) must be nonzero. Equation (3.14) then says that this quantity must be a conformal Killing vector field; but this is impossible, since, as we have shown, conformal Killing vector fields must be made out of $T_a^{(l,q)}$'s and $L_a^{(l,q)}$'s with $l=1$, and these are already excluded in the construction of the longitudinal harmonics.

The results obtained in this paper[8] and in Ref. 1 are summarized in Table I.

ACKNOWLEDGMENTS

We would like to thank Paul Townsend and James York for helpful discussions.

Work was supported in part by the Robert A. Welch Foundation, NSF PHY8304629 and NSF PHY8205717.

[1]M. A. Rubin and C. R. Ordóñez, J. Math. Phys. 25, 2888 (1984). All of our differential-geometric conventions are the same as in this paper.

[2]C. R. Ordóñez and M. A. Rubin, "Graviton Dominance in Quantum Kaluza–Klein Theory," University of Texas Theory Group preprint.

[3]A. Chodos and E. Myers, Ann. Phys. (N.Y.) 156, 412 (1984).

[4]W. V. D. Hodge, *Theory and Application of Harmonic Integrals* (Cambridge University, Cambridge, 1952); J. W. York, J. Math. Phys. 14, 456 (1973); S. M. Christensen and M. J. Duff, Nucl. Phys. B 170 [FS1], 480 (1980); M. Christensen, M. J. Duff, G. W. Gibbons, and M. Roček, "One-Loop Effects in Supergravity with a Cosmological Constant" (unpublished).

[5]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).

[6]K. Yano, *Integral Formulas in Riemannian Geometry* (Dekker, New York, 1970); M. J. Duff, P. K. Townsend, and P. van Nieuwenhuizen, Phy. Lett. B 122, 232 (1983).

[7]L. Castellani, R. D.'Auria, P. Fré, K. Pilch, and P. van Niewenhuizen, "The Complete Mass Spectrum on S_7," State University of New York at Stony Brook preprint.

[8]After obtaining our results we received a preprint (Ref. 3), which obtains by different means the same results. More recently, harmonic analysis of tensor and spinor fields on general coset spaces has been studied by the authors of Ref. 7, as well as: R. D'Auria and P. Fré, "On the Fermion Mass-Spectrum of Kaluza–Klein Supergravity," State University of New York at Stony Brook preprint ITP-SB-83-57; L. Castellani, R. D'Auria, P. Fré, K. Pilch, and P. van Nieuwenhuizen, "Fermion Masses for the Englert Solution on the Round 7-Sphere" and "The Bosonic Mass Formula for Freund–Rubin solutions of $d=11$ Supergravity on General Coset Manifolds," State University of New York at Stony Brook preprints ITP-SB-83-60 and ITP-SB-83-61; K. Pilch and A. N. Schellekens, "Formulae for the Eigenvalues of the Laplacian on Tensor Harmonics on Symmetric Coset Spaces," State University of New York at Stony Brook preprint ITP-SB-84-20.

Invariants for dissipative nonlinear systems by using rescaling

Marc R. Feix[a] and H. Ralph Lewis
Los Alamos National Laboratory, Center for Nonlinear Studies and CTR Division, Los Alamos, New Mexico 87545

(Received 5 January 1984; accepted for publication 28 June 1984)

A rescaling transformation of space and time is introduced in the study of nonlinear dissipative systems that are described by a second-order differential equation with a friction term proportional to the velocity, $\beta(t)v$. The transformation is of the form $(x,t)\to(\xi,\theta)$, where $x=\xi C(t)+\alpha(t), d\theta=dt/A^2(t)$. This rescaling is used to find each potential for which there exists an exact invariant quadratic in the velocity and to find the invariant. The invariants are found explicitly for a power-law potential, $\gamma(t)x^{m+1}/(m+1)$, and an arbitrary coefficient of friction $\beta(t)$. We show in an example how the rescaling transformation can be chosen to give an asymptotic solution of the equation in cases where the exact invariant does not exist. For certain parameters, the asymptotic solution is a self-similar solution that is an attractor for all initial conditions. The technique of applying a rescaling transformation has been useful in other problems and may have additional practical applications.

I. INTRODUCTION

In recent years considerable work has been devoted to finding exact invariants for Hamiltonian systems. One-dimensional systems have received the most attention (cf., for example, Refs. 1–3) and there has been some work on three-dimensional systems.[4] A major motivation for studying invariants of dynamical systems is the possibility for application in plasma physics and other self-consistent many-body problems; and applications of exact invariants have been made to the Vlasov–Poisson equations recently.[5,6] In this paper we discuss the extension to dissipative systems of some ideas and results that have arisen in connection with finding exact invariants for nondissipative systems.

For the Hamiltonian

$$H=\tfrac{1}{2}p^2+V(q,t), \tag{1.1}$$

all of the potentials $V(q,t)$ have been found such that there exists an exact invariant that is quadratic in the momentum p (see Ref. 3). For those potentials, the invariants can be obtained and the system rendered autonomous by making the generalized canonical transformation[7] from (q,p,t) to (Q,P,T) given by

$$Q=(q-\alpha)/\rho,\ P=[\rho(p-\dot{\alpha})-\dot{\rho}(q-\alpha)],$$
$$T=\int^t\rho^{-2}(t')dt', \tag{1.2}$$

where $\alpha(t)$ and $\rho(t)$ are arbitrary functions of time and an overdot denotes differentiation with respect to time. This transformation is a type of rescaling of the variables (q,p,t). Results for the Hamiltonian (1.1) can also be applied immediately to a certain class of dissipative systems because the more general Hamiltonian

$$H=a(t)p^2+b(q,t)p+c(q,t) \tag{1.3}$$

can be transformed to (1.1) by the generalized canonical transformation[7]

$$Q=q,\ P=p+\frac{b}{2a},\ T=\int^t a(t')dt'. \tag{1.4}$$

This transformation is a generalization of one used by Kanai.[8]

The original derivation of the quadratic invariants for the Hamiltonian (1.1) did not start from the transformation (1.2); rather, the transformation was a result. A different point of view, which has been adopted before in connection with problems related to the Hamiltonian (1.1),[9] is to begin with a rescaling transformation. This will be our starting point for discussing dissipative systems. We shall generalize (1.2) by rescaling space and time in terms of three time-dependent functions instead of two; however, we shall not assume an *a priori* rescaling of momentum. A rescaling transformation of configuration space variables and time only does not require that the system under study be Hamiltonian. It can be used with an arbitrary system. In general, our objective in introducing such a transformation is to obtain equations in the new variables for which an exact or approximate invariant can be found. It also may be that the new equations can be treated more conveniently than the old equations even if an invariant cannot be found. This possibility is important because we can expect that explicit exact invariants will be found only for severely restricted classes of dynamical systems, no matter what method be used.

A rescaling transformation can sometimes be used to remove dissipation terms from the equations of motion, as is the case with a damped linear oscillator. In other cases it can be useful to introduce or modify dissipation by means of rescaling. For example, friction in the new equations may allow asymptotic solution of the equations for large values of the new time variable. Such asymptotic solutions can represent nontrivial attractors in terms of the original variables.[10,11]

The extent of possible applications of rescaling to dissipative systems is not known. Many more useful applications may yet be found. An important area of practical interest, where a multidimensional generalization of the transforma-

[a] Permanent address: Physique Mathematique, Modelisation et Simulation, Université d'Orleans, 45045 Orleans Cedex, France.

tion used in this paper may be useful, is resistive magnetohydrodynamics.

In this paper we illustrate a particular rescaling procedure for dissipative systems by discussing examples. In Sec. II we introduce rescaling for a second-order ordinary differential equation that describes the motion in one dimension of a particle moving under the influence of a potential in the presence of friction. This is the equation of motion for a particular Hamiltonian system of the type represented by (1.3). In Sec. III, we consider a particular class of such equations that are associated, for example, with nonlinear small oscillations about an equilibrium point, which occur when the curvature of the potential at the equilibrium point vanishes. We find the cases for which our rescaling transformation leads to an autonomous system and we solve the equations explicitly and in detail. For those cases, the energy for the new equations is an exact invariant, quadratic in the velocity, and our results are an explicit example of the known general results for the Hamiltonians (1.1) and (1.3). Our results are also a generalization of a result due to Sarlet and Bahar.[12] We give an example in this class of equations from which the solution of the Emden equation for two different exponents can be obtained. In Sec. IV we consider cases for which the rescaling transformation does not lead to an autonomous system, but does allow us to find an invariant associated with the long-time asymptotic behavior of solutions. We summarize our conclusions in Sec. V.

II. THE RESCALING

We consider second-order ordinary differential equations of the form

$$\frac{d^2x}{dt^2}+\beta(t)\frac{dx}{dt}+\frac{\partial\phi}{\partial x}=0, \tag{2.1}$$

where $\phi(x,t)$ is a space- and time-dependent potential and $\beta(t)$ is a time-dependent coefficient of friction, which may be positive or negative. We introduce a rescaling transformation from (x,t) to (ξ,θ) according to

$$x=\xi C(t)+\alpha(t),\ dt=A^2(t)d\theta, \tag{2.2}$$

where $C(t)$, $\alpha(t)$, and $A(t)$ are three arbitrary functions. We shall call $\xi(\theta)$ the new coordinate and θ the new time. The equation of motion in the new variables is

$$\frac{d^2\xi}{d\theta^2}+\hat{\beta}\frac{d\xi}{d\theta}+\frac{\partial\hat{\phi}}{\partial\xi}=0, \tag{2.3}$$

where

$$\hat{\beta}(\theta)=A^2\left[\beta+2\left(\frac{1}{C}\frac{dC}{dt}-\frac{1}{A}\frac{dA}{dt}\right)\right], \tag{2.4}$$

$$\hat{\phi}(\xi,\theta)=\frac{A^4}{C^2}\phi+\frac{1}{2}\frac{A^4}{C}\left(\frac{d^2C}{dt^2}+\beta\frac{dC}{dt}\right)\xi^2+\frac{A^4}{C}\left(\frac{d^2\alpha}{dt^2}+\beta\frac{d\alpha}{dt}\right)\xi. \tag{2.5}$$

The transformations defined by (2.2) form a group. From (2.2) it is easily shown that if C_1, A_1, and α_1 characterize one transformation and C_2, A_2, and α_2 characterize another, then the result of applying transformation 1 followed by transformation 2 is another transformation of the same form, characterized by $C=C_1C_2$, $A=A_1A_2$, and $\alpha=\alpha_1+C_1\alpha_2$.

If $\hat{\beta}=0$ and $\hat{\phi}$ is independent of θ, then the energy in the new frame,

$$\epsilon\equiv\frac{1}{2}\left(\frac{d\xi}{d\theta}\right)^2+\hat{\phi}, \tag{2.6}$$

will be an exact invariant. Setting $\hat{\beta}=0$ in (2.4), we can immediately integrate the equation to find

$$A(t)=C(t)\exp[\tfrac{1}{2}B(t)], \tag{2.7}$$

where

$$B(t)=\int^t\beta(t')dt'. \tag{2.8}$$

By using (2.7) in (2.5), we can find the form which ϕ must have in order that $\hat{\phi}$ be independent of θ:

$$\phi(x,t)=\frac{1}{C^2\exp(2B)}\hat{\phi}\left(\frac{x-\alpha}{C}\right)-\frac{1}{2}\left(\frac{\ddot{C}+\beta\dot{C}}{C}\right)(x-\alpha)^2-(\ddot{\alpha}+\beta\dot{\alpha})(x-\alpha), \tag{2.9}$$

where $\hat{\phi}$ is an arbitrary function of its argument. The exact invariant ϵ can be expressed in terms of x and t through the relation

$$\frac{dx}{dt}=\frac{C}{A^2}\frac{d\xi}{d\theta}+\xi\dot{C}+\dot{\alpha}. \tag{2.10}$$

The result is

$$\epsilon=\frac{1}{2}\exp(2B)\left[C\left(\frac{dx}{dt}-\dot{\alpha}\right)-\dot{C}(x-\alpha)\right]^2+\hat{\phi}\left(\frac{x-\alpha}{C}\right). \tag{2.11}$$

Equations (2.9) and (2.11) are an example of the results found previously concerning invariants quadratic in the momentum (or velocity) for Hamiltonians of the form (1.1) and (1.3) (see Ref. 3).

III. APPLICATION TO A PARTICULAR CLASS OF EQUATIONS

Consider the potential

$$\phi(x,t)=\gamma(t)[x^{m+1}/(m+1)],\quad m\neq 1. \tag{3.1}$$

This potential is associated with small nonlinear oscillations about an equilibrium at $x=0$ if m is an integer larger than unity. Such oscillations occur if the curvature of the potential at the equilibrium point vanishes. With this potential (2.1) becomes

$$\frac{d^2x}{dt^2}+\beta(t)\frac{dx}{dt}+\gamma(t)x^m=0. \tag{3.2}$$

For certain functions $\gamma(t)$ and $\beta(t)$, this equation is of the form discussed in the previous section for which the rescaling transformation renders the equation autonomous with an exact energy invariant. In order for the potential to correspond to the allowed potentials given by (2.9), we must take

$$\alpha(t)\equiv 0,\ \hat{\phi}(\xi)=[K/(m+1)]\xi^{m+1}+(L/2)\xi^2, \tag{3.3}$$

where K and L are arbitrary constants. The requirement that

(2.9) be satisfied separately for each power of x leads to the following conditions on γ, β, and C:

$$\gamma(t) = K/C^{m+3}\exp[2B(t)], \tag{3.4}$$

$$\frac{d^2C}{dt^2} + \beta\frac{dC}{dt} = \frac{L}{C^3\exp[2B(t)]}. \tag{3.5}$$

Equation (3.5) can be solved for C in terms of β. Then (3.4) gives the family of functions $\gamma(t)$ which, for a given $\beta(t)$, make the potential (3.1) be in the class of allowable potentials given by (2.9). For such a $\gamma(t)$, (3.2) is transformed into

$$\frac{d^2\xi}{d\theta^2} + L\xi + K\xi^m = 0 \tag{3.6}$$

by the rescaling transformation. This equation can be solved easily and, for $m = 2$ and $m = 3$, the final quadrature can be performed in terms of elliptic functions.

We can find the general solution of (3.5). First we notice that the general solution for $L = 0$ is given by

$$C(t) = J(t) \tag{3.7a}$$

and that a class of solutions for $L = -b^2/4$ is given by

$$C^2(t) = J(t), \tag{3.7b}$$

where $J(t)$ is defined by

$$J(t) = a + b\int^t \exp[-B(t')]dt', \tag{3.8}$$

and a and b are constants. We next make the ansatz that the general solution of (3.5) can be written as

$$C^2(t) = MJ^2(t) + NJ(t), \tag{3.9}$$

where M and N are constants. It will turn out that this ansatz is correct. Differentiating (3.9) twice we obtain

$$2C\dot{C} = 2MJ\dot{J} + N\dot{J}, \tag{3.10}$$

$$2(C\ddot{C} + \dot{C}^2) = 2M\dot{J}^2 + 2MJ\ddot{J} + N\ddot{J}. \tag{3.11}$$

Multiplying (3.10) by β and adding (3.11) we find

$$2[C(\ddot{C} + \beta\dot{C}) + \dot{C}^2] = (2M + N)(\ddot{J} + \beta\dot{J}) + 2M\dot{J}^2. \tag{3.12}$$

Since $\ddot{J} + \beta\dot{J} = 0$, (3.12) simplifies to

$$C(\ddot{C} + \beta\dot{C}) = -\dot{C}^2 + M\dot{J}^2. \tag{3.13}$$

From (3.10) we have $\dot{C} = (1/2C)\dot{J}(2MJ + N)$ and we can rewrite (3.13) as

$$C(\ddot{C} + \beta\dot{C}) = (\dot{J}^2/4C^2)[4MC^2 - (2MJ + N)^2]. \tag{3.14}$$

Taking (3.9) into account and the fact that $\dot{J} = b \times\exp[-B(t)]$, (3.14) becomes

$$\ddot{C} + \beta\dot{C} = -\{(b^2/4)/C^3\exp[2B(t)]\}N^2. \tag{3.15}$$

Thus, the general solution of (3.5) is obtained by taking $N = 1$ and $-b^2/4 = L$:

$$C^2 = MJ^2 + J, \tag{3.16a}$$

$$J(t) = a + (-4L)^{1/2}\int^t \exp[-B(t')]dt', \tag{3.16b}$$

where M and a are arbitrary constants. The fact that there are two arbitrary constants shows that this is indeed the general solution.

Combining (3.4) and (3.16), we find that there is an energy invariant for (3.2) when $\gamma(t)$ is given by

$$\gamma(t) = K/(MJ^2 + J)^{(m+3)/2}\exp[2B(t)]. \tag{3.17}$$

Written explicitly, the invariant is

$$\begin{aligned}\epsilon = {}& \tfrac{1}{2}J(MJ+1)\exp[2B(t)] \\ & \times\left[\frac{dx}{dt} - \frac{\dot{J}(1+2MJ)}{2J(1+MJ)}x\right]^2 - \frac{b^2}{8}\frac{x^2}{J(MJ+1)} \\ & + \frac{K}{m+1}\frac{x^{m+1}}{(MJ^2+1)^{(m+1)/2}}.\end{aligned} \tag{3.18}$$

This result is a generalization of a result given by Sarlet and Bahar,[12] who also obtained this invariant for (3.2), except that the class of functions $\gamma(t)$ that they found is a subclass of the functions given by (3.17). Their results correspond to the case $M = 0$, which gives the solution (3.7b) for $C(t)$.

An example: Consider the case

$$\beta(t) = \frac{\lambda}{t},\ \exp[B(t)] = t^{\lambda},\ J(t) = a + \frac{b}{1-\lambda}t^{1-\lambda}. \tag{3.19}$$

From (3.17) we have

$$\begin{aligned}\gamma(t) = K\{ & [M(a + [b/(1-\lambda)]t^{l-\lambda})^2 \\ & + (a + [b/(1-\lambda)]t^{1-\lambda})]^{(m+3)/2}t^{2\lambda}\}^{-1}.\end{aligned} \tag{3.20}$$

Two interesting limiting cases can be obtained from (3.20). The first is

$$M = 0,\quad a = 0,\quad b = 1-\lambda, \tag{3.21}$$

which leads to

$$\gamma(t) = Kt^{[\lambda(m-1)-(m+3)]/2}. \tag{3.22}$$

The rescaling functions $A(t)$ and $C(t)$ are

$$A(t) = t^{1/2},\ C(t) = t^{(1-\lambda)/2}. \tag{3.23}$$

The form of $A(t)$ indicates that the new time, θ, is at the limit of logarithmic compression. That is,

$$\theta - 1 = \log t, \tag{3.24}$$

where θ is normalized such that $t = 1$ corresponds to $\theta = 1$. This limit of logarithmic compression is found frequently.[13] The invariant for this case is

$$\begin{aligned}\epsilon = {}& t^{\lambda-1}[\tfrac{1}{2}v^2t^2 + \tfrac{1}{2}xvt(\lambda - 1)] \\ & + [K/(m+1)](xt^{(\lambda-1)/2})^{m+1},\end{aligned} \tag{3.25}$$

where

$$v \equiv \frac{dx}{dt}. \tag{3.26}$$

If we set $\lambda = 2$ and $m = 5$ in this example and replace t by r, the radial coordinate in a spherical coordinate system, then (3.2) becomes the Emden equation with exponent 5, which arises in the study of spherically symmetric stellar equilibria,[14]

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dx}{dr}\right) + x^m = \frac{d^2x}{dr^2} + \frac{2}{r}\frac{dx}{dr} + x^m = 0. \tag{3.27}$$

The fact that the energy invariant exists for the Emden equation with exponent 5 was recognized by Chandrasekhar.[14]

The second interesting limiting case is obtained from

$$a=0 \text{ and } b\to 0,\ M\to\infty \text{ with } Mb^2\to(1-\lambda)^2. \quad (3.28)$$

With these limits, we obtain

$$\gamma(t)=Kt^{\lambda(m+1)-(m+3)}, \quad (3.29)$$

$$A(t)=t^{1-\lambda/2},\ C(t)=t^{1-\lambda}, \quad (3.30)$$

$$\epsilon=\tfrac{1}{2}[vt+(\lambda-1)x]^2+[K/(m+1)](xt^{\lambda-1})^{m+1}. \quad (3.31)$$

From (3.29) we see that $\lambda=2$ and $m=1$ gives the Emden equation with exponent unity. Thus, we have also found the energy invariant that Chandrasekhar[14] obtained for that equation.

IV. ASYMPTOTIC INVARIANTS

We consider the equation

$$\frac{d^2x}{dt^2}+\frac{\lambda}{t}\frac{dx}{dt}+\mu t^r(\operatorname{sgn} x)|x|^m=0, \quad (4.1)$$

which has the same form as the limiting cases considered in the example in Sec. III. [See (3.22) and (3.29).] For this equation we have

$$\beta(t)=\lambda/t,\ \exp[B(t)]=t^\lambda,\ \gamma(t)=\mu t^r. \quad (4.2)$$

We make a particular choice of transformation functions $A(t)$ and $C(t)$ that is convenient for obtaining exact energy invariants for (4.1) as well as for finding asymptotic results when an exact energy invariant does not exist. This choice, which has been useful in some other problems,[11] is based on a splitting of the $\hat\phi(\xi,\theta)$ defined by (2.5) into two parts. We define a "rescaled potential" $\hat\phi_1$ and a "transformation potential" $\hat\phi_2$ by

$$\hat\phi_1(\xi,\theta)=(A^4/C^2)\phi, \quad (4.3)$$

$$\hat\phi_2(\xi,\theta)=\frac{1}{2}\frac{A^4}{C}\left(\frac{d^2C}{dt^2}+\beta\frac{dC}{dt}\right)\xi^2+\frac{A^4}{C}\left(\frac{d^2\alpha}{dt^2}+\beta\frac{d\alpha}{dt}\right)\xi, \quad (4.4)$$

so that

$$\hat\phi(\xi,\theta)=\hat\phi_1(\xi,\theta)+\hat\phi_2(\xi,\theta). \quad (4.5)$$

For treating (4.1), we select $C(t)$ and $A(t)$ such that the rescaled potential is θ-independent and we choose a relation between $C(t)$ and $A(t)$.

To begin with, we relate $C(t)$ and $A(t)$ according to (2.7), so that $\hat\beta=0$. Then, choosing $\alpha(t)=0$ and $C(t)$ according to

$$C^{m+3}t^{2\lambda+r}=1, \quad (4.6)$$

we find

$$\hat\phi_1(\xi)=\mu[|\xi|^{m+1}/(m+1)], \quad (4.7)$$

$$\hat\phi_2(\xi,\theta)=\frac{\xi^2}{2}\frac{(2\lambda+r)[r+m+3-\lambda(m+1)]}{(m+3)^2}\times\left[1+(\theta-1)\frac{m+3+2r-\lambda(m-1)}{m+3}\right]^{-2}, \quad (4.8)$$

where θ is normalized such that $\theta=1$ when $t=1$. The relation between θ and t is

$$(\theta-1)\frac{m+3+2r-\lambda(m-1)}{m+3}=t^{[m+3+2r-\lambda(m-1)]/(m+3)}-1. \quad (4.9)$$

From (4.8) we recognize the two limiting cases given by (3.22) and (3.29) for which an exact invariant can be found when $\gamma(t)$ has the form μt^r. For $r=[\lambda(m-1)-(m+3)]/2$ [corresponding to (3.22)], $\hat\phi_2$ is θ-independent. For $r=\lambda(m+1)-(m+3)$ [corresponding to (3.29)], $\hat\phi_2$ is zero. For both cases, the energy ϵ is an exact invariant since $\hat\beta=0$ and $\hat\phi$ is θ-independent.

Now let us consider the case where the energy invariant does not exist and, to begin with, we suppose that we have

$$r>[\lambda(m-1)-(m+3)]/2. \quad (4.10)$$

Moreover, now and henceforth we assume

$$m>0,\ \lambda>1,\ \mu>0. \quad (4.11)$$

From (4.9) we see that $\theta\to\infty$ when $t\to\infty$; and from (4.8) we see that $\hat\phi_2\to 0$ when $\theta\to\infty$. Asymptotically, the motion in the new space is frictionless in a θ-independent potential. Therefore, ϵ (the energy in the new space) is an asymptotic invariant. We notice also that the region in the (r,m) plane described by (4.10) and (4.11) includes the straight line $r=\lambda(m+1)-(m+3)$. (See Fig. 1.) On this line the asymptotic invariant is also the exact invariant corresponding to (3.29), for which $\hat\phi_2$ is zero.

At this point we notice that, for $m<1$, a negative transformation potential can play a role asymptotically even though it goes to zero as θ^{-2} for each ξ. In this case, there is a potential crest that can push a particle into the outer region where the dominating potential is the transformation potential that goes to minus infinity for large $|\xi|$. Then, in order to find the asymptotic behavior, we simply have to solve (4.1) neglecting the potential. Obviously, this will depend on initial conditions and it indicates a bifurcation of the particle trajectories, some particles escaping the physical potential, others oscillating in the rescaled potential. To delineate the region for which such behavior can take place, it is sufficient to compare, at large times, the velocity of the moving crest with the velocity of particles in the outer region. For $r=-2$ these velocities are equal for all values of λ when $m<1$. Moreover, comparison of these velocities for $r<-2$ and $r>-2$ shows that the bifurcation can only take place for values of r and m inside the triangle bounded by the three straight lines $m=0$, $r=-2$, and $r=[m(\lambda-1)-(\lambda+3)]/2$. (See Fig. 1.)

We now turn to the case

$$r<[\lambda(m-1)-(m+3)]/2, \quad (4.12)$$

which is the opposite of (4.10). The rescaling transformation that we used when (4.10) was satisfied is inconvenient for analyzing the long-time behavior when (4.12) holds. We need to find another "new space" where the asymptotic properties can be deduced from physical arguments. We give up the requirement $\hat\beta=0$, instead selecting $C(t)$ and $A(t)$ such that the rescaled potential and the transformation potential are both θ-independent. The transformation given by

$$C(t)=t^{(r+2)/(1-m)},\ A(t)=t^{1/2},\ \alpha(t)=0 \quad (4.13)$$

will suffice. It is interesting to note that the logarithmic compression of time associated with $A(t)$ is the same as we had from (4.9) in the limiting case $2r = \lambda(m-1) - (m+3)$.

The total new potential is now

$$\hat{\phi}(\xi) = \frac{\mu}{m+1} |\xi|^{m+1} - \frac{\xi^2}{2} \frac{r+2}{(m-1)^2} [\lambda(m-1) - r - m - 1]. \tag{4.14}$$

From (2.4) and (4.13) we have for the new friction

$$\hat{\beta}(\theta) = t\left[\frac{\lambda}{t} + 2\left(\frac{\dot{C}}{C} - \frac{\dot{A}}{A}\right)\right] = \frac{3 + 2r + m - \lambda(m-1)}{1-m}. \tag{4.15}$$

For a given λ, (4.14) and (4.15) partition the (m,r) parameter plane by means of the four straight lines

$$r = [(\lambda - 1)m - (\lambda + 3)]/2,$$
$$r = (\lambda - 1)m - (\lambda + 1), \quad r = -2, \quad m = 1, \tag{4.16}$$

as illustrated in Fig. 1. There are four zones, labeled I, II, III, and IV. In zones I and II, the coefficient of friction is positive, while it is negative in zones III and IV. The potentials $\hat{\phi}$ and the signs of the coefficients of friction are indicated in Fig. 2.

We begin with zone II. Because the friction is positive, the particle will asymptotically reach the neighborhood of $\xi = 0$, where, since $m > 1$, the dominant potential in ξ space is the transformation potential. This means that the motion for $t \to \infty$ is described by (4.1) without the potential term,

$$\frac{d^2x}{dt^2} + \frac{\lambda}{t}\frac{dx}{dt} = 0. \tag{4.17}$$

The solution is

$$x = x_0 t^{1-\lambda} + y_0, \tag{4.18}$$

where x_0 and y_0 are arbitrary constants. In zones III and IV, the coefficient of friction is negative and we shall have $|\xi| \to \infty$ for large values of θ or t. However, because $m < 1$ in these zones, the dominant potential in ξ space for large $|\xi|$ is again the transformation potential and the asymptotic solutions will again be given by (4.17) and (4.18).

In zone I, we have both a positive coefficient of friction and a total potential in ξ space with two wells, as is indicated in Fig. 2. Consequently, the motion in the new space is bifurcated and the asymptotic state is a static equilibrium at the bottom of one of the potential wells, located at $\pm \xi_0$. Therefore, the asymptotic solution for x is

$$x = \xi_0 t^{-[(r+2)/(m-1)]}. \tag{4.19}$$

The choice of $C(t)$ in (4.13), which has the same time dependence as (4.19), could have been suggested by a short consideration of the stretching group symmetries that are admitted by (4.1). For zones I and IV, there is a self-similar

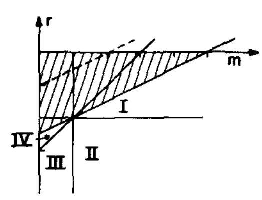


FIG. 1. The (r,m) space (drawn for the case $\lambda = 2$). In the hatched region we have ϵ asymptotically invariant. The rest of the space is divided into four zones. The equations for the three lines are $r = (m-5)/2$, $r = m - 1$, and $r = m - 3$.

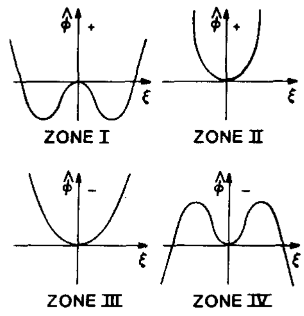


FIG. 2. The total potential $\hat{\phi}(\xi)$ in the four zones. The signs indicated are the signs of the coefficients of friction. In zone I, the points $\pm \xi_0$ are attractors; in zone IV they are unstable.

solution of (4.1) with this time dependence and, therefore, it could have been supposed that this time dependence would play an important role in describing the solution for those zones. The stretching group transformation is

$$t = c^\delta \bar{t}, \quad x = c^\eta \bar{x}. \tag{4.20}$$

In order to leave (4.1) invariant, we must have

$$\eta = -[(r+2)/(m-1)]\delta. \tag{4.21}$$

This leads to a group invariant $x/t^{(\eta/\delta)}$ and, consequently, to a self-similar solution of the form

$$x = Et^{-[(r+2)/(m-1)]}, \tag{4.22}$$

where E is a solution of

$$\mu E^m - E[(r+2)/(1-m)^2][\lambda(m-1) - r - m - 1] = 0. \tag{4.23}$$

In zones I and IV, the real solutions of (4.23) are $\pm \xi_0$. The rescaling process has told us about the stability of the self-similar solutions. In zone I, not only is the self-similar solution stable, it is also an attractor for all possible initial conditions. In zone IV, the solution is obviously unstable due to the form of the potential and the sign of $\hat{\beta}$ (see Fig. 2). The ability of rescaling to give information about the stability of self-similar solutions has also been noticed in a problem of stellar dynamics.[15]

It is interesting to note that the point $(r = -2, m = 1)$ is a pivotal point for the diagram in Fig. 1. That point corresponds to a linear harmonic oscillator with a coefficient of friction that decreases as t^{-1}, in which case a stretching group exists from which the solution can be found easily. It was noticed earlier by Besnard *et al.*[9] and by Sarlet and Bahar[12] that the linear oscillator is a limiting case in the treatment of undamped motion in a power-law potential. When damping is present, the linear oscillator again appears as a special case because the rescaled potential and the transformation potential are both quadratic in ξ. This means that the conditions (3.4) and (3.5) are replaced by the single condition

$$\frac{d^2C}{dt^2} + \beta(t)\frac{dC}{dt} + \gamma(t)C = \frac{L}{C^3 \exp[2B(t)]}, \tag{4.24}$$

where L is an arbitrary constant. Then an exact energy invariant can be found for all functions $\gamma(t)$ and $\beta(t)$. In contrast, for the nonlinear power-law case $(m = 1)$, the invariant only exists when $\gamma(t)$ and $\beta(t)$ are suitably restricted.

V. CONCLUSION

We have discussed the application of rescaling transformations to dissipative systems by means of some simple examples. First we showed how to use rescaling to obtain the potentials for motion in one dimension that allow an invariant that is quadratic in the velocity and to obtain those invariants. This result has been obtained before by other means. We applied the result in detail to time-dependent power-law potentials with damping. This allowed us to obtain exact invariants for a restricted class of time dependences. Then we showed how the rescaling method could be used to obtain time-asymptotic solutions of the equations for some time dependences that did not allow the exact energy invariants.

Rescaling transformations are potentially useful in many applications. The basic idea, which we have illustrated, is to choose the transformation functions in such a way that the new equations are more readily analyzed. A possibly important application may be to equations of resistive magnetohydrodynamics. Another possibly important application is to numerical computations. In particular cases, it may be possible to transform equations to a form for which simpler, more accurate, or more stable numerical methods are available. For example, numerical solution of (4.1) may be rendered simpler by rescaling in such a way that the new coefficient of friction vanishes and the rescaled potential is θ-independent. If $\beta(t)$ and $\gamma(t)$ are slowly varying, then in the new space there will be periodic motion on which is superimposed a small perturbation due to the transformation potential.

ACKNOWLEDGMENTS

The authors thank A. Munier for an interesting discussion.

M. R. Feix thanks the Center for Nonlinear Studies and the Controlled Thermonuclear Research Division of the Los Alamos National Laboratory for financial support.

[1]W. Sarlet, Int. J. Non-Linear Mech. **18**, 259 (1983).
[2]J. R. Ray and J. L. Reid, J. Math. Phys. **21**, 2054 (1979).
[3]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[4]H. R. Lewis, J. Math. Phys. **25**, 1139 (1984).
[5]H. R. Lewis and K. R. Symon, Phys. Fluids **27**, 192 (1984).
[6]B. Abraham-Shrauner, Phys. Fluids **27**, 197 (1984).
[7]A. Munier, J. R. Burgan, M. R. Feix, and E. Fijalkow, J. Math. Phys. **22**, 1219 (1981).
[8]E. Kanai, Prog. Theor. Phys. **5**, 440 (1968).
[9]D. Besnard, J. R. Burgan, M. R. Feix, E. Fijalkow, and A. Munier, J. Math. Phys. **24**, 1123 (1983).
[10]J. R. Burgan, M. R. Feix, E. Fijalkow, and A. Munier, J. Plasma Phys. **29**, 139 (1983).
[11]J. L. Duranceau, S. Bouquet, E. Fijalkow, and M. R. Feix, Phys. Lett. A **98**, 309 (1983).
[12]W. Sarlet and L. Y. Bahar, Int. J. Non-Linear Mech. **15**, 133 (1980).
[13]J. R. Burgan, J. Gutierrez, A. Munier, E. Fijalkow, and M. R. Feix, *Strongly Coupled Plasmas*, edited by G. Kalman and P. Carini (Plenum, New York, 1978), pp. 597–639.
[14]S. Chandrasekhar, *Introduction to the Study of Stellar Structure* (Dover, New York, 1957).
[15]A. Munier and M. R. Feix, Astrophys. J. **267**, 344 (1983).

On the analysis of relaxation in electrolytes

Harold Levine
Department of Mathematics, Stanford University, Stanford, California 94305

(Received 16 April 1984; accepted for publication 15 June 1984)

Debye and Falkenhagen [Phys. Z. **29**, 121 (1928)] analyzed a linear initial/boundary value problem for a differential equation of diffusion type to model the phenomenon of relaxation in electrolytes; specifically, they sought to characterize the disappearance in time of the radially symmetrical and static charge distribution surrounding an individual motionless ion after the latter is instantaneously removed. A detailed reexamination discloses the existence of multiple solutions for the posed problem, with agreement as regards the initial condition and disparity as regards behavior at the central location. A regular solution during the entire relaxation regime is exhibited and offered in place of the classical one, due to Debye and Falkenhagen, which retains a singular nature at the site originally occupied by the reference ion.

I. INTRODUCTION

The modern theory of liquid electrolytes, given both a qualitative and quantitative basis through the perceptive approach of Debye and Hückel,[1] envisages that any individual (reference) ion has a surrounding "atmosphere" or cloud with the opposite sign of charge; for dilute solutions the spherically symmetrical and static space-charge distribution about the central ion (at $r = 0$) is obtained from the function

$$\psi(r) = C(e^{-\kappa r}/r) \tag{1}$$

which satisfies the ordinary differential equation

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) - \kappa^2\psi = 0, \quad r>0 \tag{2}$$

and vanishes in the limit $r\to\infty$. Soon thereafter Debye[2] cited the importance of examining temporal changes in this model and wrote, "Suppose a charge (of small dimensions) has been long enough at some point of the liquid, so that its ionic atmosphere has reached its equilibrium value. Now let us annihilate this central charge at the instant $t = 0$. We will ask how the charge density of the atmosphere spreads out to zero density. The mathematical expression for the charge density or the potential as a function of the distance from the center and the time turns out not to be very simple. The principal point however is that there exists an essential time constant governing the decrease."

Debye and Falkenhagen[3] presented the first analysis of a relaxation time (θ) for the whole ionic atmosphere, relying on the partial differential equation

$$\frac{1}{s^2}\frac{\partial}{\partial s}\left(s^2\frac{\partial f}{\partial s}\right) - f = \frac{\partial f}{\partial \tau}, \quad s>0, \quad \tau>0, \tag{3}$$

which involves a pair of dimensionless variables $s = \kappa r$ and $\tau = t/\theta$; the explicit solution

$$\begin{aligned} f(s,\tau) &= \frac{e^{-s}}{(\pi)^{1/2}s}\int_{\sqrt{\tau}-\frac{s}{2(\tau)^{1/2}}}^{\infty} e^{-\zeta^2}\,d\zeta \\ &= \frac{e^{-s}}{2s}\operatorname{erfc}\left(\sqrt{\tau} - \frac{s}{2(\tau)^{1/2}}\right) \end{aligned} \tag{4}$$

is stated without derivation and justified on the grounds that

$$f(s,0) = e^{-s}/s \tag{5}$$

in accord with the initial (equilibrium) form (1), and that

$$f(s,\infty) = 0$$

as the ultimate disappearance of the atmosphere implies. Falkenhagen's book[4] contains the assertion that (4) is the only solution which satisfies all the necessary conditions, namely (3) and (5). The expression (4) appears again in the published version of lectures at Harvard by Debye[5] with the appertaining statement, "We shall not give here the mathematical steps for obtaining this solution, but the fact that it satisfies the boundary conditions is easily verified."

The entire magnitude of electricity in the ionic cloud can be calculated from its density, $s^2 f(s,\tau)$, and is thus proportional to

$$Q(\tau) = \int_0^\infty s^2 f(s,\tau)\,ds; \tag{6}$$

on utilizing the distribution function (5), the initial value

$$Q(0) = \int_0^\infty s^2 f(s,0)\,ds = 1 \tag{7}$$

follows directly. Given the determination (4) and the reduction which is described in the Appendix, Part (I), two results are established, viz.

$$\begin{aligned} Q(\tau) &= \frac{1}{2} + \sqrt{\frac{\tau}{\pi}}\,e^{-\tau} - \frac{1}{(\pi)^{1/2}}\int_0^{\sqrt{\tau}} e^{-\zeta^2}\,d\zeta, \quad \tau>0 \\ &= \frac{1}{2} - \frac{2}{3(\pi)^{1/2}}\tau^{3/2} + \frac{2}{5(\pi)^{1/2}}\tau^{5/2} - \cdots, \quad \tau\to 0 \end{aligned} \tag{8}$$

and

$$\lim_{\tau\to 0} Q(\tau) = \tfrac{1}{2} \neq Q(0). \tag{9}$$

The discontinuity revealed by (9), together with the nonuniform limit relations

$$\lim_{s\to 0}\lim_{\tau\to 0} sf(s,\tau) = \lim_{s\to 0} e^{-s} = 1$$

and

$$\lim_{\tau\to 0}\lim_{s\to 0} sf(s,\tau) = \lim_{\tau\to 0}\frac{1}{(\pi)^{1/2}}\int_{\sqrt{\tau}}^{\infty} e^{-\zeta^2}\,d\zeta = \frac{1}{2}$$

prompt a reassessment of the problem *ab initio* and of the solution (4) in particular.

 0022-2488/85/010074-11$02.50

Appropriate specifications for the basic differential equation (3) comprise a single initial condition at $\tau = 0$ [namely (5)], since a first-order τ-derivative occurs, and a separate pair of boundary conditions at the endpoints $s = \infty, 0$, inasmuch as a second-order s-derivative occurs; however, Debye and Falkenhagen simply invoke the requirement that the solution must approach zero when $s \to \infty$ and omit any specific reference to the behavior at $s = 0$. Indeed, the value of $sf(s,\tau)$ when $s \to 0$ can be left arbitrary, and thus the problem posed for the relaxation time considerations admits an entire family of solutions.

A general approach which features preliminary transformation of the initial/boundary value problem is described in Sec. II and a particular solution that manifests continuity of the total charge up to the initial instant of time is found. This solution is rederived in Sec. III where the consequences of applying various direct analytical procedures are drawn. Finally, in Sec. IV the Debye–Falkenhagen solution is reconstructed and compared with other versions that assume the same initial form; furthermore, the existence of a noteworthy solution of the differential equation for the relaxation regime, namely one which has null values initially and finally, is established.

II. GENERAL ASPECTS OF THE INITIAL/BOUNDARY VALUE PROBLEM

The pertinent equation (3) which enables a study of transients in a single species electrolyte, is convertible by simple transformations to the ordinary one-dimensional diffusion equation on a half line; thus, let

$$f(s,\tau) = \phi(s,\tau)/s \tag{10}$$

and the equation for $\phi(s,\tau)$ becomes

$$\frac{\partial^2 \phi}{\partial s^2} - \phi = \frac{\partial \phi}{\partial \tau}. \tag{11}$$

Next, applying the representation

$$\phi(s,\tau) = e^{-\tau}\psi(s,\tau) \tag{12}$$

it is found that

$$\frac{\partial^2 \psi}{\partial s^2} = \frac{\partial \psi}{\partial \tau}, \quad s > 0, \tau > 0. \tag{13}$$

Standard analysis (and, specifically, the use of a Fourier sine integral) furnishes a solution of (13),

$$\psi(s,\tau) = \frac{s}{2(\pi)^{1/2}} \int_0^\tau \psi(0,\tau') \frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{3/2}} d\tau' + \frac{se^\tau}{2(\pi)^{1/2}} \int_\tau^\infty \frac{e^{-\tau'-s^2/4\tau'}}{\tau'^{3/2}} d\tau', \tag{14}$$

which satisfies the conditions

$$\begin{aligned} &\psi(s,0) = \phi(s,0) = e^{-s}, \quad s > 0, \\ &\psi(\infty,\tau) = 0, \quad \tau > 0, \end{aligned} \tag{15}$$

and allows for an arbitrary choice of the boundary value $\psi(0,\tau), \tau > 0$. The direct inferences from (14), namely

$$\lim_{s \to 0} \psi(s,\tau) = \psi(0,\tau), \quad \tau > 0$$

and

$$\psi(s,0) = \frac{s}{2(\pi)^{1/2}} \int_0^\infty \tau'^{-3/2} e^{-\tau'-s^2/4\tau'} d\tau' = e^{-s},$$

manifest its self-consistency; and the result of differentiation therein, viz.

$$\begin{aligned} \chi(\tau) &= -\frac{\partial}{\partial s}\psi(s,\tau)\Big|_{s=0} \\ &= \frac{1}{\pi^{1/2}} \frac{d}{d\tau} \int_0^\tau \frac{\psi(0,\tau')}{(\tau-\tau')^{1/2}} d\tau' + \frac{e^\tau}{\pi^{1/2}} \int_\tau^\infty \frac{d}{d\tau'}\left(\frac{1}{(\tau')^{1/2}}\right) e^{-\tau'} d\tau' \\ &= \frac{1}{\pi^{1/2}} \frac{d}{d\tau} \int_0^\tau \frac{\psi(0,\tau')}{(\tau-\tau')^{1/2}} d\tau' - \frac{1}{(\pi\tau)^{1/2}} + \frac{e^\tau}{(\pi)^{1/2}} \int_\tau^\infty \frac{e^{-\tau'}}{(\tau')^{1/2}} d\tau', \quad \tau > 0 \end{aligned} \tag{16}$$

is reserved for later comment [cf. (19)].

There exists an analogous solution of the partial differential equation (13), wherein the boundary derivative $\chi(\tau)$ takes the place of the function $\psi(0,\tau)$, with the other requirements unchanged; this is expressed by

$$\psi(s,\tau) = \frac{1}{(\pi)^{1/2}} \int_0^\tau \chi(\tau') \frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{1/2}} d\tau' + \frac{e^\tau}{\pi^{1/2}} \int_\tau^\infty e^{-\tau'-s^2/4\tau'} \frac{d\tau'}{(\tau')^{1/2}} \tag{17}$$

and the requisite properties

$$\lim_{s \to 0}\left(-\frac{\partial \psi}{\partial s}\right) = \chi(\tau), \quad \tau > 0,$$

$$\psi(s,0) = \frac{1}{\pi^{1/2}} \int_0^\infty e^{-\tau'-s^2/4\tau'} \frac{d\tau'}{(\tau')^{1/2}} = e^{-s}, \quad s > 0$$

are confirmed. The particular version of the representation (17) that emerges after making the choice $s = 0$, i.e.,

$$\psi(0,\tau) = \frac{1}{\pi^{1/2}} \int_0^\tau \frac{\chi(\tau')d\tau'}{(\tau-\tau')^{1/2}} + \frac{1}{(\pi)^{1/2}} e^\tau \int_\tau^\infty \frac{e^{-\tau'}}{(\tau')^{1/2}} d\tau', \quad \tau > 0 \tag{18}$$

constitutes an integral equation for $\chi(\tau)$; and the solution (which is obtainable by employing a conventional Laplace transform procedure)

$$\psi(\tau) = \frac{1}{\pi^{1/2}} \frac{d}{d\tau} \int_0^\tau \frac{\psi(0,\tau')}{(\tau-\tau')^{1/2}} d\tau' - \frac{2}{\pi} \int_0^\infty \frac{e^{-\tau\sigma^2}\sigma^2 d\sigma}{1+\sigma^2}, \quad \tau > 0 \tag{19}$$

agrees with (16), inasmuch as

$$\int_0^\infty \frac{e^{-\tau\sigma^2}\sigma^2 d\sigma}{1+\sigma^2} = \frac{1}{2}\sqrt{\frac{\pi}{\tau}} - \frac{1}{2}\sqrt{\pi} e^\tau \int_\tau^\infty \frac{e^{-\tau'}}{(\tau')^{1/2}} d\tau'. \tag{20}$$

A singular component of $\chi(\tau)$ in the limit $\tau \to 0$, which stems from the second term of (19), can, in principle, be removed by a particular determination of the function $\psi(0,\tau)$.

If, in fact, the representation

$$\psi(0,\tau) = \frac{e^{\tau}}{\pi^{1/2}} \int_{\tau}^{\infty} e^{-\tau'} \frac{d\tau'}{(\tau')^{1/2}} \tag{21}$$

obtained from the explicit term in (18) is utilized for evaluating the common integral in (16) and (19), the result

$$\frac{1}{\pi^{1/2}} \int_0^{\tau} \frac{\psi(0,\tau')}{(\tau-\tau')^{1/2}} d\tau' = 1 - \frac{e^{\tau}}{\pi^{1/2}} \int_{\tau}^{\infty} \frac{e^{-\tau'}}{(\tau')^{1/2}} d\tau' \tag{22}$$

implies, together with (16), that $\chi(\tau) = 0$. The ready consequence of (13), namely

$$\chi(\tau) = \frac{d}{d\tau} \int_0^{\infty} \psi(s,\tau) ds \tag{23}$$

yields, furthermore

$$\int_0^{\infty} \chi(\tau) d\tau = -\int_0^{\infty} \psi(s,0) ds = -1, \tag{24}$$

given the condition (15); and the outcome of integrating (19), namely

$$\int_0^{\infty} \chi(\tau) d\tau = -1 + \lim_{\tau\to\infty} \frac{1}{\pi^{1/2}} \int_0^{\tau} \frac{\psi(0,\tau')}{(\tau-\tau')^{1/2}} d\tau' = 0$$

is compatible with a null value of $\chi(\tau)$ when (22) holds.

The appertaining solutions of (3), generated from (14) and (17) through the relationship

$$f(s,\tau) = e^{-\tau} \psi(s,\tau)/s \tag{25}$$

comprise, in additive manner, terms that have a null initial value and the others,

$$f(s,\tau) = \frac{1}{2(\pi)^{1/2}} \int_{\tau}^{\infty} e^{-\tau' - s^2/4\tau'} \frac{d\tau'}{\tau'^{3/2}}, \tag{26}$$

$$f(s,\tau) = \frac{1}{\pi^{1/2} s} \int_{\tau}^{\infty} e^{-\tau' - s^2/4\tau'} \frac{d\tau'}{\tau'^{1/2}}, \tag{27}$$

respectively, which assume the common initial value

$$f(s,0) = e^{-s}/s, \quad s > 0;$$

a distinction between the latter as functions of s emerges, insofar as (26) is regular at $s = 0$ for all $\tau > 0$, while (27) implies that

$$\lim_{s\to 0} s f(s,\tau) = \frac{1}{\pi^{1/2}} \int_{\tau}^{\infty} e^{-\tau'} \frac{d\tau'}{(\tau')^{1/2}} \to 1, \quad \tau \to 0.$$

If the measure of charge $Q(\tau)$ defined in (6) is calculated on the basis of the expression (27) it turns out that

$$Q(\tau) = 1 + 2\sqrt{\frac{\tau}{\pi}} e^{-\tau} - \frac{2}{\pi^{1/2}} \int_0^{\sqrt{\tau}} e^{-\zeta^2} d\zeta = 1 - \frac{4}{3(\pi)^{1/2}} \tau^{3/2} + \frac{4}{5(\pi)^{1/2}} \tau^{5/2} - \dots, \quad \tau \to 0, \tag{28}$$

and thus [in contrast with (9)] the limit $Q(\tau) \to 1, \tau \to 0$ manifests a continuous transition to the initial value.

Having regard for the general expression [furnished by (14) and (25)]

$$f(s,\tau) = \frac{1}{2(\pi)^{1/2}} \int_0^{\tau} \phi(0,\tau') \times \frac{\exp(-(\tau-\tau') - s^2/4(\tau-\tau'))}{(\tau-\tau')^{3/2}} d\tau' + \frac{1}{2(\pi)^{1/2}} \int_{\tau}^{\infty} e^{-\tau' - s^2/4\tau'} \frac{d\tau'}{\tau'^{3/2}} \tag{29}$$

it is easy to establish the connection

$$Q(\tau) = e^{-\tau} + \int_0^{\tau} \phi(0,\tau') e^{-(\tau-\tau')} d\tau' \tag{30}$$

between $Q(\tau)$ and $\phi(0,\tau) = \lim_{s\to 0} s f(s,\tau)$; furthermore, (30) implies that

$$\frac{dQ}{d\tau} + Q = \phi(0,\tau)$$

and, in particular, that

$$\left(\frac{dQ}{d\tau}\right)_{\tau=0} = -1 + \phi(0,0). \tag{31}$$

The determination which follows from (27),

$$\phi(0,\tau) = \frac{2}{\pi^{1/2}} \int_{\sqrt{\tau}}^{\infty} e^{-\zeta^2} d\zeta, \tag{32}$$

yields $\phi(0,0) = 1$ and hence, $(dQ/d\tau)_{\tau=0} = 0$, as is consistent with (28).

Let

$$\mathscr{F}(s) = \int_0^{\infty} f(s,\tau) d\tau \tag{33}$$

and the relation

$$\lim_{s\to 0}\left(s^2 \frac{d\mathscr{F}}{ds}\right) + \int_0^{\infty} s^2 \mathscr{F}(s) ds = \int_0^{\infty} s^2 f(s,0) ds = 1 \tag{34}$$

is a straightforward consequence of the basic differential equation (3) for $f(s,\tau)$. Utilizing (29) it turns out that

$$
\begin{aligned}
\mathscr{F}(s) &= \int_0^\infty f(s,\tau)d\tau \\
&= \int_0^\infty \frac{\phi(s,\tau)}{s}\,d\tau \\
&= \frac{1}{2(\pi)^{1/2}}\int_0^\infty d\tau \int_\tau^\infty \frac{e^{-\tau'-s^2/4\tau'}}{\tau'^{3/2}}\,d\tau' + \frac{1}{2(\pi)^{1/2}}\int_0^\infty d\tau \int_0^\tau \phi(0,\tau')\,\frac{\exp(-(\tau-\tau')-s^2/4(\tau-\tau'))}{(\tau-\tau')^{3/2}}\,d\tau' \\
&= \frac{1}{2(\pi)^{1/2}}\int_0^\infty \frac{d\tau}{\tau^{1/2}}\,e^{-\tau-s^2/4\tau} + \frac{1}{2(\pi)^{1/2}}\int_0^\infty \phi(0,\tau')d\tau' \int_{\tau'}^\infty \frac{\exp(-(\tau-\tau')-s^2/4(\tau-\tau'))}{(\tau-\tau')^{3/2}}\,d\tau \\
&= \frac{1}{2(\pi)^{1/2}}\int_0^\infty \frac{d\tau}{\tau^{1/2}}\,e^{-\tau-s^2/4\tau} + \frac{1}{2(\pi)^{1/2}}\int_0^\infty \phi(0,\tau')d\tau' \int_0^\infty \frac{e^{-\zeta-s^2/4\zeta}}{\zeta^{3/2}}\,d\zeta \\
&= \frac{1}{2(\pi)^{1/2}}\int_0^\infty \frac{d\tau}{\tau^{1/2}}\,e^{-\tau-s^2/4\tau} - \frac{1}{2(\pi)^{1/2}}\int_0^\infty \phi(0,\tau')d\tau'\left(\frac{2}{s}\frac{d}{ds}\int_0^\infty \frac{e^{-\zeta-s^2/4\zeta}}{\zeta^{1/2}}\,d\zeta\right) \\
&= e^{-s}\left\{\frac{1}{2} + \frac{1}{s}\int_0^\infty \phi(0,\tau)d\tau\right\},
\end{aligned}
\tag{35}
$$

and the mutual compatibility of (34) and (35), regardless of the specific nature of $\phi(0,\tau)$, is confirmed.

After (17) and (25) are cited to justify the representation

$$
f(s,\tau) = \frac{1}{(\pi)^{1/2}s}\int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{(\tau')^{1/2}} + \frac{e^{-\tau}}{(\pi)^{1/2}s}\int_0^\tau \chi(\tau')\,\frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{1/2}}\,d\tau'
\tag{36}
$$

and the function $\mathscr{F}(s)$ expressed in accordance with (33), the outcome is

$$
\begin{aligned}
\mathscr{F}(s) &= \frac{1}{(\pi)^{1/2}s}\int_0^\infty d\tau \int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{(\tau')^{1/2}} + \frac{1}{(\pi)^{1/2}s}\int_0^\infty e^{-\tau}d\tau \int_0^\tau \chi(\tau')\,\frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{1/2}}\,d\tau' \\
&= \frac{1}{(\pi)^{1/2}s}\int_0^\infty \tau^{1/2}e^{-\tau-s^2/4\tau}d\tau + \frac{1}{(\pi)^{1/2}s}\int_0^\infty e^{-\tau}\frac{d}{d\tau}\int_0^\tau \chi(\tau')\,\frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{1/2}}\,d\tau'\,d\tau \\
&= \tfrac{1}{2}(1+1/s)e^{-s},
\end{aligned}
\tag{37}
$$

independently of $\chi(\tau)$, and the relations (34) and (37) evidence full agreement.

To verify that the representation (14) of $\psi(s,\tau)$ conforms with Eq. (24) which involves $\psi(\tau) = -(\partial/\partial s)\psi(s,\tau)|_{s=0}$, consider

$$
\begin{aligned}
\int_0^\infty \psi(s,\tau)d\tau &= \frac{s}{2(\pi)^{1/2}}\int_0^\infty d\tau \int_0^\tau \phi(0,\tau')\,\frac{\exp(\tau'-s^2/4(\tau-\tau'))}{(\tau-\tau')^{3/2}}\,d\tau' + \frac{s}{2(\pi)^{1/2}}\int_0^\infty e^\tau \int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{\tau'^{3/2}}\,d\tau \\
&= \frac{s}{2(\pi)^{1/2}}\int_0^\infty \phi(0,\tau')e^{\tau'}\,d\tau' \int_{\tau'}^\infty e^{-s^2/4(\tau-\tau')}\frac{d\tau}{(\tau-\tau')^{3/2}} \\
&\quad + \frac{s}{2(\pi)^{1/2}}\left\{-\int_0^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{\tau'^{3/2}} + \int_0^\infty e^{-s^2/4\tau}\frac{d\tau}{\tau^{3/2}}\right\} \\
&= \int_0^\infty e^\tau\phi(0,\tau)d\tau + 1 - e^{-s} \\
&= \int_0^\infty \psi(0,\tau)d\tau + 1 - e^{-s}
\end{aligned}
\tag{38}
$$

on the hypothesis that the latter integral exists. Differentiation in (38) with respect to s then yields, in the limit $s\to 0$, the integral condition (34).

III. PARTICULAR ANALYTICAL TECHNIQUES

The analysis in the preceding section, based on conversion of the original partial differential equation (3) to the simpler form (13), reveals the existence of whole families of solutions that satisfy a prescribed (and common) initial condition. It is instructive next to examine the results yielded by other procedures for resolving the given equation, bearing in mind the absence thus far of a direct link with the Debye–Falkenhagen solution (4).

A Laplace transform approach may be directly invoked for the system

$$
\frac{1}{s^2}\frac{\partial}{\partial s}\left(s^2\frac{\partial f}{\partial s}\right) - f = \frac{\partial f}{\partial \tau},\quad s>0,\ \tau>0
\tag{3'}
$$

$$
f(s,0) = e^{-s}/s,\quad f(s,\tau)\to 0,\quad s\to\infty
\tag{5'}
$$

with an assigned initial condition relating to the variable τ; thus, on designating

$$
\bar{f}(s,p) = \int_0^\infty e^{-p\tau}f(s,\tau)d\tau
\tag{39}
$$

it follows from (3) and (5) that

$$\left(\frac{d^2}{ds^2}+\frac{2}{s}\frac{d}{ds}-\gamma^2\right)\bar{f}=-\frac{e^{-s}}{s}, \tag{40}$$

where

$$\gamma^2=1+p. \tag{41}$$

Let

$$\bar{f}(s,p)=\bar{\phi}(s,p)/s \tag{42}$$

and

$$\left(\frac{d^2}{ds^2}-\gamma^2\right)\bar{\phi}=-e^{-s},\quad s>0. \tag{43}$$

An appropriate solution of (43) takes the form

$$\begin{aligned}\bar{\phi}(s,p)&=A\cosh\gamma s+\frac{1}{\gamma}\int_0^s e^{-s'}\{\cosh\gamma s\sinh\gamma s'-\sinh\gamma s\cosh\gamma s'\}ds'\\&=e^{\gamma s}\left\{\frac{A}{2}-\frac{1}{2\gamma}\frac{1}{\gamma+1}\right\}+e^{-\gamma s}\left\{\frac{A}{2}-\frac{1}{2\gamma}\frac{1}{\gamma-1}\right\}+e^{-s}\left\{\frac{1}{2\gamma(\gamma-1)}-\frac{1}{2\gamma(\gamma+1)}\right\}\end{aligned} \tag{44}$$

exclusive of a solution $B\sinh\gamma s$ to the corresponding homogeneous equation which generates a regular component of $\bar{f}$ at $s=0$. The arbitrary coefficient A in (44) is fixed by the asymptotic behavior $\bar{\phi}\to 0$, $s\to\infty$, whence

$$A=(1/\gamma)[1/(\gamma+1)]$$

and

$$\bar{\phi}(s,p)=e^{-s}/p-e^{-\sqrt{p+1}s}/p(p+1)^{1/2}, \tag{45}$$

having regard for the relation (41) between γ and p. On applying the inverse of the transformation (39) to the function specified by (42) and (45), it follows that

$$f(s,\tau)=\frac{1}{2\pi is}\int_C e^{p\tau}\cdot\frac{1}{p}\left\{e^{-s}-\frac{e^{-\sqrt{p+1}s}}{(p+1)^{1/2}}\right\}dp=\frac{1}{\pi s}\int_1^\infty e^{-\chi\tau}\frac{\cos\sqrt{\chi-1}s}{\chi(\chi-1)^{1/2}}d\chi \tag{46}$$

after deforming the contour C (on which $p=p_0+i\delta, p_0>0,\ -\infty<\delta<\infty$) along the sides of a branch cut at the section $-\infty<p<-1$ of the negative real p-axis. The function $\phi(s,\tau)=sf(s,\tau)$ characterized by (46) satisfies the differential equation

$$\frac{d\phi}{d\tau}=-\frac{1}{(\pi\tau)^{1/2}}e^{-\tau-s^2/4\tau},$$

whence

$$\phi(s,\tau)=\frac{1}{\pi^{1/2}}\int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{(\tau')^{1/2}}$$

and

$$f(s,\tau)=\frac{1}{\pi^{1/2}s}\int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{(\tau')^{1/2}}$$

in conformity with the prior representation (27).

If the function $B\sinh\gamma s$ is added to (44), thereby fashioning the general solution of (43), and the condition of boundedness at $s=\infty$ is invoked for the specification of A in terms of B, it turns out that

$$f(s,\tau)=\frac{1}{(\pi)^{1/2}s}\int_\tau^\infty e^{-\tau'-s^2/4\tau'}\frac{d\tau'}{\sqrt{\tau'}}+\frac{B}{2(\pi)^{1/2}}\frac{1}{\tau^{3/2}}e^{-\tau-s^2/4\tau}, \tag{47}$$

with an arbitrary coefficient B. The second term of (47) constitutes a solution of the partial differential equation (3) which is regular for all $s\geqslant 0$ when $\tau>0$; this may be termed an instantaneous source function by virtue of its singularity at $s=0$ and null value for $s>0$ at the initial instant $\tau=0$. The deduction from (47),

$$Q(\tau)=\int_0^\infty s^2 f(s,\tau)ds=1+2\sqrt{\tau/\pi}e^{-\tau}-\frac{2}{\pi^{1/2}}\int_0^{\sqrt{\tau}}e^{-\zeta^2}d\zeta+Be^{-\tau}$$

reveals that the total charge assumes the prescribed initial value $Q(0) = 1$ only when $B = 0$ and the concomitant source term is removed.

The solution of (43) which vanishes at $s = 0$, thereby precluding the existence of a singularity for the function $\bar{f}$ in (42), namely

$$\bar{\phi}(s,p) = B \sinh \gamma s + \frac{1}{\gamma}\int_0^s e^{-s'}\{\cosh \gamma s \sinh \gamma s' - \sinh \gamma s \cosh \gamma s'\} ds'$$

is rendered precise by the assignment $B = 1/\gamma(\gamma + 1)$ and this yields

$$\bar{\phi}(s,p) = e^{-s}/p - e^{-\sqrt{p+1}s}/p, \quad \bar{\phi}(\infty,p) = 0 \tag{48}$$

in place of the determination (45). On expressing the inverse of the transform function (48) it is found that

$$\begin{aligned} f(s,\tau) &= \frac{1}{\pi s}\int_1^\infty e^{-\chi\tau}\frac{\sin\sqrt{\chi - 1}s}{\chi}d\chi \\ &= \frac{1}{2(\pi)^{1/2}}\int_\tau^\infty e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}} \\ &= f_1(s,\tau) \end{aligned}$$

after differentiating the previously established relation

$$\frac{1}{\pi}\int_1^\infty e^{-\chi\tau}\frac{\cos\sqrt{\chi - 1}s}{\chi(\chi - 1)^{1/2}}d\chi = \frac{1}{\pi^{1/2}}\int_\tau^\infty e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}}$$

with respect to s.

A different means of integrating the nonhomogeneous differential equation (43) employs an auxiliary or Green's function; specifically, given the function $G(s,s')$ which satisfies the relations

$$\left(\frac{d^2}{ds^2} - \gamma^2\right)G(s,s') = -\delta(s - s'), \quad s,s' > 0,$$
$$G(0,s') = 0, \quad G(\infty,s') = 0 \tag{49}$$

and admits the explicit (symmetric) representation

$$G(s,s') = (1/\gamma)\sinh \gamma s_< e^{-s_>\gamma} = G(s',s), \tag{50}$$

it can be verified in standard fashion that

$$\begin{aligned} \bar{\phi}(s,p) &= \int_0^\infty G(s,s')e^{-s'}ds' + \bar{\phi}(0,p)\frac{d}{ds'}G(s,0) \\ &= \bar{\phi}(0,p)e^{-\sqrt{p+1}s} - e^{-s}/p - e^{-\sqrt{p+1}s}/p \end{aligned}$$

and thus, reverting to the original variables s,τ,

$$\begin{aligned} \phi(s,\tau) &= \frac{1}{2\pi i}\int_C e^{p\tau}\bar{\phi}(s,p)dp \\ &= \frac{s}{2(\pi)^{1/2}}\int_0^\tau \phi(0,\tau') \\ &\quad\times \frac{\exp(-(\tau - \tau') - s^2/4(\tau - \tau'))}{(\tau - \tau')^{3/2}}d\tau' \\ &\quad+ \frac{s}{2(\pi)^{1/2}}\int_\tau^\infty e^{-\tau' - s^2/4\tau'}\frac{d\tau'}{\tau'^{3/2}}, \end{aligned}$$

which duplicates the content of (12) and (14).

If the Green's function $G(s,s')$ is replaced by a different one, viz.

$$g(s,s') = \frac{1}{\gamma}\cosh \gamma s_< e^{-\gamma s_>}, \tag{51}$$

whose s-derivative vanishes at $s = 0$ and otherwise satisfies the same conditions as does G, the solution of (43) can be exhibited in the form

$$\begin{aligned} \bar{\phi}(s,p) &= -\frac{e^{-\gamma s}}{\gamma}\left(\frac{\partial}{\partial s}\bar{\phi}(s,p)\right)_{s=0} + \int_0^\infty g(s,s')e^{-s'}ds' \\ &= -\frac{e^{-\sqrt{p+1}s}}{(p+1)^{1/2}}\left(\frac{\partial}{\partial s}\bar{\phi}(s,p)\right)_{s=0} \\ &\quad+ \frac{e^{-s}}{p} - \frac{e^{-\sqrt{p+1}s}}{p(p+1)^{1/2}}, \end{aligned} \tag{52}$$

where the final pair of terms are precisely those entering into (45) and thus characterize the Laplace transform of the function $f(s,\tau)$ given in (27). The first term in (51) represents the Laplace transform of

$$-\frac{1}{\pi^{1/2}}\int_0^\tau \left(\frac{\partial}{\partial s}\phi(s,\tau')\right)_{s=0} \times \exp(-(\tau - \tau') - s^2/4(\tau - \tau'))\frac{d\tau'}{(\tau - \tau')^{1/2}}$$

and hence the full inverse of (52),

$$\phi(s,\tau) = \frac{1}{2\pi i}\int e^{p\tau}\bar{\phi}(s,p)dp$$

corresponds to the function described by (12) and (17).

As an alternative to the transform analysis of the system (3) and (5) with a separate and explicit initial condition, consider the single differential equation

$$\left(\nabla^2 - 1 - \frac{\partial}{\partial\tau}\right)f = -\psi\pi H(-\tau)\delta(\mathbf{s}) \tag{53}$$

whose inhomogeneous term accounts for, via the Heaviside function $H(-\tau)$, both a steady source regime ($\tau < 0$) and the abrupt disappearance of the source (at $\tau = 0$). On invoking the representation

$$H(-\tau) = \frac{1}{2\pi i}\int_{-\infty}^\infty e^{-i\omega\tau}\frac{d\omega}{\omega} = \begin{cases}1, & \tau < 0\\ 0, & \tau > 0\end{cases} \tag{54}$$

with a contour that passes below the point $\omega = 0$ and also the representation for a local singularity/source at $\mathbf{s} = 0$,

$$\delta(\mathbf{s}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{s}}\, d\mathbf{k}$$

a particular solution of (53) is readily constructed, viz.

$$f = \frac{1}{4\pi^3} \int_{-\infty}^{\infty} \frac{e^{i\mathbf{k}\cdot\mathbf{s} - i\omega\tau}}{\omega\{\omega + i(k^2+1)\}}\, d\omega\, d\mathbf{k},$$

wherein

$$k^2 = \mathbf{k}\cdot\mathbf{k}.$$

After evaluating the ω-integral it follows that

$$f = \frac{1}{2\pi^2} \int_{-\infty}^{\infty} \exp\,(i\mathbf{k}\cdot\mathbf{s} - (k^2+1)\tau)\, \frac{d\mathbf{k}}{k^2+1},$$

and thus, consequent to integrating over directions in $\mathbf{k}$ space,

$$\begin{aligned} f(s,\tau) &= \frac{2}{\pi s} \int_0^{\infty} \sin ks\, e^{-(k^2+1)\tau} \frac{k\,dk}{k^2+1} \\ &= \frac{1}{2(\pi)^{1/2}} \int_{\tau}^{\infty} e^{-\tau' - s^2/4\tau'} \frac{d\tau'}{\tau'^{3/2}}, \end{aligned} \tag{55}$$

in accord with the result

$$\begin{aligned} &\frac{\partial}{\partial\tau} \int_0^{\infty} \sin ks\, e^{-(k^2+1)\tau} \frac{k\,dk}{k^2+1} \\ &= \frac{\partial}{\partial s} \int_0^{\infty} \cos ks\, e^{-(k^2+1)\tau}\, dk \\ &= -(s/\psi)\sqrt{\pi} e^{-\tau - s^2/4\tau}/\tau^{3/2}. \end{aligned}$$

The agreement between (55) and an earlier determination (26) may be noted along with the fact that a singular behavior at $s = 0$ is manifest only when $\tau = 0$.

IV. INTERRELATIONSHIP OF SOLUTIONS AND CONCLUSIONS

The approach chosen (though not specifically presented) by Debye and Falkenhagen for securing their representation

$$f_{\mathrm{DF}}(s,\tau) = \frac{e^{-s}}{(\pi)^{1/2} s} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\, d\zeta \tag{4'}$$

is discernible from an analysis of Debye[6] concerning transient effects in a two-component electrolytic solution. His characteristically simple procedure rests on the preliminary ansatz

$$f(s,\tau) = e^{-\alpha\tau}(e^{i\kappa s}/s) \tag{56}$$

for a particular, two-parameter, separated variable solution of the differential equation (3). Inasmuch as

$$\frac{1}{s^2}\frac{d}{ds}\left(s^2 \frac{d}{ds}\right) \frac{e^{i\kappa s}}{s} = -\kappa^2 \frac{e^{i\kappa s}}{s}, \quad s > 0$$

if follows that (56) satisfies (3) provided

$$\alpha = 1 + \kappa^2$$

and thus the integral

$$f(s,\tau) = \int_{-\infty}^{\infty} A(\kappa) e^{-(1+\kappa^2)\tau} \frac{e^{i\kappa s}}{s}\, d\kappa \tag{57}$$

expresses a more general solution of the linear differential equation. A specification of the function $A(\kappa)$ is reached by enforcing the initial condition, viz.

$$\int_{-\infty}^{\infty} A(\kappa) e^{i\kappa s}\, d\kappa = \begin{cases} e^{-s}, & s > 0, \\ 0, & s < 0, \end{cases}$$

which yields

$$A(\kappa) = \frac{1}{2\pi} \int_0^{\infty} e^{-i\kappa s - s}\, ds = \frac{1}{2\pi} \frac{1}{1 + i\kappa}. \tag{58}$$

Combining (57) and (58) the representation

$$f(s,\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-(1+\kappa^2)\tau}}{1 + i\kappa} \frac{e^{+i\kappa s}}{s}\, d\kappa \tag{59}$$

emerges and this may be transformed in the following manner. Consider sf and the appertaining relation

$$\begin{aligned} \frac{\partial}{\partial s}(sf) + sf &= \frac{1}{\pi} \int_0^{\infty} e^{-(1+\kappa^2)\tau} \cos \kappa s\, d\kappa \\ &= \frac{1}{2(\pi\tau)^{1/2}} e^{-\tau - s^2/4\tau}. \end{aligned} \tag{60}$$

Since

$$\begin{aligned} g(\tau) = \lim_{s\to 0}(sf) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-(1+\kappa^2)\tau}}{1+\kappa^2}(1 - i\kappa)\, d\kappa \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-(1+\kappa^2)\tau}}{1+\kappa^2}\, d\kappa \end{aligned} \tag{61}$$

by reason of symmetry and, moreover,

$$\begin{aligned} \frac{dg}{d\tau} &= -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(1+\kappa^2)\tau}\, d\kappa \\ &= -\frac{1}{2(\pi\tau)^{1/2}} e^{-\tau}, \quad \tau > 0 \end{aligned}$$

the subsequent determination

$$g(\tau) = \frac{1}{\pi^{1/2}} \int_{\sqrt{\tau}}^{\infty} e^{-\zeta^2}\, d\zeta \tag{62}$$

provides the integration constant (relative to s) which fixes a unique solution of the differential equation (60). Thus, recasting (60) in the form

$$\begin{aligned} \frac{\partial}{\partial s}(sfe^s) &= \frac{1}{2} \frac{e^{-\tau}}{(\pi\tau)^{1/2}} e^{s - s^2/4\tau} \\ &= \frac{1}{2(\pi\tau)^{1/2}} e^{-(s-2\tau)^2/4\tau} \end{aligned}$$

and integrating

$$\begin{aligned} sf(s,\tau)e^s &= g(\tau) + \frac{1}{2(\pi\tau)^{1/2}} \int_0^s e^{-(s-2\tau)^2/4\tau}\, ds \\ &= \frac{1}{\pi^{1/2}} \int_{\sqrt{\tau}}^{\infty} e^{-\zeta^2}\, d\zeta \\ &\quad + \frac{1}{\pi^{1/2}} \int_{-\sqrt{\tau}}^{s/2\sqrt{\tau} - \sqrt{\tau}} e^{-\zeta^2}\, d\zeta \end{aligned} \tag{63}$$

after the change of variable $s - 2\tau = 2\sqrt{\tau}\zeta$ is employed. Inasmuch as

$$\int_{-a}^{b} e^{-\zeta^2}\, d\zeta = \int_{-b}^{a} e^{-\zeta^2}\, d\zeta$$

the Debye–Falkenhagen representation follows immediately from (63).

Let the variable τ' in the expression

$$f_2(s,\tau) = \frac{1}{(\pi)^{1/2}s}\int_\tau^\infty e^{-\tau' - s^2/4\tau'}\frac{d\tau'}{(\tau')^{1/2}} \tag{27'}$$

be replaced by ζ, where

$$\sqrt{\tau'} - \frac{s}{2(\tau')^{1/2}} = \zeta,$$

and thus

$$\sqrt{\tau'} = \tfrac{1}{2}(\zeta + \sqrt{\zeta^2 + 2s})$$

$$\frac{d\tau'}{(\tau')^{1/2}} = \left(1 + \frac{\zeta}{(\zeta^2 + 2s)^{1/2}}\right)d\zeta;$$

accordingly,

$$\begin{aligned} f_2(s,\tau) &= \frac{e^{-s}}{(\pi)^{1/2}s}\int_\tau^\infty \exp\left\{-\left(\sqrt{\tau'} - \frac{s}{2(\tau')^{1/2}}\right)^2\right\}\frac{d\tau'}{(\tau')^{1/2}} \\ &= \frac{e^{-s}}{(\pi)^{1/2}s}\int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2}\left(1 + \frac{\zeta}{(\zeta^2 + 2s)^{1/2}}\right)d\zeta \\ &= f_{\mathrm{DF}}(s,\tau) + f^*(s,\tau) \end{aligned} \tag{64}$$

with

$$\begin{aligned} f^*(s,\tau) &= \frac{e^{-s}}{(\pi)^{1/2}s}\int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2}\frac{\zeta}{(\zeta^2 + 2s)^{1/2}}d\zeta \\ &= (e^s/2s)\,\mathrm{erfc}(\sqrt{\tau} + s/2(\tau)^{1/2}). \end{aligned} \tag{65}$$

It is fitting that the function f^* which represents the difference between a pair of others with the same initial and final values ($\tau = 0, \infty$) vanishes itself at these epochs. A straightforward calculation [Appendix II] establishes

$$\phi^*(s,\tau) = sf^*(s,\tau) = \frac{e^{-s}}{\pi^{1/2}}\int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2}\frac{\zeta}{(\zeta^2 + 2s)^{1/2}}d\zeta$$

as a solution of the partial differential equation

$$\frac{\partial^2}{\partial s^2}\phi^* - \phi^* = \frac{\partial \phi^*}{\partial \tau},$$

whence f^* satisfies the basic equation (3) underlying the relaxation theory analysis.

An appraisal of results, now in order, affirms the earlier contention that Debye and Falkenhagen dealt with an imperfectly posed mathematical problem: The single initial condition which they held to be definitive is merely a common feature of solutions that possess different behaviors at the central or source point. The contrasting features of three definite and distinct solutions $f_1(s,\tau)$, $f_2(s,\tau)$, and $f_{\mathrm{DF}}(s,\tau)$ for small values of τ merits note; thus

$$\begin{aligned} f_1(s,\tau) &= \frac{1}{2(\pi)^{1/2}}\int_\tau^\infty e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}} \\ &= \frac{e^{-s}}{s} - \frac{1}{2(\pi)^{1/2}}\int_0^\tau e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}} \\ &\doteq \frac{e^{-s}}{s} - \frac{1}{2(\pi)^{1/2}}\int_0^\tau (1-\sigma)e^{-s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}}, \quad \tau \ll 1 \end{aligned} \tag{26'}$$

and, pursuant to the change of variable $\sigma = s^2/4\zeta^2$,

$$\begin{aligned} f_1(s,\tau) &= \frac{e^{-s}}{s} - \frac{2}{(\pi)^{1/2}s}\int_{s/2\sqrt{\tau}}^\infty \left(1 - \frac{s^2}{4\zeta^2}\right)e^{-\zeta^2}d\zeta \\ &= \frac{e^{-s}}{s} - \frac{2}{(\pi)^{1/2}s}\left(1 + \frac{s^2}{2}\right)\int_{s/2\sqrt{\tau}}^\infty e^{-\zeta^2}d\zeta \\ &\quad + \sqrt{\frac{\tau}{\pi}}\, e^{-s^2/4\tau}, \quad \tau \to 0. \end{aligned} \tag{66}$$

The latter estimate implies that

$$\lim_{\tau\to 0} f_1(s,\tau) = e^{-s}/s, \quad s > 0,$$

$$\lim_{s\to 0} sf_1(s,\tau) = 0, \quad \tau > 0,$$

whence f_1 exhibits a regular nature at $s = 0$ for $\tau > 0$; furthermore, the respective magnitudes

$$Q(\tau) = \int_0^\infty s^2 f_1(s,\tau)ds = e^{-\tau}, \quad \tau > 0$$

and

$$Q(\tau) \doteq 1 - \tau + \tfrac{1}{2}\tau^2, \quad \tau \ll 1$$

arrived at on the basis of (26) and (66) are manifestly consistent. Next, the function

$$\begin{aligned} f_2(s,\tau) &= \frac{1}{(\pi)^{1/2}s}\int_\tau^\infty e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}} \\ &= \frac{e^{-s}}{s} - \frac{1}{(\pi)^{1/2}s}\int_0^\tau e^{-\sigma - s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}} \\ &\doteq \frac{e^{-s}}{s} - \frac{1}{(\pi)^{1/2}s}\int_0^\tau (1-\sigma)e^{-s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}}, \quad \tau \ll 1 \end{aligned} \tag{27'}$$

is converted, in the manner employed for $f_1(s,\tau)$ to

$$\begin{aligned} f_2(s,\tau) &= \frac{e^{-s}}{s} - 2\sqrt{\frac{\tau}{\pi}}\,\frac{1}{s}\left(1 + \frac{s^2}{6}\right)e^{-s^2/4\tau} \\ &\quad + \frac{2}{3(\pi)^{1/2}}\frac{\tau^{3/2}}{s}e^{-s^2/4\tau} \\ &\quad + \frac{2}{\pi^{1/2}}\left(1 + \frac{s^2}{6}\right)\int_{s/2\sqrt{\tau}}^\infty e^{-\zeta^2}d\zeta, \quad \tau \to 0 \end{aligned} \tag{67}$$

and the deductions therefrom

$$\lim_{\tau\to 0} f_2(s,\tau) = \frac{e^{-s}}{s}, \quad s > 0$$

$$\lim_{s\to 0} sf_2(s,\tau) \doteq 1 - 2\sqrt{\frac{\tau}{\pi}} + \frac{2}{3(\pi)^{1/2}}\tau^{3/2}, \quad \tau \to 0$$

along with

$$\lim_{\tau\to 0}\lim_{s\to 0} sf_2(s,\tau) = \lim_{s\to 0}\lim_{\tau\to 0} sf_2(s,\tau) = 1$$

make plain the uniform continuity of $f_2(s,\tau)$ relative to the variables s,τ. The three term estimate in (28) for $Q(\tau)$ at small values of τ is correctly reproduced by integrating the pertinent representation (67) of $f_2(s,\tau)$.

A version of the Debye–Falkenhagen function,

$$f_{\mathrm{DF}}(s,\tau) = \frac{e^{-s}}{s} - \frac{e^{-s}}{(\pi)^{1/2}s}\int_{s/2\sqrt{\tau} - \sqrt{\tau}}^\infty e^{-\zeta^2}d\zeta,$$

allows an estimate

$$f_{\mathrm{DF}}(s,\tau) \doteq \frac{e^{-s}}{s} - \sqrt{\frac{\tau}{\pi}}\,\frac{1}{s^2}e^{-s^2/4\tau}, \quad \tau \to 0 \tag{68}$$

conditional on the inequality $s \gg 2\sqrt{\tau}$; and since the limits $\tau\to 0$, $s \to 0$ are not interchangeable in respect of f_{DF} a direct counterpart of the uniform estimate (67) for $f_2(s,\tau)$ is lacking.

The realization of a singularity at $s = 0$ in the general expression (29) for $f(s,\tau)$ depends on the first term therein, inasmuch as the second is regular with $\tau > 0$. Let the function

$\phi(0,\tau)$ in (29) be assigned the specific form

$$\phi(0,\tau) = \lim_{s\to 0} sf_2(s,\tau) = \frac{1}{\pi^{1/2}}\int_\tau^\infty e^{-\sigma}\frac{d\sigma}{\sqrt{\sigma}},\quad \phi(0,0)=1, \tag{69}$$

which follows from (36) if $\chi(\tau)=0$; then the appertaining convolution

$$\begin{aligned}\mathscr{F}(s,\tau) &= \int_0^\tau \phi(0,\tau')\exp\left\{-(\tau-\tau')-\frac{s^2}{4(\tau-\tau')}\right\}\frac{d\tau'}{(\tau-\tau')^{3/2}} \\ &= \frac{4}{s}\int_{s/2\sqrt{\tau}}^\infty e^{-\zeta^2-s^2/4\zeta^2}\phi\left(0,\tau-\frac{s^2}{4\zeta^2}\right)d\zeta \end{aligned}\tag{70}$$

has a derivative

$$\begin{aligned}\frac{\partial\mathscr{F}}{\partial\tau} &= \frac{1}{\tau^{3/2}}e^{-\tau-s^2/4\tau} - \frac{4}{(\pi\tau)^{1/2}s}e^{-\tau}\int_{s/2\sqrt{\tau}}^\infty e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2-s^2/4\tau)^{1/2}} \\ &= e^{-\tau-s^2/4\tau}\left(\frac{1}{\tau^{3/2}}-\frac{2}{s\tau^{1/2}}\right),\end{aligned}$$

whose reduction is accomplished with the help of the result

$$\begin{aligned}\int_\alpha^\infty e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2-\alpha^2)^{1/2}} &= \frac{1}{2}\int_{\alpha^2}^\infty e^{-\chi}\frac{d\chi}{\sqrt{\chi-\alpha^2}} = \int_0^\infty e^{-(\alpha^2+\sigma^2)}d\sigma \\ &= \frac{1}{2}\sqrt{\pi}e^{-\alpha^2}.\end{aligned}\tag{71}$$

Hence,

$$\mathscr{F}(s,\tau) = \int_0^\tau e^{-\sigma-s^2/4\sigma}\left(\frac{1}{\sigma^{3/2}}-\frac{2}{s\sigma^{1/2}}\right)d\sigma,\quad \mathscr{F}(s,0)=0 \tag{72}$$

and it follows from (29) and (72) that

$$\begin{aligned}f(s,\tau) &= \frac{1}{2(\pi)^{1/2}}\int_0^\infty e^{-\sigma-s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}} - \frac{1}{(\pi)^{1/2}s}\int_0^\tau e^{-\sigma-s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}} \\ &= \frac{e^{-s}}{s} - \frac{1}{(\pi)^{1/2}s}\int_0^\tau e^{-\sigma-s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}} \\ &= \frac{1}{(\pi)^{1/2}s}\int_\tau^\infty e^{-\sigma-s^2/4\sigma}\frac{d\sigma}{\sigma^{1/2}} = f_2(s,\tau)\end{aligned}$$

with the anticipated singularity at $s=0$.

Alternatively, let the function

$$\chi(\tau) = -\lim_{s\to 0}\frac{\partial}{\partial s}\left(e^\tau\phi(s,\tau)\right),$$

which enters into the general representation (36) for $f(s,\tau)$ be specified through the expression

$$\phi(s,\tau) = \frac{s}{2(\pi)^{1/2}}\int_\tau^\infty e^{-\sigma-s^2/4\sigma}\frac{d\sigma}{\sigma^{3/2}}$$

that is yielded by (29) when $\phi(0,\tau)$ and the first term are both given null values. Thus,

$$\begin{aligned}\chi(\tau) &= -\frac{e^\tau}{2\sqrt{\pi}}\int_\tau^\infty e^{-\sigma}\frac{d\sigma}{\sigma^{3/2}} \\ &= -\frac{1}{(\pi\tau)^{1/2}} + \frac{e^\tau}{\pi^{1/2}}\int_\tau^\infty\frac{e^{-\sigma}}{\sigma^{1/2}}d\sigma,\quad \tau>0 \\ &= -\frac{1}{(\pi\tau)^{1/2}} + e^\tau\phi(0,\tau),\end{aligned}\tag{73}$$

where $\phi(0,\tau)$ designates the function (69). A two-part reduction of the appertaining integral

$$e^{-\tau}\int_0^\tau \chi(\tau')\frac{e^{-s^2/4(\tau-\tau')}}{(\tau-\tau')^{1/2}}d\tau'$$

is indicated, and the result

$$\begin{aligned}\int_0^\tau\frac{1}{(\tau')^{1/2}}\frac{e^{-s^2/4(\tau-\tau')^{1/2}}}{(\tau-\tau')^{1/2}}d\tau' &= \frac{s}{\tau^{1/2}}\int_{s/2\sqrt{\tau}}^\infty\frac{e^{-\zeta^2}d\zeta}{\zeta(\zeta^2-s^2/4\tau)^{1/2}} \\ &= \frac{s}{2(\tau)^{1/2}}\int_{s^2/4\tau}^\infty\frac{e^{-\chi}d\chi}{\chi(\chi-s^2/4\tau)^{1/2}}\end{aligned}\tag{74}$$

bears directly thereupon. Furthermore, if

$$\begin{aligned}\mathscr{H}(s,\tau) &= \int_0^\tau\phi(0,\tau')\exp\{-(\tau-\tau')-s^2/4(\tau-\tau')\}\frac{d\tau'}{(\tau-\tau')^{1/2}} \\ &= s\int_{s/2\sqrt{\tau}}^\infty e^{-\zeta^2-s^2/4\zeta^2}\phi\left(0,\tau-\frac{s^2}{4\zeta^2}\right)\frac{d\zeta}{\zeta^2},\end{aligned}\tag{75}$$

then

$$\frac{\partial\mathscr{H}}{\partial\tau} = \frac{1}{\tau^{1/2}}e^{-\tau-s^2/4\tau} - \frac{se^{-\tau}}{2(\pi\tau)^{1/2}}\int_{s^2/4\tau}^\infty\frac{e^{-\chi}d\chi}{\chi(\chi-s^2/4\tau)^{1/2}}. \tag{76}$$

The common integral in (74) and (76) is dealt with by first defining

$$I(\lambda) = \int_{\alpha^2}^\infty\frac{e^{-\lambda\chi}}{\chi(\chi-\alpha^2)^{1/2}}d\chi$$

and noting that [cf. (71)]

$$\frac{dI}{d\lambda} = -\int_{\alpha^2}^\infty\frac{e^{-\lambda\chi}}{(\chi-\alpha^2)^{1/2}}d\chi = -\sqrt{\frac{\pi}{\lambda}}e^{-\lambda\alpha^2},$$

whence

$$I(\lambda) = \sqrt{\pi}\int_\lambda^\infty e^{-\alpha^2\zeta}\frac{d\zeta}{\zeta^{1/2}} = 2\frac{\sqrt{\pi}}{\alpha}\int_{\alpha\sqrt{\lambda}}^\infty e^{-\sigma^2}d\sigma$$

and, in particular,

$$I(1) = \int_{\alpha^2}^\infty\frac{e^{-\chi}d\chi}{\chi\sqrt{\chi-\alpha^2}} = 2\frac{\sqrt{\pi}}{\alpha}\int_\alpha^\infty e^{-\sigma^2}d\sigma. \tag{77}$$

On combining (76) and (77) there obtains

$$\frac{\partial\mathscr{H}}{\partial\tau} = \frac{1}{\tau^{1/2}}e^{-\tau-s^2/4\tau} - 2e^{-\tau}\int_{s/2\sqrt{\tau}}^\infty e^{-\sigma^2}d\sigma$$

and, accordingly,

$$\mathscr{H}(s,\tau)$$
$$= \int_0^\tau e^{-\sigma - s^2/4\sigma} \frac{d\sigma}{\sigma^{1/2}} - 2\int_0^\tau e^{-\zeta} \int_{s/2\sqrt{\zeta}}^\infty e^{-\sigma^2} d\sigma\, d\zeta$$
$$= \int_0^\tau e^{-\sigma - s^2/4\sigma} \frac{d\sigma}{\sigma^{1/2}}$$
$$+ 2e^{-\tau} \int_{s/2\sqrt{\tau}}^\infty e^{-\sigma^2} d\sigma - \frac{s}{2} \int_0^\tau e^{-\sigma - s^2 4\sigma} \frac{d\sigma}{\sigma^{3/2}}. \quad (78)$$

Thus, the representation (36) yields

$$f(s,\tau) = \frac{e^{-s}}{s} - \frac{1}{2(\pi)^{1/2}} \int_0^\tau e^{-\sigma - s^2/4\sigma} \frac{d\sigma}{\sigma^{3/2}}$$
$$= \frac{1}{2(\pi)^{1/2}} \int_\tau^\infty e^{-\sigma - s^2/4\sigma} \frac{d\sigma}{\sigma^{3/2}} = f_1(s,\tau)$$

when the results (73), (74), (75), and (77) are taken into account.

In summary, the preceding analysis establishes the fact that the solution proposed by Debye and Falkenhagen belongs to a trio, $f_1(s,\tau)$, $f_2(s,\tau)$, and $f_{\rm DF}(s,\tau)$, all of which comply with the same prescriptions at $\tau = 0$, $\tau = \infty$, and $s = \infty$; $f_1(s,\tau)$ alone is devoid of singularity at $s = 0$ when $\tau > 0$ and may therefore lay claim to a preferential status during the transient regime envisaged.

ACKNOWLEDGMENTS

Thanks are due to Professor M. Howe for suggesting a particular formulation and to Dr. W. Möhring for supplying a pertinent reference.

APPENDIX

(I) Let the Debye–Falkenhagen representation (4) for $f(s,\tau)$ be utilized in connection with the volume charge integral

$$Q(s,\tau) = \int_0^s s^2 f(s,\tau) ds;$$

then, pursuant to integration by parts,

$$Q(s,\tau) = \frac{1}{\pi^{1/2}} \int_0^s d(-e^{-s}) \left\{ s \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta \right\}$$
$$= -\frac{s}{\pi^{1/2}} e^{-s} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta + \frac{1}{\pi^{1/2}} \int_0^s e^{-s} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta\, ds$$
$$+ \frac{1}{\pi^{1/2}} \int_0^s e^{-s} \cdot \frac{s}{2(\tau)^{1/2}} \exp\left\{ -\left(\tau - s + \frac{s^2}{4\tau} \right) \right\} ds$$
$$= -\frac{se^{-s}}{\pi^{1/2}} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta + \frac{1}{\pi^{1/2}} \int_0^s d(-e^{-s}) \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta + \frac{e^{-\tau}}{2(\pi\tau)^{1/2}} \int_0^s s e^{-s^2/4\tau} ds$$
$$= \frac{1}{\pi^{1/2}} \int_{\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta - \frac{(1+s)e^{-s}}{\pi^{1/2}} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta + \sqrt{\frac{\tau}{\pi}} e^{-\tau} (1 - e^{-s^2/4\tau}),$$

whence

$$Q(\infty,\tau) = \frac{1}{\pi^{1/2}} \int_{\sqrt{\tau}}^\infty e^{-\zeta^2} d\zeta + \sqrt{\frac{\tau}{\pi}} e^{-\tau}$$
$$= \frac{1}{2} + \sqrt{\frac{\tau}{\pi}} e^{-\tau} - \frac{1}{\pi^{1/2}} \int_0^{\sqrt{\tau}} e^{-\zeta^2} d\zeta,$$

as stated in (8).

The corresponding expressions generated by the choice (27) for the function $f(s,\tau)$ are

$$Q(s,\tau) = \frac{2}{\pi^{1/2}} \int_\tau^\infty e^{-\tau'} \sqrt{\tau'} (1 - e^{-s^2/4\tau'}) d\tau'$$

and

$$Q(\infty,\tau) = \frac{4}{\pi^{1/2}} \int_{\sqrt{\tau}}^\infty \zeta^2 e^{-\zeta^2} d\zeta = -\frac{4}{\pi^{1/2}} \frac{d}{da} \left(\frac{1}{2} \sqrt{\frac{\pi}{a}} \right)_{a=1} - \frac{4}{\pi^{1/2}} \int_0^{\sqrt{\tau}} \zeta d\left(-\frac{1}{2} e^{-\zeta^2} \right)$$
$$= 1 + 2\sqrt{\frac{\tau}{\pi}} e^{-\tau} - \frac{2}{\pi^{1/2}} \int_0^{\sqrt{\tau}} e^{-\zeta^2} d\zeta,$$

in agreement with (28).

(II) A direct verification that

$$\sqrt{\pi} \phi(s,\tau) = e^{-s} \int_{\sqrt{\tau} - s/2\sqrt{\tau}}^\infty e^{-\zeta^2} \frac{\zeta}{(\zeta^2 + 2s)^{1/2}} d\zeta$$

satisfies the partial differential equation

$$\frac{\partial^2 \phi}{\partial s^2} - \phi = \frac{\partial \phi}{\partial \tau},$$

relies on the individual determinations

$$\sqrt{\pi}\,\frac{\partial \phi}{\partial s} = -e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{1/2}} - e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{3/2}}$$
$$+\frac{1}{2(\tau)^{1/2}}e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{\tau^{1/2}+s/2(\tau)^{1/2}},$$

$$\sqrt{\pi}\,\frac{\partial^2 \phi}{\partial s^2} = e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{1/2}} + e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{3/2}} - \frac{1}{2(\tau)^{1/2}}e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{\tau^{1/2}+s/2(\tau)^{1/2}}$$
$$+e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{3/2}} + 3e^{-s}\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{5/2}}$$
$$-\frac{1}{2(\tau)^{1/2}}e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{(\tau^{1/2}+s/2(\tau)^{1/2})^3}$$
$$-\frac{s}{4\tau^{3/2}}e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{\tau^{1/2}+s/2(\tau)^{1/2}} - \frac{1}{4\tau}e^{-(\tau+s^2/4\tau)}.\frac{1}{\tau^{1/2}+s/2(\tau)^{1/2}}$$
$$-\frac{1}{4\tau}e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{(\tau^{1/2}+s/2(\tau)^{1/2})^2},$$

and

$$\sqrt{\pi}\,\frac{\partial \phi}{\partial \tau} = -e^{-(\tau+s^2/4\tau)}.\frac{\sqrt{\tau}-s/2\sqrt{\tau}}{\tau^{1/2}+s(2\tau)^{1/2}}.\left\{\frac{1}{2(\tau)^{1/2}}+\frac{s}{4\tau^{3/2}}\right\}.$$

Inasmuch as

$$3\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{5/2}} = \int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\,d\{-(\zeta^2+2s)^{-3/2}\}$$
$$= e^{-(\tau-s+s^2/4\tau)}.\frac{1}{(\tau^{1/2}+s/2(\tau)^{1/2})^3} - 2\int_{\sqrt{\tau}-s/2\sqrt{\tau}}^{\infty} e^{-\zeta^2}\frac{\zeta\,d\zeta}{(\zeta^2+2s)^{3/2}}$$

it follows that

$$\sqrt{\pi}\left(\frac{\partial^2 \phi}{\partial s^2} - \phi - \frac{\partial \phi}{\partial \tau}\right)$$
$$= e^{(\tau+s^2/4\tau)}.\left\{\frac{\frac{1}{2}+s/4\tau}{(\tau^{1/2}+s/2(\tau)^{1/2})^3} - \frac{1/2\sqrt{\tau}}{(\tau^{1/2}+s/2(\tau)^{1/2})^2}\right\}$$
$$= e^{-(\tau+s^2/4\tau)}.\frac{1}{(\tau^{1/2}+s/2(\tau)^{1/2})^3}\left\{\frac{1}{2}+\frac{s}{4\tau}-\frac{1}{2(\tau)^{1/2}}\left(\sqrt{\tau}+\frac{s}{2(\tau)^{1/2}}\right)\right\}$$
$$= 0.$$

Q.E.D.

[1]P. Debye and E. Hückel, Phys. Z. **24**, 185 (1923).
[2]P. Debye, Trans. Faraday Soc. **23**, 334 (1927).
[3]P. Debye and H. Falkenhagen, Phys. Z. **29**, 121 (1928).
[4]H. Falkenhagen, *Electrolytes* (Oxford U. P., New York, 1934).
[5]A. Prock and G. McConkey, *Topics in Chemical Physics*, based on the Harvard Lectures of Peter J. W. Debye (Elsevier, New York, 1962).
[6]P. Debye, *Sommerfeld Festschrift* (Leipzig, 1928), p. 52.

Cauchy system for the resolvent of Milne's integral equation with anisotropic scattering

S. Ueno
Information Science Laboratory, Kanazawa Institute of Technology, Nonoichimachi Ishikawa 921, Japan

(Received 11 April 1980; accepted for publication 19 July 1984)

It has been shown previously that the resolvent of the inhomogeneous truncated Milne's integral equation with isotropic scattering enables us to reduce dimensionally the above auxiliary equation via Chandrasekhar's X- and Y-functions. In the present paper, extending the above procedure to the case of anisotropic scattering, we show how to use effectively the Bellman–Krein–Sobolev-like formula for the dimensional reduction of the Cauchy system of the source function via the generalized Chandrasekhar X- and Y-functions.

I. INTRODUCTION

It is well known[1–4] that an inhomogeneous truncated Milne's integral equation for the source function, i.e., the auxiliary equation, plays an important role in the theory of radiative transfer. In the case of isotropic scattering, making use of the Fredholm resolvent, an initial-value solution of the auxiliary equation was reduced to the Cauchy system of Chandrasekhar's X- and Y-functions, which fulfill the Riccati-type of integrodifferential equations.[5,6] In the case of anisotropic scattering, it usually has been the case that invariant imbedding of the source function enabled us to convert an initial-value solution of the auxiliary equation to the Cauchy system of the scattering function and that of the X- and Y-functions.[2,4,6,7] In this paper, extending our preceding procedure[5] to the case of anisotropic scattering, we show how to use powerfully an initial-value solution of the two-dimensional resolvent of the auxiliary equation, i.e., the Bellman–Krein–Sobolev-like formula, for the dimensional reduction via the generalized X- and Y-functions. In our preceding papers,[8–10] it was shown that, with the aid of invariant imbedding, the scattering fucntion of Chandrasekhar's planetary problems with the diffuse (or specular) reflector was computed. In our subsequent paper, making use of the present procedure, we shall show how to get an initial-value solution of Chandrasekhar's planetary problems with hybrid reflectors.

II. BASIC EQUATIONS

Consider the family of generalized Milne's integral equations with anisotropic scattering in a turbid slab:

$$f(t,v,x) = g(t,v) + \lambda \overline{\Lambda}_{t,v}\{f(y,w,x)\}, \tag{1}$$

where $0 \leqslant t \leqslant x$, $-1 \leqslant v \leqslant 1$, $g(t,v)$ is the forcing function, λ $(0 \leqslant \lambda \leqslant 1)$ is the constant parameter, and $\overline{\Lambda}$ is the two-dimensional truncated Hopf's operator

$$\overline{\Lambda}_{t,v}\{f(y,w)\} = \int_0^x \int_0^1 E(t,y;v,w) f(y,w) dy\, dw. \tag{2}$$

In Eq. (2) the generalized exponential function is given by

$$E(t,y;v,w) = \exp[-(|t-y|)/w] P(v,w)/w, \tag{3}$$

where the phase function $P(v,w)$ in radiative transfer satisfies the local principle of reciprocity, i.e.,

$$P(v,w) = P(w,v), \tag{4}$$

$$P(-v,w) = P(v,-w) = P(w,-v), \tag{5}$$

$$P(-v,-w) = P(v,w), \tag{6}$$

where $0 \leqslant v,w \leqslant 1$.

In Eq. (2) the plus or minus sign of w corresponds to the interval length $0 \leqslant y \leqslant t \leqslant x$ or $0 \leqslant t \leqslant y \leqslant x$, respectively. In other words, Eq. (1) takes the form

$$f(t,v,x) = g(t,v) + \lambda \int_t^x \int_0^1 E(t,y;v,-w) f(y,-w,x) dy\, dw + \lambda \int_0^t \int_0^1 E(t,y;v,w) f(y,w,x) dy\, dw. \tag{7}$$

It is assumed that the interval length x is so sufficiently small that Eq. (1) has a unique solution. Furthermore, it should be mentioned that the E-function is shift-invariant with respect to the geometric argument, whereas it is asymmetric with respect to the angular argument.

Suppose that the Fredholm resolvent $K(t,y;v,w;x)$ does exist and it satisfies the following integral equations:

$$K(t,y;v,u;x) = \lambda E(t,y;v,u) + \lambda \int_0^x \int_0^1 K(t,z;v,w;x) E(z,y;w,u) dz\, dw, \tag{8}$$

and

$$K(t,y;v,u;x) = \lambda E(t,y;v,u) + \lambda \int_0^x \int_0^1 E(t,z;v,w) K(z,y;w,u;x) dz\, dw, \tag{9}$$

where $0 \leqslant t,y \leqslant x$, $-1 \leqslant v,u \leqslant 1$.

With the aid of the resolvent K, the solution of Eq. (1) takes the form

$$f(t,v,x) = g(t,v) + \int_0^x \int_0^1 K(t,y;v,w;x) g(y,w) dy\, dw. \tag{10}$$

Upon differentiation of Eq. (1) with respect to x, Eq. (1) becomes

$$f_x(t,v,x) = \lambda \int_0^1 E(t,x;v,w) f(x,w,x) dw$$

 0022-2488/85/010085-04$02.50

$$+\lambda \int_0^x \int_0^1 E(t,y;v,w) f_x(y,w,x)\,dy\,dw, \tag{11}$$

where the subscript x represents the differentiation. On introducing the Φ-function, which governs the integral equation

$$\Phi(t,v,u,x) = \lambda E(t,x;v,u) + \lambda \int_0^x \int_0^1 E(t,y;v,w)\Phi(y,w,u,x)\,dy\,dw, \tag{12}$$

it is seen that

$$f_x(t,v,x) = \int_0^1 \Phi(t,v,w,x) f(x,w,x)\,dw. \tag{13}$$

From Eq. (10) it follows that

$$f(x,w,x) = g(x,w) + \int_0^x \int_0^1 K(x,y;w,u;x) g(y,u)\,dy\,du. \tag{14}$$

Equation (13) becomes

$$f_x(t,v,x) = \int_0^1 \Phi(t,v,w,x)\,dw \Big[g(x,w) + \int_0^x \int_0^1 K(x,y;w,u;x) g(y,u)\,dy\,du \Big]. \tag{15}$$

On the other hand, Eq. (10) may be differentiated with respect to x to get

$$f_x(t,v,x) = \int_0^1 K(t,x;v,w;x) g(x,w)\,dw + \int_0^x \int_0^1 K_x(t,y;v,w;x) g(y,w)\,dy\,dw. \tag{16}$$

From comparison of Eqs. (15) and (16) it follows that

$$\Phi(t,v,u,x) = K(t,x;v,u;x), \tag{17}$$

$$K_x(t,y;v,u;x) = \int_0^1 \Phi(t,v,w,x) K(x,y;w,u;x)\,dw, \tag{18}$$

$$K_x(t,y;v,u;x) = \int_0^1 K(t,x;v,w;x) K(x,y;w,u;x)\,dw. \tag{19}$$

Equation (19) should be solved subject to the initial conditions

$$K(t,y;v,u;y) = \Phi(t,v,u,y), \quad \text{for } 0 \leqslant t < y, \tag{20}$$

$$K(t,y;v,u;t) = \Phi(y,v,u,t), \quad \text{for } 0 < y \leqslant t, \tag{21}$$

and Eqs. (17) and (19) are the desired relations. Equation (19) is similar in form to the Bellman–Krein–Sobolev formula.[11–13]

III. CAUCHY SYSTEM FOR Φ-FUNCTION

In a manner similar to Eq. (1), we introduce an auxiliary function $J(t,v,u,x)$:

$$J(t,v,u,x) = \lambda F(x,t;v,u) + \lambda \int_0^x \int_0^1 E(t,y;v,w) J(y,w,u,x)\,dy\,dw, \tag{22}$$

$$0 \leqslant t \leqslant x, \quad -1 \leqslant v \leqslant 1, \quad 0 \leqslant u \leqslant 1,$$

where

$$F(x,t;v,u) = \exp[-(x-t)/u] P(v,u), \tag{23}$$

and the E-function is given by Eq. (3). Note that

$$\Phi(t,v,u,x) = J(t,v,u,x)/u. \tag{24}$$

With the aid of Eq. (10), the J-function is expressed in terms of the resolvent

$$J(t,v,u,x) = \lambda F(x,t;v,u) + \lambda \int_0^x \int_0^1 K(t,y;v,w;x) F(x,y;w,u)\,dy\,dw. \tag{25}$$

On differentiating Eq. (22) with respect to x, it becomes

$$J_x(t,v,u,x) = -J(t,v,u,x)/u + \int_0^1 \Phi(t,v,w,x) J(x,w,u,x)\,dw, \tag{26}$$

where

$$J(x,v,u,x) = \lambda F(x,x;v,u) + \lambda \int_0^x \int_0^1 \Phi(y,v,w,x) F(x,y;w,u)\,dy\,dw. \tag{27}$$

Equation (26) is the desired initial-value solution of the J-function, whose form in the conservative case is similar in form to that given by Kagiwada and Kalaba.[8]

In particular, for $t = 0$ we have

$$J_x(0,v,u,x) = -J(0,v,u,x)/u + \int_0^1 \Phi(0,v,w,x) J(x,w,u,x)\,dw. \tag{28}$$

Inserting $(x-t)$ in place of t in Eq. (22), it can be rewritten in the form

$$\begin{aligned} J(x-t,v,u,x) &= \lambda F(x,x-t;v,u) + \lambda \int_0^x \int_0^1 E(x-t,y;v,w) J(y,w,u,x)\,dy\,dw \\ &= \lambda \exp[-t/u] P(v,u) + \lambda \int_0^x \int_0^1 E(t,y;v,w) J(x-y,w,u,x)\,dy\,dw. \end{aligned} \tag{29}$$

Differentiation with respect to x provides

$$\begin{aligned} J_x(x-t,v,u,x) &= \lambda \int_0^1 E(t,x;v,w) J(0,w,u,x)\,dw \\ &\quad + \lambda \int_0^x \int_0^1 E(t,y;v,w) J_x(x-y,w,u,x)\,dy\,dw. \end{aligned} \tag{30}$$

On keeping Eq. (12) in mind, after some minor rearrangements of terms, the solution of the above equation becomes

$$J_x(x-t,v,u,x) = \int_0^1 \Phi(t,v,w,x) J(0,w,u,x)\,dw. \tag{31}$$

Putting $t = 0$ in Eq. (31), it becomes

$$J_x(x,v,u,x) = \int_0^1 \Phi(0,v,w,x) J(0,w,u,x)\,dw. \tag{32}$$

The generalized X- and Y-functions are defined by the relations

$$X(x,v,u) = J(x,v,u,x)/\lambda, \tag{33}$$

$$Y(x,v,u) = J(0,v,u,x)/\lambda. \tag{34}$$

On recalling Eqs. (24), (28), and (32), a Cauchy system for the X- and Y-functions takes the form

$$X_x(x,v,u) = \lambda \int_0^1 Y(x,v,w)Y(x,w,u)\frac{dw}{w}, \tag{35}$$

$$Y_x(x,v,u) = -Y(x,v,u)/u + \lambda \int_0^1 Y(x,v,w)X(x,w,u)\frac{dw}{w}, \tag{36}$$

together with the initial conditions

$$X(0,v,u) = P(v,u), \tag{37}$$

$$Y(0,v,u) = P(v,u). \tag{38}$$

On recalling Eqs. (24) and (26), we have

$$\Phi_x(t,v,u,x) = -\Phi(t,v,u,x)/u + \left(\frac{\lambda}{u}\right)\int_0^1 \Phi(t,v,w,x)X(x,w,u)dw. \tag{39}$$

Once X- and Y-functions have been determined by Eqs. (35) and (36), Eq. (39) permits us to compute the Φ-function. Then, with the aid of Eq. (19), we can compute the two-dimensional resolvent $K(t,y;v,u;x)$.

IV. EXPANSION IN LEGENDRE POLYNOMIALS OF THE PHASE FUNCTION

The results obtained in the preceding sections pertain to an arbitrary phase function. A simplification occurs if the phase function may be expanded in Legendre polynomials

$$P(v,w;\phi-\phi') = \sum_{m=0}^{\infty}(2-\delta_{0m}) \times \sum_{i=m}^{n} C_i^m P_i^m(v)P_i^m(w)\cos m(\phi-\phi'), \tag{40}$$

where δ_{0m} is the Kronecker delta function, P_i^m is the associated Legendre function of degree i and order m, and

$$C_i^m = C_i(i-m)!/(i+m)! \quad (i=m,m+1,\dots,n; m=0,1,2,\dots,n). \tag{41}$$

In this case the auxiliary function with the azimuthal arguments takes the form

$$J(t,v,u,x;\phi) = J^0(t,v,u,x) + 2\sum_{m=1}^{n} J^m(t,v,u,x)\cos m\phi, \tag{42}$$

where, for the sake of simplicity, ϕ_0 is put to be zero. The quantities $J^m(t,v,u,x)$ are the coefficients of the azimuthal expansion of the total auxiliary function. In a manner similar to Eq. (22), the Fourier component J^m-function fulfills the basic integral equation

$$J^m(t,v,u,x) = \lambda F^m(x,t;v,u) + \lambda \int_0^x \int_0^1 E^m(t,y;v,w)J^m(y,w,u,x)dy\,dw, \tag{43}$$

where $m = 0,1,2,\dots,n$, $0 \leqslant t \leqslant x$, $-1 \leqslant v \leqslant 1$, $0 \leqslant u \leqslant 1$,

$$F^m(x,t;v,u) = \exp[-(x-t)/u]P^m(v,u), \tag{44}$$

and

$$E^m(t,y;v,w) = \exp[-(|t-y|)/w]P^m(v,w)/w. \tag{45}$$

In Eqs. (44) and (45) $P^m(v,w)$ is denoted by

$$P^m(v,w) = \sum_{i=m}^{n} C_i^m P_i^m(v)P_i^m(w), \tag{46}$$

for $m = 0,1,2,\dots,n$

Once the auxiliary function has been determined by Eq. (42) for each m, the Fourier component of the intensity of radiation is found via the conversion of the auxiliary functions. Equation (42) determines the mth component of the total auxiliary function as a function of two arguments t and v, whereas u and x are parameters. However, the reduction of two arguments into a single argument results in computational simplification as below:

$$J^m(t,v,u,x) = \sum_{i=m}^{n} C_i^m P_i^m(v)J_i^m(t,u,x). \tag{47}$$

By substitution of Eqs. (45) and (46) into Eq. (43), we get

$$J_i^m(t,u,x) = \lambda F_i^m(x,t,u) + \lambda \sum_{j=m}^{n} \int_0^x \int_0^1 E_j^m(t,y,w)J_j^m(y,w,x)dw\,dy, \tag{48}$$

where

$$F_i^m(x,t,u) = P_i^m(u)\exp[-(x-t)/u], \tag{49}$$

and

$$E_j^m(t,y,w) = C_j^m P_i^m(w)P_j^m(w)\exp[-|t-y|/w]/w, \tag{50}$$

for $i = m,m+1,\dots,n$.

In such a way $J^m(t,v,u,x)$ is expressed in terms of the components $J_i^m(t,u,x)$ via Eq. (47), while for the evaluation of the J_i^m-function we should solve the system of Eq. (48). Then, for the complete solution of the problem under consideration, it is required to solve the system of Eq. (48) for all m. Hence, the complete solution of the problem for large n with the aid of the J_i^m-function becomes untractable from computational aspects.

Putting $t = x$ and $t = 0$ in Eq. (47), respectively, we get

$$J^m(x,v,u,x) = \lambda \sum_{i=m}^{n} C_i^m P_i^m(v)X_i^m(u,x), \tag{51}$$

$$J^m(0,v,u,x) = \lambda \sum_{i=m}^{n} C_i^m P_i^m(v)Y_i^m(u,x), \tag{52}$$

where

$$X_i^m(u,x) = J_i^m(x,u,x)/\lambda, \tag{53}$$

$$Y_i^m(u,x) = J_i^m(0,u,x)/\lambda. \tag{54}$$

From the physical aspects, the X_i^m- and Y_i^m-functions determine the probability of emergence of photons from the slab boundaries. On the other hand, recalling Eqs. (33) and (34), we have

$$X^m(x,v,u) = J^m(x,v,u,x)/\lambda, \tag{55}$$

$$Y^m(x,v,u) = J^m(0,v,u,x)/\lambda. \tag{56}$$

On keeping in mind Eqs. (51), (52), (55), and (56), we obtain

$$X^m(x,v,u) = \sum_{i=m}^{n} C_i^m P_i^m(v)X_i^m(u,x), \tag{57}$$

$$Y^m(x,v,u) = \sum_{i=m}^{n} C_i^m P_i^m(v) Y_i^m(u,x). \tag{58}$$

An allowance for Eqs. (35) and (36) enables us to get a Cauchy system for the X- and Y-functions,

$$\frac{dX_i^m(u,x)}{dx} = \lambda \sum_{j=m}^{n} (-1)^{i+j} C_j^m Y_j^m(u,x) \times \int_0^1 P_j^m(w) Y_i^m(w,x) \frac{dw}{w}, \tag{59}$$

$$\frac{dY_i^m(u,x)}{dx} = \frac{-Y_i^m(u,x)}{u} + \lambda \sum_{j=m}^{n} C_j^m X_j^m(u,x) \times \int_0^1 P_j^m(w) Y_i^m(w,x) \frac{dw}{w}, \tag{60}$$

for $i = m, m+1, \ldots, n$, $m = 0,1,2,\ldots,n$. The system of Eqs. (59) and (60) should be solved subject to the initial conditions

$$X_i^m(u,0) = P_i^m(-u), \tag{61}$$

$$Y_i^m(u,0) = P_i^m(-u), \tag{62}$$

where $0 \leqslant u \leqslant 1$.

V. DISCUSSION

In the present paper, with the aid of invariant imbedding, we derived a Cauchy system for the two-dimensional resolvent of the inhomogeneous truncated Milne's integral equation with anisotropic scattering. It is of interest to mention that, whereas the resolvent kernel of Milne's equation with isotropic scattering is shift-invariant with respect to the geometrical argument, the kernel of resolvent in the case of anisotropic scattering is of convolutional structure with respect to the geometrical argument and furthermore is of asymmetric character with respect to the angular argument. Then, the generalized Bellman–Krein–Sobolev formula is expressed in terms of the integration with respect to the angular argument of the product of Φ-functions. The dimensional reduction of the generalized X- and Y-functions is made when the phase function is expanded in Legendre polynomials. In our subsequent paper, making use of the present procedure, the scattering and transmission functions of Chandrasekhar's planetary problem with diffuse-and-specular reflectors will be dealt with.

ACKNOWLEDGMENT

This research was supported in part by the Ministry of Education, Japan, under Grant No. 403556.

[1] *Theoretical Astrophysics*, edited by V. A. Ambarzumian (Pergamon, London, 1958).

[2] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[3] Ida W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge U.P., London, 1960).

[4] V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, NJ, 1963).

[5] R. Bellman, H. Kagiwada, R. Kalaba, and S. Ueno, J. Math. Phys. **9**, 906 (1968).

[6] H. Kagiwada, R. Kalaba, and S. Ueno, *Multiple Scattering Processes: Inverse and Direct* (Addison–Wesley, Reading, MA, 1975).

[7] H. Kagiwada and R. Kalaba, J. Math. Phys. **11**, 1975 (1970).

[8] H. Kagiwada and R. Kalaba, J. Quant. Spectrosc. Radiat. Transfer **11**, 1101 (1971).

[9] R. Bellman and S. Ueno, Astrophys. Space Sci. **16**, 241 (1972).

[10] S. Ueno and S. Mukai, J. Quant. Spectrosc. Radiat. Transfer **13**, 126 (1973).

[11] R. Bellman, Proc. Am. Math. Soc. **8**, 435 (1957).

[12] M. G. Krein, Dokl. Akad. Nauk SSSR **100**, 413 (1955).

[13] S. L. Sobolev, Usp. Math. Nauk **9**, 234 (1954).

Bound and radiation fields of a pointlike SO(3) monopole

M. Sorg
Institut für Theoretische Physik, Universität Stuttgart, Pfaffenwaldring 57, D-7000 Stuttgart 80, West Germany

(Received 6 February 1984; accepted for publication 15 June 1984)

The macroscopic field of a pointlike, arbitrarily moving monopole is split up into its bound and radiation parts. The SO(2) fiber bundle, having the total monopole field as its curvature, is embedded into an SO(3) bundle (Georgi–Glashow model) such that the SO(2) monopole field is composed in an SO(3) invariant manner by the microscopic gauge and Higgs fields: The Higgs field constitutes the bound part and the gauge field (Yang–Mills field) is responsible for the radiation part of the macroscopic monopole field. The role played by the field equations for the validity of separate local energy-momentum conservation laws is discussed extensively.

I. INTRODUCTION AND SURVEY OF RESULTS

Since the pioneering work of t'Hooft[1] the properties of static smooth monopole solutions, emerging in nonabelian gauge theories, have been understood very well (see Ref. 2 for a review). It seems, however, that only little is known about nonstatic, smoothly localized solutions, where the monopole(s) are arbitrarily accelerated. The present paper is concerned with a single arbitrarily moving SO(3) monopole in the pointlike approximation.

From the macroscopic point of view, a pointlike monopole presents no difficulties. The macroscopic monopole field is the magnetic analog of the well-known Liénard–Wiechert field for a single electric point charge. The properties of this field were investigated thoroughly during the past decade from the point of view of traditional field theory.[3] At a first glance, it may seem that a point monopole in Maxwell's SO(2) electrodynamics is the same as a point monopole in SO(3) electrodynamics (Georgi–Glashow model), because the new effects emerging in nonabelian generalizations of ordinary electrodynamics turn up only when one considers smooth extended solutions to the field equations. This may be true as long as one deals with static solutions; but if one considers arbitrarily accelerated monopoles, the more general nonabelian description opens up new aspects also in the pointlike case. Here, the new approach refers to the separation of the macroscopic monopole field into bound and radiation parts.

The main aim of the present paper is to demonstrate, by means of embedding the (macroscopic) SO(2) bundle into a (microscopic) SO(3) bundle, that the bound and radiation parts of the macroscopic monopole field can be considered as being differently composed by the microscopic fields: The (microscopic) Higgs field constitutes the bound part of the macroscopic field, which appears to be very plausible, because the bound field contributes to the energy-momentum of the material source. On the other hand, the microscopic gauge field (Yang-Mills field) produces, via its longitudinal component, the macroscopic radiation field, which is responsible mainly for the interaction of the source with its surroundings. This result is obtained by an appropriate choice of the connection in the embedding bundle. The basic restriction of this choice of connection is the invariance requirement for the macroscopic field.

The paper is organized as follows.

In Sec. II we shortly review the geometry of the monopole field in terms of the fiber bundle language. It is demonstrated that this field can be considered as the Euler class in a two-dimensional real vector bundle over space-time. The total monopole field is split up into its bound and radiation parts and the geometric meaning of the latter is discussed. This discussion reveals the geometric origin of the partial disentanglement of bound and radiation fields with respect to the separate validity of Maxwell's field equations for each part (Bianchi identities).

The partial disentanglement of bound and radiation fields becomes complete if one considers the corresponding energy-momentum densities, which satisfy separate continuity equations for the bound and radiative case. We shortly discuss the question as to what extent the Bianchi identity and the dynamical part of the field equations are really necessary for the validity of the separate conservation laws.

After this extensive discussion of bound and radiation fields of the pointlike SO(2) monopole, we turn to the main aim of the paper by proposing that the bound and radiation parts of the macroscopic field be differently composed of microscopic fields: The bound part is due to the Higgs field, while the radiative part is produced by the longitudinal component of the Yang–Mills field. This identification program implies the use of a higher gauge group [SO(3)] containing the lower one [SO(2)] as a subgroup, such that the corresponding SO(2) bundle is the reduction of the embedding SO(3) bundle.

In Sec. III, the bundle embedding, which is the process inverse to the reduction SO(3)→SO(2), is studied in some detail. If SO(2) electrodynamics is considered as a reduced SO(3) system, the macroscopic monopole field **F** as the subbundle curvature can be expressed in a manifestly gauge-invariant SO(3) form. The bundle embedding must necessarily lead to very special SO(3) field configurations, because the exact symmetry group is only a submanifold of the proper gauge group. Such configurations are the vacua of Higgs and Yang–Mills type. The relation between these two vacuum field configurations is studied in some detail along with the invariance of the Euler class with respect to a transition from one configuration to the other.

In Sec. IV, an intermediate configuration between the

 0022-2488/85/010089-10$02.50

two types of vacua can be found which allows the desired identification of the bound and radiation fields with certain combinations of the microscopic fields. Starting from the Higgs vacuum configuration, the necessary change of connection is performed in such a way that the curvature in the embedding SO(3) bundle becomes connected to the acceleration of the point monopole. This means that the (microscopic) Yang–Mills fields, being responsible for the (macroscopic) radiation field, are due to the acceleration of the particle. This fits very well into the usual understanding of the relation between radiation and acceleration in the classical domain.

In Sec. V, the continuity equations for the three different kinds of energy-momentum densities arising from the three different kinds of fields (Yang–Mills field and its longitudinal and transverse component) are studied. It turns out that all three kinds of densities satisfy their corresponding continuity equations despite the fact that the field equations are not valid. Whereas for the Yang–Mills field and its longitudinal component (= radiation field) one half of the field equations is satisfied via a Bianchi identity, the transverse part does not even satisfy a Bianchi identity. Nevertheless, a conservation law is valid for the transverse energy-momentum density.

An integration (Appendix) of the three kinds of densities yields the corresponding four-momenta. The momenta due to the transverse and longitudinal part of the Yang–Mills field turn out to be equal.

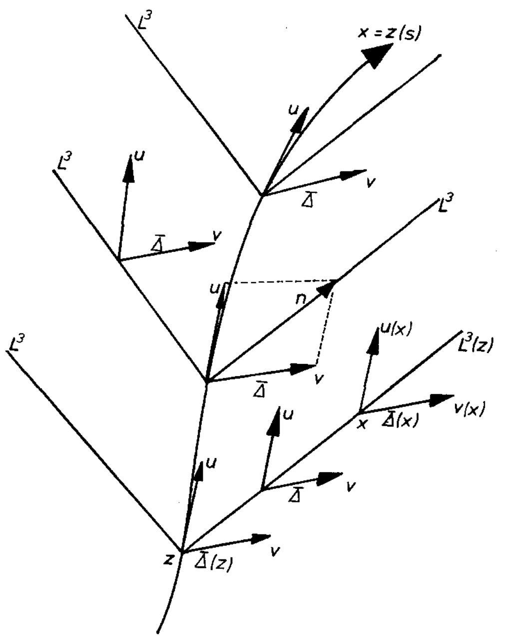


FIG. 1. The bundle hierarchy $\tilde{\tau}_4 \subset \bar{\tau}_4 \subset \tau_4$: The unit vector $\mathbf{u}(\mathbf{z}) := [d\mathbf{z}(s)/ds]$ $(\mathbf{u}\cdot\mathbf{u} = +1)$ is tangent to the monopole world line $\mathfrak{L}$ at some point $\mathbf{z}$. It is orthogonal to the three-dimensional plane $\bar{\Delta}(\mathbf{z})$ which itself cuts the world line $\mathfrak{L}$ orthogonally at that event $\mathbf{z}$ and is spanned by the orthonormal vectors $\{\mathbf{v};\mathbf{k},\mathbf{h}\}(\mathbf{z})$. Hence, $\mathbf{e}(\mathbf{z}) = \{\mathbf{u};\mathbf{v},\mathbf{k},\mathbf{h}\}$ is an orthonormal tetrad at $\mathbf{z}$. A section $\mathbf{e}(x)$ in the principal SO(1,3) bundle Λ_4 associated to the (trivial) tangent bundle τ_4 of Minkowski space M_4 is obtained from $\mathbf{e}(\mathbf{z})$ by parallel transport of $\mathbf{e}(\mathbf{z})$ (with respect to the canonical connection in τ_4) along all future null geodesics emanating from the vertex point $\mathbf{z}$ on $\mathfrak{L}$. Thus the tetrad field $\mathbf{e}(x)$ is constant over any future light cone $L^3(\mathbf{z})$ and varies only when the vertex point $\mathbf{z}$ is shifted along the world line. The subframe $\bar{\mathbf{e}}(x) := \{\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$ locally spans the distribution $\bar{\Delta}(x)$ and simultaneously represents a section in the principle bundle $\bar{\Lambda}_4$ associated to the three-plane bundle $\bar{\tau}_4$. The section $\mathbf{e}(x)$ can be used to calculate the connection $\bar{\omega}$ in $\bar{\tau}_4$ (3.4a) and (3.6) as well as the extrinsic curvature $\{\mathbb{B}_i\}$ (3.4b) and (3.12). The above construction evidently implies $\nabla\mathbf{u} = \mathbf{d}s \otimes \partial\mathbf{u}/\partial s \equiv \mathbf{n} \otimes \dot{\mathbf{u}}$ and therefore the curvature in $\bar{\tau}_4$ vanishes (3.13) on account of (3.12). However, the further reduction $\bar{\Delta} \to \tilde{\Delta}$ leads to the nontrivial bundle $\tilde{\tau}_4$ with the connection being given by $\mathbb{A}_1$ (2.12) and its curvature $\mathbb{F}$ representing the electromagnetic field (2.9). Thus, the electromagnetic SO(2) bundle $\tilde{\tau}_4$ can be embedded hierarchically into two trivial bundles $\bar{\tau}_4$ and τ_4.

II. BOUND AND RADIATION FIELDS IN SO(2) ELECTRODYNAMICS

As a preparation to the splitting of the field into bound and radiation parts in Yang–Mills–Higgs theory, we briefly give a survey of this phenomenon within the framework of abelian electrodynamics including the recent fiber bundle approach to this problem.[4]

A. Field of an arbitrarily moving monopole

The electromagnetic field $\mathbb{F}$,

$$\mathbb{F} = \tfrac{1}{2} F_{\mu\nu} \mathbf{d}x^\mu \wedge \mathbf{d}x^\nu, \tag{2.1}$$

of a pointlike magnetic monopole is a solution to Maxwell's homogenous equations $(\delta := -*\mathbf{d}*)$

$$\delta *\mathbb{F} = 0, \tag{2.2a}$$

$$\delta\, \mathbb{F} = 0, \tag{2.2b}$$

which are valid off the world line of the point charge. This field $\mathbb{F}$ is the magnetic analog of the well-known Liénard–Wiechert field, which is generated by an electric monopole in arbitrary motion. The magnetic monopole field $\mathbb{F}$ is given explicitly by[4]

$$\mathbb{F} = \rho^{-2}\, \tilde{\mathbb{L}} + \rho^{-1} \mathbf{n} \wedge \tilde{\mathbb{L}}(\dot{\mathbf{u}}). \tag{2.3}$$

Here, $\rho(x)$ is the retarded distance of $x \in M_4^-$ from the particle world line $\mathfrak{L}$: $\mathbf{x} = \mathbf{z}(s)$; $\mathbf{n}$ is a lightlike one-form which can be written as (Fig. 1)

$$\mathbf{n} = \mathbf{u} + \mathbf{v} \tag{2.4}$$

where $\mathbf{u}$ is the (timelike) velocity one-form and $\mathbf{v}$ is the dual of a (spacelike) "radial" unit vector $\mathbf{v}$ $[\mathbf{v}(\mathbf{v}) = -1]$ being Fermi-Walker-transported along the particle world line. The two-form $\tilde{\mathbb{L}}$

$$\tilde{\mathbb{L}} = -\mathbf{k} \wedge \mathbf{h} \tag{2.5}$$

is a representation of the generator for (local) SO(2) rotations

$$\mathbf{k}' = \cos\alpha \mathbf{k} + \sin\alpha \mathbf{h},$$

$$\mathbf{h}' = -\sin\alpha \mathbf{k} + \cos\alpha \mathbf{h}, \tag{2.6}$$

within the distribution $\tilde{\Delta}$. This distribution is spanned (locally) by the unit sections $\mathbf{k}(x)$ and $\mathbf{h}(x)$ in the two-dimensional vector bundle $\tilde{\tau}_4$ which was used in Ref. 4 for the geometrization of the Liénard–Wiechert field. Therefore, the value of $\tilde{\mathbb{L}}$ upon the four-acceleration $\dot{\mathbf{u}} (\equiv d\mathbf{u}/ds)$ of the monopole implies a projection onto $\tilde{\Delta}$ and an SO(2) rotation within $\tilde{\Delta}$:

$$\tilde{\mathbb{L}}(\dot{\mathbf{u}}) = (\mathbf{k}\dot{\mathbf{u}})\mathbf{h} - (\mathbf{h}\dot{\mathbf{u}})\mathbf{k}. \tag{2.7}$$

The electromagnetic field (2.3) is the Euler class of the vector bundle $\tilde{\tau}_4$ and hence is a member of the cohomology group $H^2(M_4^-, \mathbb{Z})$ containing all closed two-forms over M_4^-:

$$\mathbf{dF} = 0. \tag{2.8}$$

which is the same as the first half (2.2a) of the field equations. Equation (2.8) may be verified by observing that F is also the curvature of the connection ($\mathbb{A}_1$) in $\tilde{\tau}_4$:

$$\mathbf{F} = \mathbf{d}\,\mathbb{A}_1. \tag{2.9}$$

Thus (2.8) turns out as the Bianchi identity for F in $\tilde{\tau}_4$. However, F does not admit a *global* "potential" $\mathbb{A}_1$ over M_4^-, otherwise the magnetic charge g would be zero due to Stokes theorem applied to an arbitrary two-cycle $C^2 \subset M_4^-$ $(\partial C^2 = \varnothing)$:

$$g := \frac{1}{4\pi}\oint_{C^2} \mathbf{F} \rightarrow \frac{1}{4\pi}\oint_{\partial C^2} \mathbb{A}_1 = 0. \tag{2.10}$$

On the contrary, we shall verify below that $g = -1$. The relation (2.9) is valid only *locally* on patches chosen appropriately over M_4^-.

The potential $\mathbb{A}_1$ transforms just as a connection one-form in the intersections of the patches when an SO(2) gauge transformation (2.6) is applied:

$$\mathbb{A}_1 \rightarrow \mathbb{A}_1' = \mathbb{A}_1 + \mathbf{d}\alpha. \tag{2.11}$$

This transformation law becomes immediately evident by observing that the connection $\mathbb{A}_1$ can be expressed by the two $\tilde{\tau}_4$ sections $\mathbf{k}(x)$ and $\mathbf{h}(x)$ as

$$\mathbb{A}_1 = (\mathbf{k}\cdot\nabla\mathbf{h}), \tag{2.12}$$

where the SO(2) gauge transformation (2.6) applies.

Whereas the first half (2.2a) of the field equations was shown to be a Bianchi identity, the second half (2.2b) results from the fact that the (Poincaré) dual of the electromagnetic field admits a *global* potential ${}^{\mathrm{LW}}\mathbb{A}$

$$^{*}\mathbf{F} = \mathbf{d}\,{}^{\mathrm{LW}}\mathbb{A}, \tag{2.13}$$

which is the well-known Liénard–Wiechert potential

$${}^{\mathrm{LW}}\mathbb{A} = -\rho^{-1}\mathbf{u}. \tag{2.14}$$

We shall reconsider the meaning of the field equations subsequently in connection with the continuity equations for the energy-momentum densities, but first let us make some remarks on the splitting of the field F into bound and radiation parts.

B. Splitting of the field

A natural splitting of the total monopole field (2.3) is based upon the separation into long range $(\sim\rho^{-1})$ and short range $(\sim\rho^{-2})$ terms. Consequently, one writes the total field F as a sum of bound (^{b}F) and radiation (^{r}F) parts:

$$\mathbf{F} = {}^b\mathbf{F} + {}^r\mathbf{F}, \tag{2.15a}$$

$${}^b\mathbf{F} = \rho^{-2}\tilde{\mathbb{L}} \tag{2.15b}$$

$${}^r\mathbf{F} = \rho^{-1}\mathbf{n}\wedge\tilde{\mathbb{L}}(\dot{\mathbf{u}}). \tag{2.15c}$$

In the context of this splitting, there is an interesting point: The two fields ^{b}F and ^{r}F are partly decoupled in the sense that one-half of Maxwell's equations (2.2b) is satisfied by each one of the partial fields separately; i.e.,

$$\delta\,{}^r\mathbf{F} = 0, \tag{2.16a}$$

$$\delta\,{}^b\mathbf{F} = 0. \tag{2.16b}$$

Here, only (2.16a) is the relevant equation, (2.16b) is a consequence of (2.16a) and (2.2b). The geometric approach[5] to the monopole field reveals the origin of this phenomenon. The dual (^{*r}F) of the radiation part (^{r}F) is the curvature in the bundle $\hat{\tau}_4$ normal to the original bundle $\tilde{\tau}_4$ which has the total field F as its curvature:

$${}^{*r}\mathbf{F} = -\mathbf{d}\mathbb{B}_1. \tag{2.17}$$

The connection in $\hat{\tau}_4$ is given by the one-form $\mathbb{B}_1$ and so the Bianchi identity in $\hat{\tau}_4$ agrees with (2.16a).

In addition, ^{*r}F is the exterior derivative of the mean curvature one-form for the integral surfaces of the distribution $\tilde{\Delta}$.

We now come to the question of to what extent the field equations are necessary for the validity of the continuity equations for the energy-momentum densities.

C. Splitting of the energy-momentum density

The partial disentanglement of bound and radiation fields expressed by (2.16) becomes complete if one considers the energy-momentum densities produced by those fields. Since the total energy-momentum density $\mathbb{T}$ is quadratic in the monopole field strength F,

$$\mathbb{T} = T^{\mu}{}_{\nu}\mathbf{E}_{\mu}\otimes\mathbf{d}x^{\nu}, \tag{2.18a}$$

$$-8\pi T^{\mu}{}_{\nu} = F^{\mu\rho}F_{\nu\rho} + {}^{*}F^{\mu\rho}\,{}^{*}F_{\nu\rho}, \tag{2.18b}$$

the splitting (2.15a) of the total field F induces a splitting of $\mathbb{T}$ into a radiative part (${}^r\,\mathbb{T}$) and the bound part (${}^b\mathbb{T}$):

$$\mathbb{T} = {}^r\,\mathbb{T} + {}^b\mathbb{T}, \tag{2.19a}$$

$$-8\pi\,{}^r T^{\mu}{}_{\nu} = {}^rF^{\mu\rho}\,{}^rF_{\nu\rho} + {}^{*r}F^{\mu\rho}\,{}^{*r}F_{\nu\rho}. \tag{2.19b}$$

Observe that ${}^r\,\mathbb{T}$ contains only the long-range terms $(\sim\rho^{-2})$, whereas ${}^b\mathbb{T}$ is built from the properly bound part of the field $(\sim\rho^{-4})$ and the mixed terms $(\sim\rho^{-3})$.

The energy-momentum conservation law is usually expressed locally by the vanishing of the divergence $\nabla\cdot\mathbb{T}$ which may be written as a one-form equation

$$\nabla\cdot\mathbb{T} \equiv (\partial_{\mu}T^{\mu}{}_{\nu})\mathbf{d}x^{\nu} = 0. \tag{2.20}$$

The remarkable point in the present context is now that the validity of the field equations (2.2) is not really necessary for the vanishing of the tensor divergence (2.20). In fact, one has the identity

$$-4\pi\nabla\cdot\mathbb{T} \equiv {}^{*}(\delta\mathbf{F}\wedge{}^{*}\mathbf{F}) - {}^{*}(\delta{}^{*}\mathbf{F}\wedge\mathbf{F}), \tag{2.21}$$

and thus the tensor divergence vanishes not only when both field equations [(2.2a) and (2.2b)] are satisfied; this divergence vanishes also when the wedge products, occurring in the two terms on the right of (2.21), vanish separately. However, the most general case is when even the wedge products are not zero themselves but only their difference vanishes. So we see that one can actually dispense with the field equations and nevertheless one can have energy momentum conservation!

A physically relevant application of this effect is encountered when (2.21) is written down for the radiation part ^{r}F (2.15c) of the monopole field (2.3). In this case, the first term on the right of (2.21) vanishes on account of the Bianchi identity (2.16a) in the normal bundle $\hat{\tau}_4$. An explicit computation of the second term on the right of (2.21) reveals that the wedge product occurring there vanishes also.[3] Hence, the radiation part of the energy-momentum density satisfies the continuity equation separately,

$$\nabla\cdot\mathbb{T} = 0, \tag{2.22}$$

without the field equations (2.2) being fully satisfied by ${}^r\mathbb{F}$.

This is a very plausible result if one interprets the classical mechanism of the generation of the radiation field in quantum mechanical terms: Once the photons have been generated by the source, they move outward with the velocity of light and independent of the subsequent behavior of the source and its bound field. An interpretation in purely classical terms is also possible but involves a more subtle argument concerning the geometry of light cones.[6]

D. The bound and radiation parts as composite fields

The separate continuity equation (2.22) for the radiative energy-momentum density ${}^r\mathbb{T}$ is extremely plausible and therefore was welcome for the understanding and description of the electromagnetic field of a single monople. However, it seems somewhat unsatisfactory that the electromagnetic field as an entity is broken apart into two parts with rather distinct physical meanings: The bound part contributes to the energy momentum of the source and thus increases its inertia and weight, whereas the radiative part is shot out into space in order to interact with the remaining matter in the neighborhood of the source. It seems desirable to introduce two conceptually different fields in order to describe those two different aspects of the Maxwell field.

Following this line of thinking, one wants to have two (microscopic) fields which together constitute the (macroscopic) Maxwell field $\mathbb{F}$ of the monopole in such a way that one of the microscopic fields produces the macroscopic radiation part ${}^r\mathbb{F}$ whereas the other is responsible for the bound part ${}^b\mathbb{F}$. A possibility for such a construction arises when one embeds the fiber bundle $\tilde{\tau}_4$ for the macroscopic Maxwell field into a higher-dimensional fiber bundle $\bar{\tau}_4$. Thus, one needs a constraint field $\mathbf{v}(x)$ in $\bar{\tau}_4$, which defines the reduction $\bar{\tau}_4\to\tilde{\tau}_4$. The curvature fields $\{\mathbb{F}_i\}$ in $\bar{\tau}_4$ should then cooperate with the constraint field $\mathbf{v}$ in order to produce the macroscopic field $\mathbb{F}$ in a way that the higher-order gauge field $(\mathbb{F}_i v^i)$ generates the radiation field ${}^r\mathbb{F}$ whereas $\mathbf{v}(x)$ acquires the meaning of a microscopic matter field and produces the bound macroscopic field. In this way, the contribution of the bound electromagnetic field to the mass of the source would actually be due to the microscopic matter field.

As a specification of this program, let us embed the SO(2) bundle $\tilde{\tau}_4$ into an SO(3) bundle $\bar{\tau}_4$. For the geometrized monopole field $\mathbb{F}$ (2.3), this embedding bundle $\bar{\tau}_4$ may be constructed by means of the three-distribution $\bar{\Delta}$, which itself is spanned (locally) by the space-like frame vectors $\{\mathbf{e}_i(x), i=1,2,3\}\equiv\{\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$, being orthogonal to the four-velocity field $\mathbf{u}(x)$. Instead of reducing the tangent bundle τ_4 of modified Minkowski space $M_4^-\,(=M_4\setminus\mathfrak{L})$ to the SO(2) bundle $\tilde{\tau}_4$, we reduce it first to the SO(3) bundle $\bar{\tau}_4$ which contains $\tilde{\tau}_4$ as a subbundle. This subbundle $\tilde{\tau}_4$ is specified by the "Higgs vector" $\mathbf{v}(x)=\{v^i\}(x)$. Considering the latter one as a τ_4-section, it agrees with the first spacelike frame vector $\mathbf{v}(x)$ being orthogonal to $\tilde{\Delta}$ within $\bar{\Delta}$ (Fig. 1).

For the associated principal bundles, this reduction process means that we first reduce the Lorentz group SO(1,3) to its subgroup SO(3) and then the electromagnetic SO(2) bundle is viewed as a further reduction of the SO(3) system.

Consequently, the curvature $\mathbb{F}$ in the SO(2) subbundle can be expressed by the objects in the embedding SO(3) bundle and is given explicitly by[7]

$$\mathbb{F} = v^i\mathbb{F}_i - \tfrac{1}{2}\epsilon_{ijk}v^k(\mathbb{D}v^i)\wedge(\mathbb{D}v^j) \tag{2.23a}$$

$$(\mathbb{D}v^i \equiv \mathbf{d}v^i + \epsilon^{ij}{}_k v^k\mathbb{A}_j). \tag{2.23b}$$

Here, $\mathbb{D}$ denotes the covariant derivative in $\bar{\tau}_4$.

According to our philosophy concerning the field splitting (2.15a) we now try the identifications

$${}^r\mathbb{F} = v^i\mathbb{F}_i, \tag{2.24a}$$

$${}^b\mathbb{F} = -\tfrac{1}{2}\epsilon_{ijk}v^k(\mathbb{D}v^i)\wedge(\mathbb{D}v^j). \tag{2.24b}$$

The longitudinal part of the "Yang–Mills field" $\mathbb{F}_i$ as the *microscopic* SO(3) gauge field produces the *macroscopic* SO(2) gauge field. Similarly, the Higgs field $\mathbf{v}(x)$ as the microscopic matter field produces the macroscopic bound part ${}^b\mathbb{F}$ contributing to the mass of the source.

In the following, we shall determine the connection $\{\mathbb{A}_i(x)\}$ and its curvature fields $\{\mathbb{F}_i(x)\}$ in the embedding bundle $\bar{\tau}_4$ such that the desired identifications are actually verified.

III. SO(2) ELECTRODYNAMICS AS A REDUCED SO(3) SYSTEM

If one wants to consider SO(2) electrodynamics as a reduced SO(3) system, one has to expect specific restrictions for the SO(3) objects, such as connection $\{\mathbb{A}_i\}$ and curvature $\{\mathbb{F}_i\}$. The reason for this expectation is that some of the degrees of freedom of the higher-dimensional gauge group must be frozen in order that the higher-dimensional representation [here SO(3)] becomes strictly equivalent to the lower-dimensional case [SO(2)]. One of those specific configurations is known as "Higgs vacuum"; this field configuration is also the starting point for the construction (2.24) to be considered below. But the bundle embedding $\tilde{\tau}_4\to\bar{\tau}_4$ first leads to another special type of SO(3) field configuration, which may be called "Yang–Mills vacuum." So we first describe the latter one and then perform the transition to the Higgs vacuum configuration. However, before relating these two different vacua to each other, some remarks must be made about the general properties of an SO(3) bundle $\bar{\tau}_4$ resulting from its embedding into the trivial tangent bundle τ_4 over space-time M_4^-.

A. Bundle embedding

The geometry of the tangent subbundle $\bar{\tau}_4$ exhibits some constraints resulting from the fact that its embedding bundle τ_4 as the tangent bundle of M_4^- is trivial. The triviality of τ_4 is expressed by the vanishing of the curvature Ω of the canonical connection ω over flat Minkowski space M_4^-:

$$\Omega := \mathbf{d}\omega + \omega\wedge\omega = 0. \tag{3.1}$$

Here, Ω and ω are $so(1,3)$-valued two-forms and one-forms, respectively. According to the property (3.1), any representative of the canonical connection ω is of the form of a pure gauge

$$\omega = \Lambda^{-1}\cdot \mathbf{d}\Lambda, \tag{3.2}$$

where Λ is an element of the (proper) Lorentz group SO(1,3). The decomposition of the $so(1,3)$ element ω (3.2) with respect to a basis $\{L^i, l^i\}$ of generators of the Lorentz group is given by[4]

$$\omega = \mathbb{A}_i L^i + \mathbb{B}_i l^i, \tag{3.3}$$

where $\{L^i\}$ are the generators of the rotation subgroup SO(3) and $\{l^i\}$ denote the Lorentz boost operators. The connection coefficients $\mathbb{A}_i$, $\mathbb{B}_i$ of the canonical connection ω are expressed by the tetrad vectors $\{\mathbf{e}_\mu(x)\} = \{\mathbf{u};\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$ as[8]

$$\mathbb{A}_i = \tfrac{1}{2}\epsilon_i{}^{jk}(\mathbf{e}_j\cdot\nabla\mathbf{e}_k), \tag{3.4a}$$

$$\mathbb{B}_i = -(\mathbf{e}_i\cdot\nabla\mathbf{e}_0), \tag{3.4b}$$

and satisfy the relations

$$\mathbf{d}\mathbb{A}_i + \tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{A}_j\wedge\mathbb{A}_k = \tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{B}_j\wedge\mathbb{B}_k, \tag{3.5a}$$

$$\mathbf{d}\mathbb{B}_i + \epsilon_i{}^{jk}\mathbb{A}_j\wedge\mathbb{B}_k = 0. \tag{3.5b}$$

These relations can now be exploited for the geometry in the SO(3) subbundle $\bar\tau_4$. The connection $\bar\omega$ in $\bar\tau_4$ is obtained by projecting the canonical connection one-form ω (3.3) onto the $so(3)$ subalgebra of the Lorentz Lie algebra $so(1,3)$:

$$\bar\omega = \mathbb{A}_i L^i. \tag{3.6}$$

The $so(3)$ curvature fields $\{\mathbb{F}_i\}$ of this connection are given by

$$\mathbb{F}_i = \mathbf{d}\mathbb{A}_i + \tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{A}_j\wedge\mathbb{A}_k, \tag{3.7}$$

which may be also put in the form

$$\mathbb{F}_i = \tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{B}_j\wedge\mathbb{B}_k, \tag{3.8}$$

when (3.5a) is used.

The connection coefficients $\mathbb{B}_i$ acquire the meaning of extrinsic curvature fields for the distribution $\bar\Delta$ spanned locally by the subframe vectors $\{\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$ and (3.5b) is written more conveniently in SO(3) covariant form as

$$\mathbb{D}\mathbb{B}_i := \mathbf{d}\mathbb{B}_i + \epsilon_i{}^{jk}\mathbb{A}_j\wedge\mathbb{B}_k \equiv 0. \tag{3.9}$$

This means that the extrinsic curvature fields $\mathbb{B}_i$ are covariantly constant. This covariant constancy can be used for a very short and elegant verification of the Bianchi identity in $\bar\tau_4$:

$$\mathbb{D}\mathbb{F}_i := \mathbf{d}\mathbb{F}_i + \epsilon_i{}^{jk}\mathbb{A}_j\wedge\mathbb{F}_k \equiv 0 \tag{3.10}$$

if one uses (3.8) for the curvature fields $\mathbb{F}_i$. The Bianchi identity (3.10) in $\bar\tau_4$ ensures also the validity of the Bianchi identity (2.8) for the macroscopic field (2.23), which is readily verified by observing

$$\mathbb{D}\mathbb{D}v^i = \epsilon^{ij}{}_k v^k\mathbb{F}_j. \tag{3.11}$$

Let us make some final remarks concerning the embedding $\bar\tau_4\to\tau_4$. Since the connection $\bar\omega$ (3.6) in $\bar\tau_4$ was obtained by projecting the connection (ω) of a trivial embedding bundle (τ_4), there are some special features which are not present in the general case. The special form (3.8) of the curvature fields $\mathbb{F}_i$ and the covariant constancy of the extrinsic curvature fields $\mathbb{B}_i$ are only valid on account of the special connection $\bar\omega$ (3.6). On the other hand, the relation between fields $\mathbb{F}_i$ and "potentials" $\mathbb{A}_i$ (3.7), the formula (2.23) for the subbundle curvature $\mathbf{F}$, the Bianchi identities (2.8) and (3.10), and the double operation of the covariant derivative $\mathbb{D}$ in (3.11) are valid for *any* connection in $\bar\tau_4$. We have to remember this fact later on when we change the connection $\bar\omega$ in $\bar\tau_4$ in order to obtain the identification (2.24), which cannot be obtained by a projectively generated connection such as $\bar\omega$ (3.6). We are demonstrating this now by applying the present embedding mechanism to the bundle $\bar\tau_4$ with its curvature $\mathbf{F}$ being the monopole field (2.3).

B. Yang–Mills vacuum and charge quantization

In Ref. 4 a special section $\mathbf{e}(x) = \{\mathbf{u};\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$ in the trivial principal bundle Λ_4 associated to the tangent bundle τ_4 over Minkowski space M_4^- has been used to perform the splitting of τ_4 into the Whitney sum of $\bar\tau_4$ and $\hat\tau_4$. Utilizing now the subframe $\mathbf{e}(x) := \{\mathbf{v},\mathbf{k},\mathbf{h}\}(x)$ as a section in $\bar\Lambda_4$ associated to $\bar\tau_4$ we find that the curvature fields ($\mathring{\mathbb{F}}_i$) of the projectively generated connection $\bar\omega$ (3.6) in $\bar\tau_4$ vanish. This can be immediately seen by means of (3.8) and the specific result found for the extrinsic curvature fields $\mathbb{B}_i$

$$\mathbb{B}_i = -(\mathbf{e}_i\cdot\dot{\mathbf{u}})\mathbf{n}, \tag{3.12}$$

which are all proportional to the single one-form $\mathbf{n}$. Thus $\bar\tau_4$ is a trivial bundle

$$\mathring{\mathbb{F}}_i \equiv 0, \tag{3.13}$$

and consequently the Euler class (2.23) of the subbundle $\bar\tau_4$ acquires the specific shape

$$\mathbf{F}\to\mathring{\mathbf{F}} = -\tfrac{1}{2}\epsilon_i{}^{jk}(\mathbb{D}v^i)\wedge(\mathbb{D}v_j)v_k. \tag{3.14}$$

This means that the macroscopic field $\mathbf{F}$ is built up solely by the Higgs field $\mathbf{v}$, wheras the Yang–Mills field contribution is zero. Such a field configuration we call a "Yang–Mills vacuum," which is just the opposite of a Higgs vacuum where the macroscopic field $\mathbf{F}$ is built up by the Yang–Mills field alone (see below).

In the case of vanishing curvature (3.13), one can always find an SO(3) gauge transformation $S(x) = \{S^i{}_j(x)\}\epsilon$SO(3)

$$\mathbf{e}_i' = S^j{}_i\mathbf{e}_j, \tag{3.15}$$

such that the new potentials $\{\mathring{\mathbb{A}}_i\}$

$$\mathring{\mathbb{A}}_i = S^j{}_i\mathbb{A}_j + \Sigma_i \quad (\Sigma_i\bar L^i = S^{-1}\cdot\mathbf{d}S) \tag{3.16}$$

vanish[9] everywhere in M_4^-:

$$\mathring{\mathbb{A}}_i \equiv 0. \tag{3.17}$$

Therefore, the SO(3) covariant derivative $\mathbb{D}$ can be replaced by the ordinary one ($\mathbf{d}$), and the macroscopic field (3.14) assumes an especially simple form

$$\mathring{\mathbf{F}} = -\tfrac{1}{2}\epsilon_{ijk}v^k(\mathbf{d}v^i)\wedge(\mathbf{d}v^j) \equiv -\mathbf{d}S^2. \tag{3.18}$$

It just becomes (the pullback of) the surface element $\mathbf{d}S^2$ on a unit sphere S^2 contained in the typical fiber $E(-3)$ of $\bar\tau_4$ [observe $v^iv_i = -1$, the fiber metric is here $\bar g_{ij} = \mathrm{diag}(-1,-1,-1)$]. In view of this fact, it is especially easy to find the monopole charge g (2.10) as

$$g = \frac{1}{4\pi}\oint\mathbf{F} = -1, \tag{3.19}$$

because the Higgs vector $\mathbf{v}$ covers the unit sphere S^2 just once.[10] This is seen immediately by remembering that the potentials $\mathbb{A}_i$, in the gauge where $v^i = \delta^i{}_1$, have been found originally to be given in S^2-coordinates (ϑ,φ) as

$$\mathbb{A}_1 = \cos\vartheta\, \mathbf{d}\varphi, \quad \mathbb{A}_2 = -\sin\vartheta\, \mathbf{d}\varphi \quad \mathbb{A}_3 = \mathbf{d}\vartheta. \tag{3.20}$$

Consequently the Euler class $\mathring{\mathbb{F}}$ (2.9) is

$$\mathring{\mathbb{F}} = -\sin\vartheta\, \mathbf{d}\vartheta \wedge \mathbf{d}\varphi \tag{3.21}$$

and the charge integral (3.19) measures the surface area of the unit sphere S^2.

Observe, however, that the magnetic charge g is always quantized ($g\in\mathbb{Z}$) if the monopole field $\mathbb{F}$ is an Euler class being always of the form (2.23). This result holds also if (3.13) is not valid, which is seen most readily if the field (2.23) is put into the form

$$\mathbb{F} = -\mathbf{d}S^2 + \mathbf{d}\mathbb{A}\{\mathbf{v}\} \quad (\mathbb{A}\{\mathbf{v}\} := v^i\mathbb{A}_i). \tag{3.22}$$

Only the first term on the right of (3.22) contributes to the charge and ensures that g is an integer (= winding number of the map $\mathbf{v}: M_4^- \to S^2$). The second term is evidently a total differential and cannot contribute to g as a consequence of Stokes theorem (2.10). In this way, the property of the macroscopic field $\mathbb{F}$ of being an Euler class and therefore also being an element of the cohomology group $H^2(M_4^-,\mathbb{Z})$ has been demonstrated in general.

C. Transition to a Higgs vacuum configuration

The Higgs vacuum configuration for the SO(3) field variables $\{\mathbb{A}_i, \phi_i\}$

$$\phi_i := \phi v_i, \quad (\phi^i\phi_i) = -\phi^2 < 0, \tag{3.23}$$

is a special solution to the Yang–Mills–Higgs field equations[2]

$$\mathbb{D}^*\mathbb{F}_i = \epsilon_{ijk}\phi^k\, {}^*\mathbb{D}\phi^j, \tag{3.24a}$$

$$\mathbb{D}^*\mathbb{D}\phi^i = \frac{\partial\, {}^*\mathbb{V}}{\partial\phi_i}, \tag{3.24b}$$

following from the variational principle

$$\delta\int {}^*\mathscr{L} = 0, \tag{3.24c}$$

where the Lagrangian four-form density ${}^*\mathscr{L}$ is given by

$${}^*\mathscr{L} = \tfrac{1}{2}\mathbb{F}^i\wedge\mathbb{F}_i + \tfrac{1}{2}(\mathbb{D}\phi^i)\wedge{}^*(\mathbb{D}\phi_i) - {}^*\mathbb{V}. \tag{3.25}$$

The second half of the field equations for the Yang–Mills fields $\mathbb{F}_i$ is given by the Bianchi identity (3.10) quite analogously to the abelian case (2.2a). The particular solution to the field equations (3.24) called "Higgs vacuum" (denoted hereafter by a bar) may be characterized now by the requirement that the Higgs vector field $\boldsymbol{\phi}(x)$ be a covariantly constant section in $\bar{\tau}_4$:

$$\bar{\mathbb{D}}\phi^i = 0, \tag{3.26a}$$

$$\mathbb{V}(\phi_i) = 0. \tag{3.26b}$$

If the Higgs vacuum extends up to the location of the monopole, one deals with a point monopole and one can dispense with the potential $\mathbb{V}$. The Higgs vacuum conditions (3.26) now assume the shape

$$\phi^i = \bar{\phi}v^i, \tag{3.27a}$$

$$\bar{\phi} = \text{const}, \quad \mathbb{V}(\bar{\phi}) = 0, \tag{3.27b}$$

$$\bar{\mathbb{D}}v^i = 0. \tag{3.27c}$$

For the sake of simplicity, we assume $\bar{\phi} = 1$ in the following.

The requirement (3.27c) restricts the potentials $\mathbb{A}_i$ to the form

$$\bar{\mathbb{A}}_i = -\epsilon_{ijk}v^j\,\mathbf{d}v^k - \mathbb{A}\{\mathbf{v}\}v_i, \tag{3.28}$$

where $\mathbb{A}\{\mathbf{v}\}$ has already been defined in (3.22). As a consequence of this restricted form of the connection coefficients $\bar{\mathbb{A}}_i$, the curvature fields assume the very specific shape

$$\bar{\mathbb{F}}_i = -\mathbb{F}v_i, \tag{3.29}$$

where $\mathbb{F}$ is just the Euler class (2.23) of the subbundle $\tilde{\tau}_4 \subset \bar{\tau}_4$ defined by the Higgs vector $\mathbf{v}(x)$. Using the definition (3.7) of the curvature fields $\mathbb{F}_i$ the Euler class $\mathbb{F}$ is found in the version (3.22) which may be easily transformed back to (2.23). Observe that in Yang–Mills–Higgs theory the bundle $\bar{\tau}_4$ is not considered as embedded in a trivial SO(1,3) bundle τ_4 as was the case in our approach to the monopole field. Consequently, there are no extrinsic curvature fields $\mathbb{B}_i$ in $\bar{\tau}_4$ as long as one does not make use of the embedding $\bar{\tau}_4 \to \tau_4$.

After the Higgs vacuum configurations have been introduced within the framework of Yang–Mills–Higgs theory, the question must be faced now how the Yang–Mills vacuum, found in our approach, can be transformed to a Higgs vacuum? The Higgs vacuum shall turn out as the more convenient starting point for searching for the splitting (2.24) of the monopole field.

The desired transition is readily obtained by observing that for *any* configuration $\{\mathbb{A}_i\}$ the following change of connection ($\mathbb{A}_i \to \bar{\mathbb{A}}_i$) in $\bar{\tau}_4$ results in a Higgs vacuum:

$$\mathbb{A}_i \to \bar{\mathbb{A}}_i = \mathbb{A}_i + \epsilon_{ijk}v^k\mathbb{D}v^j. \tag{3.30}$$

This assertion is readily proved by simply calculating the covariant derivative $\bar{\mathbb{D}}v^i$ [cf. (2.23b)] of the Higgs vector field by means of the new connection $\bar{\mathbb{A}}_i$ instead of the old one ($\mathbb{A}_i$). This procedure directly leads to (3.27c). In the special case of our monopole field (2.3), where the connection $\{\mathring{\mathbb{A}}_i\}$ and its curvature $\{\mathring{\mathbb{F}}_i\}$ are zero [cf. (3.13) and (3.17)], the Higgs vacuum potentials $\{\bar{\mathbb{A}}_i\}$ become especially simple:

$$\bar{\mathbb{A}}_i = \epsilon_{ijk}v^k\mathring{\mathbb{D}}v^j. \tag{3.31}$$

Further, the covariant derivative of the Higgs vector field $\mathbf{v}$, agreeing with the first frame vector $\mathbf{v}(x)$ of the section $\mathbf{e}(x)$ in the trivial principal bundle Λ_4 via the embedding $\bar{\tau}_4 \to \tau_4$, is found after some calculations to be

$$\mathring{\mathbb{D}}v^i = \tilde{P}^i{}_j(\rho^{-1}\mathbf{e}^j + \mathbb{B}^j). \tag{3.32}$$

Here, the projector $\tilde{P}$ refers to the distribution $\tilde{\Delta}$ embedded into $\bar{\Delta}$

$$\tilde{P}^i{}_j = \delta^i{}_j + v^iv_j = -k^ik_j - h^ih_j. \tag{3.33}$$

Further, the one-forms $\{\mathbf{e}^j\}$ are the dual objects with respect to the triad vectors $\{\mathbf{e}_i\} = \{\mathbf{v},\mathbf{k},\mathbf{h}\}$ defining the SO(3) gauge $\bar{\mathbf{e}}(x)$:

$$\mathbf{e}^j(\mathbf{e}_i) = \delta^j{}_i. \tag{3.34}$$

The Higgs vacuum potentials $\bar{\mathbb{A}}_i$ (3.31) for the monopole field become in the chosen gauge

$$\bar{\mathbb{A}}_i = \epsilon_{ijk} v^k (\rho^{-1} \mathbf{e}^j + \mathbb{B}^j). \tag{3.35}$$

By direct use of (3.7) it can be shown that the curvature fields $\bar{\mathbb{F}}_i$ of the Higgs vacuum potentials $\bar{\mathbb{A}}_i$ (3.35) really lead to

$$\bar{\mathbb{F}}_i = -v_i \mathbb{F},$$

with $\mathbb{F}$ being the macroscopic monopole field (2.3).

In the next section, we start from the Higgs vacuum potentials (3.35) and obtain the splitting (2.24) into bound and radiation fields.

IV. BOUND AND RADIATION FIELDS OF THE SO(3) MONOPOLE

The vacuum configurations of Yang–Mills and Higgs type discussed in the preceding section are opposite to each other in the sense that, in the Yang–Mills vacuum case, the macroscopic monopole field $\mathbb{F}$ (2.3) is due to the Higgs field alone [cf. (3.14)]; whereas, in the Higgs vacuum case, the field $\mathbb{F}$ is generated by the Yang–Mills field alone [cf. (3.29)]. In the following, we look for an intermediate configuration such that both microscopic fields $(\mathbb{F}_i, v^i)$ contribute to the macroscopic field $\mathbb{F}$ in the way described in (2.24). That intermediate configuration is found by a change of connection similar to that one which led from Yang–Mills vacuum to Higgs vacuum [cf. (3.30)]. However, the macroscopic field $\mathbb{F}$ must be invariant with respect to such a change of connection. Hence, let us first study this invariance condition.

A. Invariance of the Euler class

Starting from a Higgs vacuum configuration we try to change the connection $\{\bar{\mathbb{A}}_i\}$ in $\bar{\tau}_4$ in the following way generalizing (3.30):

$$\bar{\mathbb{A}}_i \rightarrow \mathbb{A}_i = \bar{\mathbb{A}}_i + \mathbb{C}_i . \tag{4.1}$$

Here the one-forms $\mathbb{C}_i$ change tensorially with respect to an SO(3) gauge transformation (3.15) (as do the curvature fields $\mathbf{B}_i$ and $\mathbb{F}_i$ also):

$$\mathbb{C}'_i = \mathbb{C}_j S^j{}_i . \tag{4.2}$$

By means of the new connection $\{\mathbb{A}_i\}$ (4.1) the new curvature fields $\mathbb{F}_i$ are found via (3.7) as

$$\mathbb{F}_i = \bar{\mathbb{F}}_i + \bar{\mathbb{D}}\mathbb{C}_i + \tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{C}_j \wedge \mathbb{C}_k . \tag{4.3}$$

Consequently, the new Euler class due to the (unchanged) Higgs vector field $\mathbf{v}$ becomes [cf. (2.23)]

$$\mathbb{F} = \bar{\mathbb{F}} + \mathbf{d}(v^i \mathbb{C}_i). \tag{4.4}$$

Thus, the invariance ($\mathbb{F} \equiv \bar{\mathbb{F}}$) of the Euler class demands that $(v^i \mathbb{C}_i)$ be a total differential

$$v^i \mathbb{C}_i = \mathbf{d}\Psi, \tag{4.5}$$

where $\Psi(x)$ denotes some space-time function. We do not discuss here the most general change of connection leaving invariant the Euler class; rather we restrict ourselves to

$$v^i \mathbb{C}_i = 0. \tag{4.6}$$

Observe, that the transition (3.30) to the Higgs vacuum is just of this type. Further, the new covariant derivative of the Higgs vector field $\mathbf{v}(x)$ becomes

$$\mathbb{D}v^i = \bar{\mathbb{D}}v^i + \epsilon^{ij}{}_k v^k \mathbb{C}_j , \tag{4.7}$$

which may be inverted to yield [observe (3.27c)]

$$\mathbb{C}_i = -\epsilon_{ijk} v^k \mathbb{D}v^j . \tag{4.8}$$

Using the result (4.8), the new fields (4.3) can be cast into a more concise form. We have

$$\tfrac{1}{2}\epsilon_i{}^{jk}\mathbb{C}_j \wedge \mathbb{C}_k = -v_i\{\tfrac{1}{2}\epsilon_{jkl} v^l (\mathbb{D}v^j) \wedge (\mathbb{D}v^k)\}, \tag{4.9}$$

which just agrees with the bound field (2.24b). Consequently, the new fields (4.3) assume the special shape

$$\mathbb{F}_i = -v_i{}^{\prime}\mathbb{F} + \bar{\mathbb{D}}\mathbb{C}_i . \tag{4.10}$$

This form of the field is stringent if the assumption (4.6) is made and the starting configuration is a Higgs vacuum.

B. The bound field

For the desired identification (2.24b) of the bound field ${}^b\mathbb{F}$ (2.15b) we must explicitly find the new covariant derivative of the Higgs vector field $\mathbf{v}$. A solution to this problem is obtained readily by observing that in the static case (nonaccelerated monopole: $\dot{u} \equiv 0$) the radiation field ${}^r\mathbb{F}$ (2.15c) must vanish and the monopole field is then produced by the Higgs vector field $\mathbf{v}(x)$ alone. Since in this case the extrinsic curvature fields $\mathbf{B}_i$ (3.12) also vanish, we are left in (3.32) with

$$\mathring{\bar{\mathbb{D}}}v^i \rightarrow \rho^{-1}\tilde{P}^i{}_j \mathbf{e}^j . \tag{4.11}$$

Therefore, we now directly put[11]

$$\mathbb{D}v^i = \rho^{-1}\tilde{P}^i{}_j \mathbf{e}^j \tag{4.12}$$

and further

$$\mathbb{A}_i = \epsilon_{ijk} v^k \mathbb{B}^j . \tag{4.13}$$

These connection coefficients vanish whenever the particle is not accelerated. This means that the curvature of this connection can only be different from zero when the particle is accelerated and consequently irradiates electromagnetic energy momentum. Hence, the curvature fields are expected to be closely related to the radiation field, and this just tends into the desired direction.

Finally, combining (4.8) with (4.12) yields

$$\mathbb{C}_i = -\epsilon_{ijk} v^k \rho^{-1} \mathbf{e}^j . \tag{4.14}$$

Thus, (4.1) is really satisfied by (3.35), (4.13), and (4.14).

In this way, we have reached the second half of our aim, namely to find that SO(3) connection $\{\mathbb{A}_i\}$ which brings about the identification (2.24b). The solution is given in (4.13) along with (4.12). Now we turn to the first part (2.24a) of the problem.

C. The radiation field

In order to show that the change of connection (4.1), as given in (4.14), really produces the macroscopic radiation field ${}^r\mathbb{F}$ (2.15c) via (2.24a), we simply compute the curvature fields $\mathbb{F}_i$ of the connection (4.13) by means of (3.7). Observe that the form (3.8) for the Yang–Mills fields $\mathbb{F}_i$ is not valid here, because the modified connection (4.1) is no longer generated projectively from the canonical connection ω in τ_4 (see Ref. 12).

After some computations, the curvature fields $\mathbb{F}_i$ of the connection $\{\mathbb{A}_i\}$ (4.13) are found to be

$$\mathbb{F}_i = (\nu^j \mathbb{B}_j)\wedge \mathbb{C}_i - \nu_i(\mathbb{B}_j\wedge \mathbb{C}^j) + \nu_i(\nu^j \mathring{\mathbb{F}}_j). \tag{4.15}$$

Since the curvature fields $\{\mathring{\mathbb{F}}_j\}$ of the trivial connection $\{\mathring{\mathbb{A}}_i\}$ vanish [cf. (3.13)], a comparison of (4.15) with (4.10) yields for the radiation field ${}^r\mathbb{F}$ (see Ref. 13)

$${}^r\mathbb{F} = \mathbb{B}_j \wedge \mathbb{C}^j, \tag{4.16}$$

and for the covariant derivative of the one-forms $\mathbb{C}_i$,

$$\bar{\mathbb{D}}\mathbb{C}_i = (\nu^j \mathbb{B}_j)\wedge \mathbb{C}_i. \tag{4.17}$$

Moreover, the (Poincaré) dual of (4.17) is found to be

$$*\bar{\mathbb{D}}\mathbb{C}_i = (\nu^j\mathbb{B}_j)\wedge \mathbb{D}\nu_i, \tag{4.18}$$

which we shall need when studying the field equations below.

Now, inserting the extrinsic curvature fields $\mathbb{B}_i$ from (3.12) and the connection one-forms $\mathbb{C}_i$ from (4.14) into the result (4.16) one really finds the radiation field ${}^r\mathbb{F}$ as given by (2.15c). This can be seen immediately because the $\mathbb{C}_i$ (4.14) contain the $\mathfrak{so}(2)$ generator $\tilde{\mathbb{L}}$ in its SO(3) form[9]

$$\mathbb{C}_i = \rho^{-1}\tilde{L}_{ij}\mathbf{e}^j, \tag{4.19a}$$

$$\tilde{L}^i{}_j = (\nu_k \bar{L}^k)^i{}_j = -\nu_k \epsilon^{ki}{}_j = h^i k_j - k^i h_j, \tag{4.19b}$$

which also occurs in (2.15c).

Thus, we have also reached the second goal, namely the identification of the macroscopic radiation field ${}^r\mathbb{F}$ with the "longitudinal" part of the Yang–Mills field (2.24a):

$${}^r\mathbb{F} = \nu^i\mathbb{F}_i = \mathbb{B}_j\wedge \mathbb{C}^j = \rho^{-1}\mathbf{n}\wedge \tilde{\mathbb{L}}(\dot{\mathbf{u}}). \tag{4.20}$$

V. FIELD EQUATIONS AND CONSERVATION LAWS

After the different geometric meanings of the bound and radiation field of the monopole have been revealed, we study now the implications of the nonabelian field equations in connection with the splitting of the energy-momentum density as described in Sec. II C. There, the separate conservation law (2.22) for the radiative energy-momentum density ${}^r\mathbb{T}$ was partly a consequence of the Bianchi identity in the normal bundle $\hat{\tau}_4$ [cf. (2.16a) and (2.17)] and partly it came about as a happy circumstance: The vanishing of the second term on the right of the conservation law (2.21) could be shown only by means of an explicit computation. But a geometric necessity for the vanishing of this second term could not be found within the SO(2) framework. In this sense, there is no satisfactory geometric explanation for the separate conservation law (2.22) if one remains within the SO(2) symmetry.

The situation can be improved somewhat if one reconsiders this problem within the embedding SO(3) geometry.

Forming the energy-momentum density ${}^{\mathrm{YM}}\mathbb{T}$ of the Yang–Mills field $\{\mathbb{F}_i\}$ (4.10) similarly as in the abelian case (2.18),

$$8\pi\, {}^{\mathrm{YM}}T_{\mu\nu} = F_{i\mu}{}^{\rho}F^i{}_{\nu\rho} + {}^*F_{i\mu}{}^{\rho\,*}F^i{}_{\nu\rho}, \tag{5.1}$$

the density splits up into two parts on account of the orthogonality relation (4.6) and the covariant constancy (3.27c) of the Higgs vector field $\mathbf{v}(x)$:

$${}^{\mathrm{YM}}\mathbb{T} = {}^r\mathbb{T} + {}^{\perp}\mathbb{T}. \tag{5.2}$$

Here, the energy-momentum density ${}^{\perp}\mathbb{T}$ is due to the transverse part of the field $\{\mathbb{F}_i\}$

$$8\pi\, {}^{\perp}T_{\mu\nu} = f_{i\mu}{}^{\rho}f^i{}_{\nu\rho} + {}^*f_{i\mu}{}^{\rho\,*}f^i{}_{\nu\rho}, \tag{5.3a}$$

$$\mathbf{f}_i = \tfrac{1}{2}f_{i\mu\nu}\,\mathbf{d}x^\mu\wedge \mathbf{d}x^\nu \equiv \bar{\mathbb{D}}\mathbb{C}_i \tag{5.3b}$$

$$(\nu^i\mathbf{f}_i = 0).$$

If we now form the divergence of Eq. (5.2),

$$\nabla\cdot{}^{\mathrm{YM}}\mathbb{T} = \nabla\cdot{}^r\mathbb{T} + \nabla\cdot{}^{\perp}\mathbb{T}, \tag{5.4}$$

the terms referring to the Yang–Mills part (${}^{\mathrm{YM}}\mathbb{T}$) and the transverse part (${}^{\perp}\mathbb{T}$) can be shown to vanish separately:

$$\nabla\cdot{}^{\mathrm{YM}}\mathbb{T} = 0, \tag{5.5a}$$

$$\nabla\cdot{}^{\perp}\mathbb{T} = 0. \tag{5.5b}$$

Hence we are left with the desired conservation law (2.22) for the radiative part ${}^r\mathbb{T}$. On the other hand, the conservation laws (5.5) can be closely related to the vanishing of the curvature fields $\mathring{\mathbb{F}}_i$ for the Yang–Mills vacuum [cf. (3.13)]; in this sense one has traced back the separate conservation law for ${}^r\mathbb{T}$ to the existence of a Yang–Mills vacuum configuration in the embedding bundle $\bar{\tau}_4$. We shall study this effect now in some detail.

A. Conservation law for the Yang–Mills density ${}^{\mathrm{YM}}\mathbb{T}$

The divergence of the energy-momentum density ${}^{\mathrm{YM}}\mathbb{T}$ of the Yang–Mills fields $\mathbb{F}_i$ may be written in a form analogous to the abelian case (2.21) as ($\Theta := -{}^*\mathbb{D}^*$)

$$4\pi\nabla\cdot{}^{\mathrm{YM}}\mathbb{T} = {}^*(\Theta\mathbb{F}_i\wedge {}^*\mathbb{F}^i) - {}^*(\Theta\, {}^*\mathbb{F}_i\wedge \mathbb{F}^i). \tag{5.6}$$

Now the second term vanishes again on account of the Bianchi identity (3.10) and for the first term one finds

$$\Theta\mathbb{F}_i = 2\rho^{-2}\epsilon_{ij}{}^k\nu^j\mathbb{B}_k. \tag{5.7}$$

Therefore, using the curvature fields $\mathring{\mathbb{F}}_i$ in the bundle $\bar{\tau}_4$ in the form (3.8), we find for the first term on the right of (5.6)

$$\Theta\mathbb{F}_i\wedge {}^*\mathbb{F}^i = 2\rho^{-2}(\mathbb{D}\nu^i)\wedge \mathring{\mathbb{F}}_i. \tag{5.8}$$

The Yang–Mills vacuum has vanishing curvature $\mathring{\mathbb{F}}_i = 0$ (3.13) and thus one really finds the assertion (5.5a) satisfied. Observe again, that the continuity equation (5.5a) holds despite the fact that the homogeneous field equations (3.24a) are not valid here [cf. (5.7)].

B. Conservation law for the transverse density ${}^{\perp}\mathbb{T}$

As for the continuity equation (5.5b) for the energy-momentum density ${}^{\perp}\mathbb{T}$ one starts again with the identity ($\bar{\Theta} := -{}^*\bar{\mathbb{D}}^*$)

$$4\pi\nabla\cdot{}^{\perp}\mathbb{T} = {}^*(\bar{\Theta}\,\mathbf{f}_i\wedge {}^*\mathbf{f}^i) - {}^*(\bar{\Theta}\,{}^*\mathbf{f}_i\wedge \mathbf{f}^i). \tag{5.9}$$

Since the two-forms $\mathbf{f}_i \equiv \bar{\mathbb{D}}\mathbb{C}_i$ are not gauge fields (as are the curvature fields $\mathbb{F}_i$), there does not exist such a relation as a Bianchi identity, which automatically ensures the vanishing of the second term on the right of the divergence equations. Therefore, we have to show here explicitly that the second term nevertheless vanishes.

Indeed, an actual computation shows that the extrinsic curvature fields again emerge in the form

$$\bar{\Theta}\,{}^*\mathbf{f}_i = \rho^{-2}\tilde{P}_i{}^j\mathbb{B}_j, \tag{5.10}$$

and hence the second term in (5.9) is found to be

$$\bar{\Theta}\,{}^*\mathbf{f}_i\wedge \mathbf{f}^i = -\rho^{-2}(\mathbb{D}\nu^i)\wedge \mathring{\mathbb{F}}_i, \tag{5.11}$$

which again vanishes on account of (3.13). Here, we have used also the covariant derivative $\bar{\mathbb{D}}$ referring to the Higgs vacuum connection $\{\bar{\mathbb{A}}_i\}$ (3.35). The reason for this choice is that the $\mathbf{f}_i$ are the exterior covariant derivatives of a tensor one-form $\mathbb{C}_i$ (5.3b) with respect to the Higgs vacuum connection $\{\bar{\mathbb{A}}_i\}$. This simplifies the calculations by observing that the Higgs vacuum fields $\bar{\mathbb{F}}_i$ occurring on the double differentiation process

$$\bar{\mathbb{D}}\bar{\mathbb{D}}\mathbb{C}_i = \epsilon_i{}^{jk}\bar{\mathbb{F}}_j \wedge \mathbb{C}_k \tag{5.12}$$

are especially simple [cf. (3.29)].

Similarly, a detailed computation of the first term on the right of (5.9) gives

$$\bar{\Theta}\mathbf{f}_i = \rho^{-2}\epsilon_i{}^{jk}\nu_k\mathbb{B}_j \tag{5.13}$$

and therefore

$$\bar{\Theta}\mathbf{f}_i \wedge {}^*\mathbf{f}^i = \rho^{-2}\epsilon^{jkl}\nu_l\mathbb{C}_j \wedge \mathring{\mathbb{F}}_k, \tag{5.14}$$

which again vanishes on account of (3.13).

In this way, we have verified the continuity equation for the transverse density ${}^{\perp}\mathbb{T}$ (5.5b). Both conservation laws (5.5a) and (5.5b) could be related to the existence of a Yang–Mills vacuum in the embedding bundle $\bar{\tau}_4$. This is then also true for the radiative density ${}^r\mathbb{T}$ which is given in terms of ${}^{\mathrm{YM}}\mathbb{T}$ and ${}^{\perp}\mathbb{T}$ by Eq. (5.2). Observe that the radiative conservation law (2.22) follows now from (5.4) and (5.5).

C. Radiative four-momentum ${}^r\mathbf{P}$

Since both ${}^{\mathrm{YM}}\mathbb{T}$ and ${}^{\perp}\mathbb{T}$ form together the radiative density ${}^r\mathbb{T}$, the question now arises to what extent this is also true for the corresponding integral quantities. The total irradiated four-momentum ${}^r\mathbf{P}$ of the monopole is given by an integration of ${}^r\mathbb{T}$ over a suitable three-chain $C^3 \subset M_4$ [cf. (5.2)].

$${}^r\mathbf{P} = \int_{C^3} {}^{*r}\mathbb{T} = \int_{C^3} {}^{*\mathrm{YM}}\mathbb{T} - \int_{C^3} {}^{*\perp}\mathbb{T} =: {}^{\mathrm{YM}}\mathbf{P} - {}^{\perp}\mathbf{P}. \tag{5.15}$$

Since all fields involved ($\mathbb{F}_i, {}^r\mathbb{F}, \mathbf{f}_i$) have explicitly been calculated in the foregoing sections, one can carry through the integration in (5.15) (see Appendix). One finds that the special partitioning (2.24) of the total Yang–Mills field $\mathbb{F}_i$ (4.10) into a longitudinal ${}^r\mathbb{F}$ and transverse part ($\bar{\mathbb{D}}\mathbb{C}_i$) just splits up the four-momentum ${}^{\mathrm{YM}}\mathbf{P}$ into equal transverse and longitudinal contributions

$$\frac{1}{2}\,{}^{\mathrm{YM}}\mathbf{P} = {}^r\mathbf{P} = {}^{\perp}\mathbf{P} = -\frac{2}{3}\int_{-\infty}^{s} ds'(\dot{\mathbf{u}}\cdot\dot{\mathbf{u}})_{(s')}\mathbf{u}(s'). \tag{5.16}$$

Here the integration runs over the past history ($-\infty < s' \leqslant s$) of the monopole up to the present moment (s). The energy W irradiated per unit time by the point monopole is given now correctly by the well-known Larmor radiation rate (apart from dimensional constants)

$$W = (\mathbf{u}\cdot{}^r\mathbf{P}) = -\tfrac{2}{3}(\dot{\mathbf{u}}\;\dot{\mathbf{u}}). \tag{5.17}$$

VI. CONCLUSIONS

The foregoing arguments present an example for the original assertion that the embedding of abelian electrodynamics into a higher-dimensional nonabelian field theory may produce (already in the classical domain) further interesting effects besides the emergence of smooth particlelike solutions. Whereas the latter effect mainly refers to the static case and avoides the divergent problems for the classical field energy, the new aspects of the embedding now aim at the radiation parts of the fields when the sources are arbitrarily accelerated.

In the case of a single, arbitrarily moving monopole, the embedding SO(2)→SO(3) could be performed in such a way that an intermediate field configuration between the two extremes of Higgs and Yang–Mills vacua arose. In the Higgs vacuum configuration (3.26), the macroscopic monopole field $\mathbb{F}$ (2.3) is produced by the Yang–Mills field $\mathbb{F}_i$ via its longitudinal component ($\nu^i\mathbb{F}_i$) [cf. (3.29)]. On the other hand, the macroscopic field $\mathbb{F}$ in a Yang–Mills vacuum configuration is traced back to the Higgs field $\mathbf{v}(x)$ [cf. (3.14)]. The existence of two *different microscopic* field configurations leading to the *same macroscopic* monopole field is due to the possibility of changing the connection $\bar{\omega}$ in the embedding SO(3) bundle $\bar{\tau}_4$ without altering the Euler class $\mathbb{F}$ of the reduced bundle $\tilde{\tau}_4$ [cf. (4.4) and (4.5)].

This freedom on the microscopic level was exploited in order to choose the microscopic fields $\mathbb{F}_i$ in such a way that the radiation part ${}^r\mathbb{F}$ (2.15c) of the macroscopic field $\mathbb{F}$ is produced by the microscopic Yang–Mills field $\mathbb{F}_i$ via its longitudinal component ($\nu^i\mathbb{F}_i$), whereas the macroscopic bound part ${}^b\mathbb{F}$ is built by the Higgs vector field $\mathbf{v}(x)$ [cf. (2.24)].

It seems to us that the macroscopic splitting of the entity $\mathbb{F}$ into its bound and radiation parts has now been linked to a more fundamental gauge-independent separation on the microscopic level, where different kinds of fields can be used for the different aspects of the macroscopic entity $\mathbb{F}$. Unfortunately, this could be demonstrated only for the single-particle case; it remains to be seen whether the present results also apply to the general multimonopole configuration.

APPENDIX: COMPUTATION OF THE FOUR-MOMENTA

Using the extrinsic curvature fields $\mathbb{B}_i$ (3.12) and the changes of connections $\mathbb{C}_i$ (4.14) yields the radiation field ${}^r\mathbb{F}$ via (4.16). Its explicit form has already been presented in (2.15c). Thus the radiative density ${}^r\mathbb{T}$ (2.19b) turns into

$${}^r\mathbb{T} = -(4\pi\rho^2)^{-1}\{(\dot{\mathbf{u}}\dot{\mathbf{u}}) + (\mathbf{v}\dot{\mathbf{u}})^2\}\mathbf{n}\otimes\mathbf{n}. \tag{A1}$$

Smilarly, inserting the transverse fields $\mathbf{f}_i = \bar{\mathbb{D}}\mathbb{C}_i$ (5.3b) as given by (4.17) into the definition of the transverse density ${}^{\perp}\mathbb{T}$ (5.3a) yields

$${}^{\perp}\mathbb{T} = 2(4\pi\rho^2)^{-1}(\mathbf{v}\dot{\mathbf{u}})^2\mathbf{n}\otimes\mathbf{n}. \tag{A2}$$

Finally, the Yang–Mills density ${}^{\mathrm{YM}}\mathbb{T}$ (5.1) is found as

$${}^{\mathrm{YM}}\mathbb{T} = (4\pi\rho^2)^{-1}\{(\mathbf{v}\dot{\mathbf{u}})^2 - (\dot{\mathbf{u}}\dot{\mathbf{u}})\}\mathbf{n}\otimes\mathbf{n}. \tag{A3}$$

Obviously, all three kinds of densities are of the same structure and the integrations (5.15)

$$\mathbf{P} = \int_{C^3} {}^*\mathbb{T} \tag{A4}$$

can be done very conveniently by means of the technique of retarded integration.[14]

In this method, the three-cell $\mathbf{d}C^3$ on the hypersurface C^3 of integration is expressed by a solid angle element $d\Omega'$

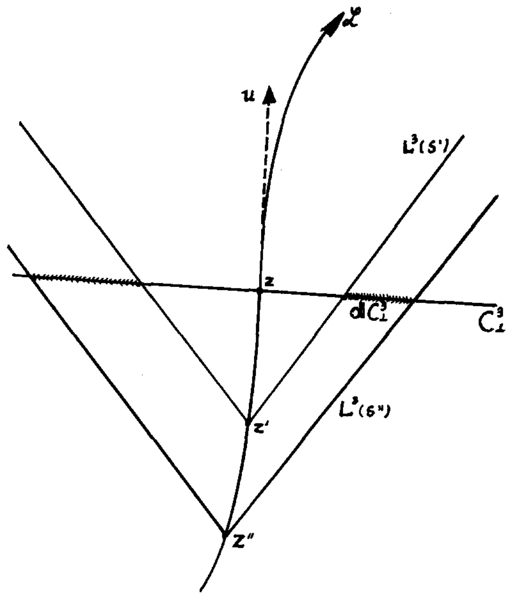


FIG. 2. The hypersurface C^3 to be used for the integrations (A4) is chosen as the three-plane $C^3_\perp$ cutting the world line $\mathfrak{L}:\mathbf{x} = \mathbf{z}(s)$ orthogonally at the event $\mathbf{z}$, where the momentum $\mathbf{P}$ is to be determined. The proper time for this event is s and the four-velocity is $\mathbf{u}$. The three-cell $\mathbb{d}\, C^3_\perp$ is cut out of $C^3_\perp$ by two light cones $L^3(s')$ and $L^3(s'')$ with vertices at the earlier events $\mathbf{z}^1 = \mathbf{z}(s')$ and $\mathbf{z}'' = \mathbf{z}(s'')$ $(s'' < s' < s)$ on the world line $\mathfrak{L}$. The integration over $C^3_\perp$ can now be split up into an angular integration over the intersections $L^3 \wedge C^3_\perp$, which have $s' = \text{const}$, and into an integration in orthogonal direction to these intersections, where proper time s' is varying. The whole three-plane $C^3_\perp$ is swept out when $\mathbf{z}'$ tends to $\mathbf{z}$ and $\mathbf{z}''$ is moved backwards along the world line up to the infinite past $(s'' \to -\infty)$.

and by the proper time element ds' at earlier times (see Fig. 2). The three-surface C^3 is chosen here to be that three-plane $C^3_\perp$ which cuts the world-line $\mathfrak{L}:\mathbf{x} = \mathbf{z}(s)$ of the monopole orthogonally in that event $\mathbf{z}$ at proper time s where the four-momentum $\mathbf{P}$ has to be calculated. The corresponding four-velocity is $\mathbf{u}$. Therefore, the three-cell $\mathbb{d}C^3_\perp$ on $C^3_\perp$ is

$$\mathbb{d}C^3_\perp = {}^*\mathbf{u}(s)\,\rho^2_{(s')}\,\{\mathbf{n}(s')\cdot\mathbf{u}(s)\}^{-1}\,ds'\,d\Omega\,'. \tag{A5}$$

If one computes now the values of the density three-forms ${}^*\mathbf{T}$ upon the three-cell $\mathbb{d}C^3_\perp$ the unwieldy factor $(\mathbf{n}(s')\cdot\mathbf{u}(s))^{-1}$ and the retarded distance $\rho_{(s')}$ drop out leaving us with the very simple angular integrations

$$\int_{C^3_\perp} {}^{*\mathrm{YM}}\mathbb{T} = \int_{-\infty}^{s} ds' \int_{\Omega'} \frac{d\Omega\,'}{4\pi}(\mathbf{v}\dot{\mathbf{u}})^2 - (\dot{\mathbf{u}}\dot{\mathbf{u}})\}_{(s')}\mathbf{n}(s'), \tag{A6a}$$

$$\int_{C^3_\perp} {}^{*r}\,\mathbb{T} = -\int_{-\infty}^{s} ds' \int_{\Omega'} \frac{d\Omega\,'}{4\pi}\{(\mathbf{v}\dot{\mathbf{u}})^2 + (\dot{\mathbf{u}}\dot{\mathbf{u}})\}_{(s')}\mathbf{n}(s'), \tag{A6b}$$

$$\int_{C^\perp_3} {}^{*\perp}\mathbf{T} = 2\int_{-\infty}^{s} ds' \int \frac{d\Omega\,'}{4\pi}(\mathbf{v}\dot{\mathbf{u}})^2_{(s')}\mathbf{n}(s'). \tag{A6c}$$

The angular integrals are readily found to be

$$\int \frac{d\Omega\,'}{4\pi}(\mathbf{v}\dot{\mathbf{u}})^2_{(s')}\mathbf{n}(s') = -\frac{1}{3}(\dot{\mathbf{u}}\dot{\mathbf{u}})_{(s')}\mathbf{u}(s'), \tag{A7a}$$

$$\int \frac{d\Omega\,'}{4\pi}\mathbf{n}(s') = \mathbf{u}(s'). \tag{A7b}$$

With these results, the four-momenta $\mathbf{P}$ (5.15) combine in the way given by (5.16).

[1] G. t'Hooft, Nucl. Phys. B **79**, 276 (1974).

[2] P. Goddard and D. I. Olive, Rep. Prog. Phys. **41**, 1357 (1978).

[3] C. Teitelboim, D. Villarroel, and C. G. van Weert, Riv. Nuovo Cimento **9**, (1980).

[4] M. Sorg. Z. Naturforsch. **37a**, 201 (1982). In that paper, the Euler class has been identified with the dual (${}^*\mathbb{F}$) of the field in order to obtain a geometrization of the well-known Liénard–Wiechert field of classical point particle electrodynamics. There, magnetic monopoles are usually excluded. In the present paper, the Euler class is identified with the field ($\mathbb{F}$) itself in agreement with the current approach in monopole literature. This change of notation is of no influence on the geometric reasoning. Throughout the paper, we absorb the coupling constant (charge unit) into the potentials (fields) such that they are of the dimension of an inverse (square) length.

[5] D. Fauth and M. Sorg, J. Math. Phys. **25**, 1301 (1984).

[6] C. Teitelboim, Phys. Rev. D **1**, 1572 (1970).

[7] The expression (2.23) for the macroscopic field $\mathbb{F}$ in terms of the microscopic fields $\{\mathbb{F}_i, v^i\}$ was first mentioned by t'Hooft.[1] Later on, t'Hooft's proposal was rejected,[2] because it leads to a pointlike distribution of magnetic charge even if one has smooth gauge fields $\mathbb{F}_i$. In the present context, we prefer the original choice of t'Hooft (2.23), because it agrees with the curvature in the subbundle $\bar{\tau}_4$ of $\bar{\tau}_4$ if the connection $\tilde{\omega}$ in $\bar{\tau}_4$ is obtained from the connection $\bar{\omega}$ in $\bar{\tau}_4$ by projection onto the subalgebra $\mathfrak{so}(2) \subset \mathfrak{so}(3)$: $\tilde{\omega} = \bar{\omega}|_{\mathfrak{so}(2)}$. This choice has the advantage that the fiber bundle for the description of the macroscopic fields follows from "microscopic" bundle in a very simple and natural way. If one does not want to join this point of view, one has to face the problem how the "macroscopic" and "microscopic" bundles should be related to each other.

[8] Putting $i = 1$ in (3.4a) yields the potential $\mathbb{A}_1$ (2.12) for the monopole field $\mathbb{F}$.

[9] The generators $\{\bar{L}^i\}$ used in (3.16) form an adjoint representation of the Lorentz group generators $\{L^i\}$ occurring in (3.3): $(\bar{L}^i)_{jk} = -\epsilon^i_{\ jk}$.

[10] The fact that the monopole charge g is negative is closely related to the integrability of the two-distribution $\bar{\Delta}$ generating the tangent subbundle $\bar{\tau}_4$. Whereas the integral surfaces $\bar{S}$ of $\bar{\Delta}$ for $g = -1$ do surely exist, the Gauss–Bonnet theorem forbids the existence of integral surfaces when $g < -1$. Moreover, it forbids the existence of simply connected integral surfaces when $g > -1$ [D. Fauth, thesis, Stuttgard, 1983 (unpublished)].

[11] The occurrence of the projector $\bar{\mathbf{P}}$ in the covariant derivative (4.12) of the Higgs vector $\mathbf{v}(x)$ ensures that the property of umbilicality of the two-distribution $\bar{\Delta}$ with respect to its normal section $\mathbf{v}$ in $\bar{\tau}_4$ is left unchanged during the change of connection $\{\mathbb{A}_i\}$ [(4.1) and (4.14)] (see Ref. 5).

[12] Observe, however, that the extrinsic curvature fields remain covariantly constant also with respect to the new connection (4.13): $\mathbb{D}\mathbb{B}_i \equiv \mathring{\mathbb{D}}\mathbb{B}_i \equiv 0$.

[13] The dual of the radiation field ${}^r\mathbb{F}$ is

$${}^{*r}\mathbb{F} = -(\mathbb{D}v^i)\wedge\mathbb{B}_i = -\mathbb{d}(v^i\mathbb{B}_i),$$

which agrees with the normal bundle curvature (2.17) (choose the special gauge $v^i = \delta^i{}_1$).

[14] M. Sorg, Z. Naturforsch. **33a**, 619 (1978).

Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform

G. Beylkin
Schlumberger–Doll Research, Old Quarry Road, Ridgefield, Connecticut 06877-4108

(Received 8 May 1984; accepted for publication 20 July 1984)

This paper treats the linearized inverse scattering problem for the case of variable background velocity and for an arbitrary configuration of sources and receivers. The linearized inverse scattering problem is formulated in terms of an integral equation in a form which covers wave propagation in fluids with constant and variable densities and in elastic solids. This integral equation is connected with the causal generalized Radon transform (GRT), and an asymptotic expansion of the solution of the integral equation is obtained using an inversion procedure for the GRT. The first term of this asymptotic expansion is interpreted as a migration algorithm. As a result, this paper contains a rigorous derivation of migration as a technique for imaging discontinuities of parameters describing a medium. Also, a partial reconstruction operator is explicitly derived for a limited aperture. When specialized to a constant background velocity and specific source–receiver geometries our results are directly related to some known migration algorithms.

I. INTRODUCTION

The interpretation of seismic reflection data, ultrasound reflectivity imaging in medical applications, and various other methods of nondestructive evaluation require a solution to the inverse scattering problem. The inverse scattering problem is nonlinear and different approximate solutions have been suggested over the years. Some of the most useful in practice are the so-called *migration* schemes in geophysics. References 1–9 contain examples of such algorithms. By a migration scheme (algorithm) in this paper we understand a *technique of imaging discontinuities* of parameters describing the medium.

It must be emphasized that the construction of these approximate solutions involves (explicitly or implicitly) two major, separate steps: the first step is a linearization of the inverse problem and the second step is the solution of the linearized inverse problem.

In this paper the linearization is accomplished by a perturbation technique equivalent to the distorted wave Born approximation. We derive an integral equation formulation of the linearized inverse scattering problem for the Helmholtz equation. Analogous integral equations can be derived for fluids with variable density and for elastic solids.

The primary concern in this paper is the solution of the linearized inverse scattering problem. It requires the inversion of an integral operator with an oscillatory kernel. This operator is related—via the one-dimensional Fourier transform—to the causal generalized Radon transform (GRT).

The appearance of the GRT has a simple physical explanation. In all cases where it is impossible to make measurements directly inside the medium of interest, the only feasible measurements are *integrals* of certain combinations of parameters describing the medium. If these integrals are line integrals or integrals over hyperplanes we are dealing with the classical Radon transform.[10] Integrals with a weight function over more general hypersurfaces represent the generalized Radon transform. (Note that the problem of recovering a function from a knowledge of integrals over geometrical objects such as hypersurfaces can be viewed as a problem in the field of integral geometry.)

To solve the linearized inverse scattering problem we invert the GRT. The inversion of the GRT is of mathematical interest by itself.[11–16] In Refs. 14–16 a general explicit technique was developed for inverting the GRT. As we show here, this technique leads to an asymptotic solution of the linearized inverse problem of wave propagation.

Miller[7] recognized that seismic imaging for the general case of variable background and irregular source–receiver geometry could be cast as the problem of inverting a GRT. He derived an approximate imaging algorithm, using weighted and filtered backprojection of the data, and applied it to both synthetic and real examples. The weighting suggested in Ref. 7 differs from what we obtain by an obliquity factor.

In this paper we give an exact, formal answer to what is the proper weighting and filtering of the data, and, most important, to what is the nature of the reconstructed image.

The inversion of the GRT requires the introduction of Fourier integral operators (FIO). A special role is played by a FIO of the form $F = R^*KR$ (see Refs. 14–16). Here, R denotes the GRT, R^* is an operator dual to R, and K is a one-dimensional convolution operator; R^* is also known as the generalized backprojection operator (GBO). The Fourier integral operator F was studied in Refs. 14 and 16, and it was shown that by properly choosing the convolution operator K and the weight function of the GBO the problem of inverting the GRT can be reduced to that of solving a Fredholm integral equation.

By modifying slightly the arguments used in Refs. 14–16 and exploiting the fact that F is "almost" the identity operator we rigorously establish a class of migration algorithms as *approximate* solutions of the linearized inverse scattering problem. The approximation amounts to using only the first term of an asymptotic expansion for the "inverse" GRT. Due to the nature of the asymptotics we are able to give a precise meaning to what is reconstructed by this first-order inversion for arbitrary configurations of sources and receivers, including the case of limited view angles. In particular, we show that the (locations of) *discontinuities* of the unknown function describing the medium are recovered, rather than the function itself.

We derive an algorithm for recovering these discontinuities for variable background velocity and an arbitrary configuration of sources and receivers. Our derivation is valid as long as certain physically meaningful conditions on the global structure of rays are satisfied.

Until now, mathematically rigorous justifications relating migration to inverse scattering have been given only for migration schemes with constant background velocities and special source-receiver geometries. Previous workers, notably Norton and Linser[17] and Rose,[18] have made the connection between the Radon transform and the linearized inverse scattering problem. Norton and Linser[17] derived explicit inversion formulas for a constant background velocity and coincident sources and receivers for plane, spherical, and cylindrical apertures. They obtained certain backprojection algorithms as approximations to the exact inversion formulas. These backprojections are migration schemes in the sense of our definition. Specializing our results to their case we obtain the backprojection algorithm of Ref. 17 for a plane aperture. However, in the case of a spherical aperture our answer is different from that of Ref. 17. Our approximation remains valid even if the point of reconstruction is not close to the center of the spherical aperture. As an additional example, we obtain a migration algorithm for sources and receivers located on a plane and separated by a fixed distance.

II. LINEARIZATION OF THE INVERSE PROBLEM

To linearize the inverse scattering problem we use a standard procedure which is essentially a small perturbation technique. Formally, this procedure can be stated as follows. Consider an equation of the form

$$Lv = g, \tag{2.1}$$

where the operator L,

$$L = L_0 + L_1,$$

is a perturbation of a known operator L_0 by an operator L_1. Assuming that we can—exactly or approximately—invert the operator L_0, we look for a solution of the equation (2.1) in the form

$$v = v^{\text{in}} + v^{\text{sc}},$$

where

$$v^{\text{in}} = L_0^{-1} g,$$

and L_0^{-1} denotes the inverse operator. Substituting this into (2.1) and applying L_0^{-1} to both sides of the equation we obtain

$$v^{\text{sc}} = -L_0^{-1} L_1 v^{\text{in}} - L_0^{-1} L_1 v^{\text{sc}}.$$

By making the (single scattering) approximation

$$v^{\text{sc}} \approx -L_0^{-1} L_1 v^{\text{in}}, \tag{2.2}$$

we linearize the relation between the function v^{sc} and the perturbation L_1.

Let us now apply this procedure to the Helmholtz equation which describes wave propagation in a fluid of constant density. Without complicating the necessary arguments we treat this equation in n space dimensions since the dimension enters only as a parameter.

First, we consider the case where the perturbation is about a constant background velocity, which we take to be 1. Assuming then that the index of refraction $n(x)$ of the medium in some region X can be written as

$$n^2(x) = 1 + f(x),$$

the goal is to characterize the function $f(x)$ from observations of the scattered field on the boundary ∂X of the region X, as generated by a known incident field. Let us assume that the incident field is due to a point source located at a point η on the boundary ∂X. The operator L_0 is the Helmholtz operator for a constant-velocity medium, i.e.,

$$L_0 = \nabla_x^2 + k^2,$$

where ∇_x^2 is the Laplacian operator in spatial variables and the perturbation is

$$L_1 = k^2 f(x). \tag{2.3}$$

The incident field is given by the Green's function

$$v^{\text{in}}(x,\eta) = -\frac{i}{4}\left(\frac{k}{2\pi|x-\eta|}\right)^{(n-2)/2} H^{(1)}_{(n-2)/2}(k|x-\eta|),$$

where $H^{(1)}_{(n-2)/2}$ is the Hankel function of the first kind. We use the first term of the asymptotic expansion of the Hankel function to approximate v^{in} by

$$v^{\text{in}}(x,\eta) \approx e^{-i(\pi/2)(n+1)/2} \frac{k^{(n-3)/2}}{2(2\pi|x-\eta|)^{(n-1)/2}} e^{ik|x-\eta|}, \tag{2.4}$$

which is exact for n = 3. Since the operator L_0^{-1} is defined in terms of the Green's function, we obtain from (2.2) the (first term of the asymptotic expansion of) integral representation of the singly scattered field v^{sc} as

$$v^{\text{sc}}(k,\xi,\eta) = \frac{(-ik)^{n-1}}{4(2\pi)^{n-1}} \int_X \frac{e^{ik|x-\xi|} e^{ik|x-\eta|}}{(|x-\xi||x-\eta|)^{(n-1)/2}} f(x)\,dx. \tag{2.5}$$

For n = 3, (2.5) yields v^{sc} as

$$v^{\text{sc}}(k,\xi,\eta) = -\frac{k^2}{16\pi^2} \int_X \frac{e^{ik|x-\xi|}}{|x-\xi|} \frac{e^{ik|x-\eta|}}{|x-\eta|} f(x)\,dx.$$

Let us now consider the case of variable background. Assuming that the index of refraction of the medium in some region X is of the form $n^2(x) = n_0^2(x) + f(x)$, where $n_0(x)$ is known, the problem, again, is to characterize the function $f(x)$ from observation of the scattered field on the boundary ∂X. Now L_0 is the operator

$$L_0 = \nabla_x^2 + k^2 n_0^2.$$

The perturbation L_1 is of the form (2.3)

Again, we choose the incident field to be due to a point source, so that

$$(\nabla_x^2 + k^2 n_0^2) v^{\text{in}}(x,\eta) = \delta(x-\eta), \tag{2.6}$$

where η indicates the position of the source. As an approximate solution of (2.6) we use, in place of (2.4), the first term of the geometrical optics approximation

$$v^{\mathrm{in}}(x,\eta) \approx e^{-i(\pi/2)(n+1)/2} k^{(n-3)/2} A^{\mathrm{in}}(x,\eta) e^{ik\phi^{\mathrm{in}}(x,\eta)}. \tag{2.7}$$

Here, $\phi^{\mathrm{in}}(x,\eta)$ is the phase function and satisfies the eikonal equation

$$(\nabla_x \phi^{\mathrm{in}})^2 = n_0^2(x). \tag{2.8}$$

Function $A^{\mathrm{in}}(x,\eta)$ is the amplitude and satisfies the transport equation along the ray connecting the source at η on the boundary ∂X and the point x inside the region X,

$$A^{\mathrm{in}} \nabla_x^2 \phi^{\mathrm{in}} + 2\nabla_x A^{\mathrm{in}} \cdot \nabla_x \phi^{\mathrm{in}} = 0. \tag{2.9}$$

We note that the factor $e^{-i(\pi/2)(n+1)/2} k^{(n-3)/2}$ in (2.7) is obtained by matching the geometrical optics approximation (2.7) with the exact solution in the neighborhood of the source for large k.

If we interchange source and receiver positions then (2.7) also yields the approximation $v^{\mathrm{out}}(x,\xi)$ for the Green's function (which defines the kernel of the operator L_0^{-1}). Thus, we arrive at

$$v^{\mathrm{out}}(x,\xi) \approx e^{-i(\pi/2)(n+1)/2} k^{(n-3)/2} A^{\mathrm{out}}(x,\xi) e^{ik\phi^{\mathrm{out}}(x,\xi)}, \tag{2.10}$$

where $\phi^{\mathrm{out}}(x,\xi)$ satisfies the eikonal equation in (2.8) and $A^{\mathrm{out}}(x,\xi)$ satisfies the transport equation along the ray connecting the point x and the receiver at the point ξ on the boundary ∂X,

$$A^{\mathrm{out}} \nabla_x^2 \phi^{\mathrm{out}} + 2\nabla_x A^{\mathrm{out}} \cdot \nabla_x \phi^{\mathrm{out}} = 0. \tag{2.11}$$

In general, one can solve the eikonal equation (2.8) by ray tracing. The transport equations in (2.9) and (2.11) then reduce to ordinary differential equations along rays.[19] If the background index of refraction $n_0(x)$ is discontinuous then the rays satisfy Snell's law on surfaces of discontinuities and appropriate transmission coefficients have to be used in computing the amplitudes on these surfaces. We formulate the assumptions we need to make about the global structure of rays—and, therefore, about the background index of refraction—in the next section.

Combining (2.2), (2.3), (2.7), and (2.10) we find the (first term of the asymtotic expansion of) integral representation of the singly scattered field

$$v^{\mathrm{sc}}(k,\xi,\eta) = (-ik)^{n-1} \int_X e^{ik\phi^{\mathrm{out}}(x,\xi)} e^{ik\phi^{\mathrm{in}}(x,\eta)} \times A^{\mathrm{out}}(x,\xi) A^{\mathrm{in}}(x,\eta) f(x) dx, \tag{2.12}$$

as a function of the receiver position ξ, the source position η, and wave number k. The integral representation (2.12) is an integral equation for the unknown function f.

Analogous integral equations can be derived for fluids with variable density and for elastic solids. We shall present the derivation elsewhere. In these cases integral equations of the type in (2.12) relate the singly scattered field to a combination of parameters characterizing the medium. For elastic media we obtain a system of four integral equations corresponding to p–p, p–s, s–p, and s–s scattered fields, and the phase functions ϕ^{in} and ϕ^{out} satisfy different eikonal equations corresponding to the indices of refraction of p and s waves. The amplitudes $A^{\mathrm{in}}(x,\eta)$ and $A^{\mathrm{out}}(x,\xi)$ satisfy the corresponding transport equations along the rays connecting points x,η and x,ξ, respectively. In the following sections we consider the integral equation (2.12) in a form which covers these cases.

III. THE INTEGRAL EQUATION OF THE LINEARIZED INVERSE PROBLEM AND THE CAUSAL GRT

It was shown in the previous section that the linearization of the inverse scattering problem leads to the integral equation

$$v(k,\xi,\eta) = (-ik)^{n-1} \int_X f(x) a(x,\xi,\eta) e^{ik\phi(x,\xi,\eta)}\, dx, \tag{3.1}$$

where X is the domain of definition of the unknown function $f(x)$, ξ and η are points on the boundary ∂X corresponding to receiver and source locations, and k is the wave number. The phase function $\phi(x,\xi,\eta)$ is the sum of two phase functions

$$\phi(x,\xi,\eta) = \hat{\phi}(x,\xi) + \tilde{\phi}(x,\eta), \tag{3.2}$$

which satisfy the eikonal equations

$$(\nabla_x \hat{\phi}(x,\xi))^2 = \hat{n}^2(x), \tag{3.3a}$$

and

$$(\nabla_x \tilde{\phi}(x,\eta))^2 = \tilde{n}^2(x). \tag{3.3b}$$

In (3.3a) and (3.3b), $\hat{n}$ and $\tilde{n}$ are indices of refraction, i.e., positive bounded functions. We have replaced ϕ^{out} and ϕ^{in} by $\hat{\phi}$ and $\tilde{\phi}$. The function $a(x,\xi,\eta)$ in (3.1) replaces the product of the amplitudes A^{out} and A^{in}

$$a(x,\xi,\eta) = A^{\mathrm{out}}(x,\xi) A^{\mathrm{in}}(x,\eta). \tag{3.4}$$

Both A^{out} and A^{in} are positive since they are solutions of the transport equations in (2.9) and (2.11). This is true even for discontinuous indices of refraction $\hat{n}$ and $\tilde{n}$ as long as the global structure of rays satisfies the assumptions formulated later in this section. Therefore, $a(x,\xi,\eta)$ is positive on $X \times \partial X \times \partial X$ and can be called a weight function. We assume, in addition, that $a(x,\xi,\eta)$ is infinitely differentiable, namely, $a(x,\xi,\eta) \in C^\infty(\bar{X} \times \partial X \times \partial X)$, where $\bar{X}$ is any compact set contained in X.

The integral equation (3.1) is related to a causal GRT. To see this, consider the transform R defined by

$$(Rf)(t,\xi,\eta) = \int f(x) a(x,\xi,\eta) \delta(t - \phi(x,\xi,\eta)) dx, \quad \text{for } t \geqslant 0,$$

$$(Rf)(t,\xi,\eta) = 0, \quad \text{for } t \leqslant 0. \tag{3.5}$$

We call the transform R in (3.5) the causal GRT for obvious reasons and note that the transform R agrees with the GRT as defined in Refs. 14–16 for $t \geqslant 0$. However, the integral in (3.5) is not defined for $t < 0$, and it is natural to set $(Rf)(t,\xi,\eta) = 0$ for $t \leqslant 0$. Since in this article we consider only the causal GRT, we drop the word causal.

The Fourier transform $(Rf)\hat{}\,(k,\xi,\eta)$ of $(Rf)(t,\xi,\eta)$ in (3.5) with respect to t yields the integral in (3.1) up to the factor $(-ik)^{n-1}$, since the function $v(k,\xi,\eta)/(-ik)^{n-1}$ can be shown to satisfy the dispersion relation if $\hat{\phi}$ and $\tilde{\phi}$ are positive. Thus,

$$v(k,\xi,\eta) = (-ik)^{n-1} (Rf)\hat{}\,(k,\xi,\eta). \tag{3.6}$$

We consider the problem of finding the function $f(x)$ in (3.1) in the following two situations: (i) the position of the source is fixed, i.e., we are given $v(k,\xi,\eta)$ for fixed η and for a set of values $\xi \in \partial X$; and (ii) the position of the receiver is a function of the position of the source, i.e., we are given $v(k,\xi(\eta),\eta)$, where $\xi(\eta)$ is a known function of η, for a set of values $\eta \in \partial X$.

We specialize the arguments of Refs. 14 and 16 to the case of the integral equation in (3.1). Having established the relation of the integral equation in (3.1) to the GRT it becomes natural to introduce the same generalized backprojection operator and Fourier integral operator used in Refs. 14–16 to study the integral equation in (3.1). This we do in Sec. IV of this article.

We make the following assumptions about the domain X, its boundary ∂X, and indices of refraction in (3.3a) and (3.3b).

Let $n(x)$ be the index of refraction in the region X and let S_x^{n-1} be the unit sphere with the center at the origin of the tangent space at the interior point $x \in X$. Here, S_x^{n-1} represents all directions at the point $x \in X$. Let $\{\gamma(x,\xi)\}$ be a family of geodesics (rays) of the metric $n(x)dx$ connecting the point x with points $\xi \in \partial X^0$, where $\partial X^0 \subset \partial X$ is an open region of the boundary. Each ray within the family has a direction $\omega \in S_x^{n-1}$ at the point x. Thus the family of rays maps directions at the point x (an open domain of the unit sphere S_x^{n-1}) into ∂X^0. *In this article we always assume that this map is an orientation-preserving diffeomorphism.* Algebraically this means that certain Jacobians do not vanish. Physically it means that if a source located at an interior point of X illuminates a region ∂X^0 on the boundary, then this region can be smoothly contracted along the rays into a part of a small sphere around the source. We note that this assumption leads to the uniqueness and stability estimate in the inverse travel time problem.[20] When the index of refraction is constant this condition is satisfied for domains which are star shaped with respect to points of reconstruction. These include all practical configurations in geophysics, tomography, and nondestructive testing.

Our next remark deals with the domains of definition of the operators that appear in this article. We always define operators on functions which belong to the class $C_0^\infty(X)$ or $C^\infty(X)$. However, we can extend the domain of definition to the appropriate dual class of generalized functions by the standard procedure (see Appendix B). Thereby, we consider such an extension automatically performed each time we define an operator in this paper.

IV. ASYMPTOTIC SOLUTION OF THE LINEARIZED INVERSE PROBLEM WITH A FIXED INCIDENT FIELD

In this section we construct an asymptotic solution of the integral equation in (3.1) given the function $v(k,\xi,\eta)$, where η—the position of the source—is fixed. For the sake of brevity the dependence on η will sometimes be suppressed. Thus, we write the integral equation in (3.1) as

$$v(k,\xi) = (Wf)(k,\xi), \tag{4.1}$$

where

$$(Wf)(k,\xi) = (-ik)^{n-1}\int_X f(x)a(x,\xi)e^{ik\phi(x,\xi,\eta)}dx. \tag{4.2}$$

In (4.1) and (4.2), $v(k,\xi)$ and $a(x,\xi)$ stand for $v(k,\xi,\eta)$ and $a(x,\xi,\eta)$ in (3.1). The phase function $\phi(x,\xi,\eta)$ is described in (3.2).

We now introduce the generalized backprojection operator R^* dual to the generalized Radon transform R. For infinitely differentiable functions $u(t,\xi) \in C^\infty(R_+ \times \partial X)$ we define R^* as

$$(R^*u)(y) = \int_{\partial X} u(t,\xi)\Big|_{t=\phi(y,\xi,\eta)} b(y,\xi)d\xi. \tag{4.3}$$

The weight function $b(y,\xi)$ is a smooth, non-negative function on $X \times \partial X$, $b(y,\xi) \in C^\infty(X \times \partial X)$, which we have chosen to be

$$b(y,\xi) = [h(y,\xi)/a(y,\xi)]\,\chi(y,\xi), \tag{4.4}$$

where $h(y,\xi)$ is the determinant

$$h(y,\xi) = \begin{bmatrix} \phi_{y_1} & \phi_{y_2} & \cdots & \phi_{y_n} \\ \hat{\phi}_{y_1\xi_1} & \hat{\phi}_{y_2\xi_1} & \cdots & \hat{\phi}_{y_n\xi_1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\phi}_{y_1\xi_{n-1}} & \hat{\phi}_{y_2\xi_{n-1}} & \cdots & \hat{\phi}_{y_n\xi_{n-1}} \end{bmatrix}, \tag{4.5}$$

and $\chi(y,\xi)$ is a cutoff function. The cutoff function $\chi(y,\xi)$ is described below and is chosen to ensure that $\chi(y,\xi)h(y,\xi) \geqslant 0$ on $X \times \partial X$. The phase functions ϕ, $\hat{\phi}$, and $\tilde{\phi}$ are related by (3.2). The functions $\hat{\phi}$ and $\tilde{\phi}$ are solutions of the eikonal equations (3.3a) and (3.3b). Our particular choice (4.4) of the weight function $b(y,\xi)$ in (4.3) makes it possible to regularize the problem of inverting the GRT as was shown in Refs. 14 and 16.

We now change the variable of integration in (4.3). For each point y in the interior of X let ω denote a point on the unit sphere S_y^{n-1}. This means that ω is a direction at the point y. For the ray with direction ω at the interior point $y \in X$ let $\xi(\omega)$ be the point of intersection of that ray with the boundary ∂X. According to the assumptions formulated in the previous section, the function $\xi = \xi(\omega)$ is invertible and has continuous partial derivatives of the first order. We choose $\omega \in S_y^{n-1}$ to be the new variable of integration in (4.3). From Lemma A in Appendix A it follows that

$$h(y,\xi)d\xi = \hat{n}^n(1 + (\tilde{n}/\hat{n})\cos\psi)d\omega, \tag{4.6}$$

where

$$\cos\psi(y,\xi) = [\nabla_y\hat{\phi}(y,\xi)\cdot\nabla_y\tilde{\phi}(y,\eta)]/\hat{n}(y)\tilde{n}(y), \tag{4.7}$$

and $d\omega$ is the standard solid angle differential form on the unit sphere S_y^{n-1}.

Substituting (4.6) in (4.3) we rewrite R^* as

$$(R^*u)(y) = \hat{n}^n(y)\int_{S_y^{n-1}} u(t,\xi)\Big|_{t=\phi(y,\xi,\eta)} \frac{\chi(y,\xi)}{a(y,\xi)} \times\left[1 + \frac{\tilde{n}(y)}{\hat{n}(y)}\cos\psi(y,\xi)\right]d\omega, \tag{4.8}$$

where $\xi = \xi(\omega)$. In this form the operator R^* can be computed explicitly by making use of ray tracing.

It remains to define the cutoff function $\chi(y,\xi)$. Given the interior point $y \in X$ and the boundary point $\xi \in \partial X$ we set $\chi(y,\xi) = 0$, if $1 + [\tilde{n}(y)/\hat{n}(y)]\cos\psi(y,\xi) \leqslant 0$. Choosing an arbitrarily small $\epsilon > 0$ we set $\chi(y,\xi) = 1$, if $1 + [\tilde{n}(y)/\hat{n}(y)]\cos\psi(y,\xi) \geqslant \epsilon$, and define $\chi(y,\xi)$ elsewhere so that it is infinitely differentiable and $0 \leqslant \chi(y,\xi) \leqslant 1$.

Let $\partial X_\eta(y)$ be the region of the boundary defined by

$$\partial X_\eta(y) = \{\xi \in \partial X: \cos\psi(y,\xi) < -[\hat{n}(y)/\tilde{n}(y)] + \epsilon\}.$$

Let $\partial X_\eta^0(y)$ be the complement of $\partial X_\eta(y)$, i.e.,

$$\partial X_\eta^0(y) = \partial X \setminus \partial X_\eta(y).$$

To describe $\partial X_\eta(y)$ and, therefore, $\partial X_\eta^0(y)$ for a given interior point $y \in X$, three cases can be distinguished.

(1) $\hat{n}(y) > \tilde{n}(y)$. In this case $\partial X_\eta(y)$ is empty since ϵ can be chosen sufficiently small.

(2) $\hat{n}(y) = \tilde{n}(y)$. In this case for sufficiently small ϵ the region $\partial X_\eta(y)$ consists of an ϵ-neighborhood of the unique point ξ_0, where $\cos \psi(y,\xi_0) = -1$.

(3) $\hat{n}(y) < \tilde{n}(y)$. In this case for any sufficiently small ϵ the region $\partial X_\eta(y)$ is a connected part of the boundary ∂X. The cutoff function $\chi(y,\xi) \equiv 1$ on $\partial X_\eta^0(y)$. If (1) holds the cutoff function $\chi(y,\xi) \equiv 1$ on all of ∂X. If (2) holds the cutoff function isolates a single point ξ_0 and is introduced here for technical reasons. In carrying out the integration in (4.8) one can set the cutoff function $\chi(y,\xi) \equiv 1$ on the boundary ∂X. If (3) holds the cutoff function $\chi(y,\xi)$ is zero on all but a small open subset of $\partial X_\eta(y)$ which is determined by the choice of ϵ. In carrying out the integration in (4.8) we can set $\epsilon = 0$ and $\chi(y,\xi) = 0$ on $\partial X_\eta(y)$.

If $v(k,\xi)$ is available only on a part of the boundary ∂X we have to modify the cutoff function. We include an ϵ-neighborhood of the region where the function $v(k,\xi)$ is not known in the set $\partial X_\eta(y)$. Then $\chi(y,\xi)$ is again defined in such a way that it is infinitely differentiable with values $0 \leqslant \chi(y,\xi) \leqslant 1$ with $\chi(y,\xi) \equiv 1$ on $\partial X_\eta^0(y) = \partial X \setminus \partial X_\eta(y)$ and $\chi(y,\xi) = 0$ on all but an arbitrary small subset within $\partial X_\eta(y)$. We denote the modified cutoff function again by $\chi(y,\xi)$.

Note, that if (3) holds—as in the case when the incoming wave is s–p converted at the point $y \in X$ in an elastic medium—the angle ψ_0, such that $\cos \psi_0 = -\hat{n}(y)/\tilde{n}(y)$ plays the role of a "critical angle" between the directions of incoming and outgoing waves at the point y.

We now consider the FIO defined by

$$(Ff)(y) = \frac{1}{(2\pi)^n} \int_0^\infty \int_{\partial X} \int_X e^{ik\Phi(x,y,\xi)} A(x,y,\xi) \times f(x) dx\, d\xi\, k^{n-1}\, dk, \tag{4.9}$$

where

$$\Phi(x,y,\xi) = \phi(x,\xi,\eta) - \phi(y,\xi,\eta). \tag{4.10}$$

The function $\phi(x,\xi,\eta)$ is described in (3.2), and

$$A(x,y,\xi) = [a(x,\xi)/a(y,\xi)] h(y,\xi) \chi(y,\xi), \tag{4.11}$$

where $a(x,\xi)$ is the weight function in (4.2). Let us also introduce the operator $\mathscr{F}^+$:

$$(\mathscr{F}^+ v)(t) = \frac{i^{n-1}}{(2\pi)^n} \int_0^\infty v(k) e^{-ikt}\, dk. \tag{4.12}$$

We have

$$F = R^* \mathscr{F}^+ W. \tag{4.13}$$

To investigate the operator F we observe that the first term in the Taylor series for $\Phi(x,y,\xi)$ is $\nabla_y \phi(y,\xi,\eta)\cdot(x-y)$ and consider the operator

$$(I_{\partial X_\eta^0} f)(y) = \frac{1}{(2\pi)^n} \int_0^\infty \int_{\partial X_\eta^0(y)} \int_X e^{ik\nabla_y\phi(y,\xi,\eta)\cdot(x-y)} \times h(y,\xi) f(x) dx\, d\xi\, k^{n-1}\, dk. \tag{4.14}$$

Changing variables of integration from k,ξ to p, where

$$p = k\nabla_y \phi(y,\xi,\eta), \tag{4.15}$$

we find that

$$dp = k^{n-1} h(y,\xi) d\xi\, dk, \tag{4.16}$$

and, thereby,

$$(I_{\partial X_\eta^0} f)(y) = \frac{1}{(2\pi)^n} \int_{\Omega_\eta(y)} \int_X e^{ip\cdot(x-y)} f(x) dx\, dp, \tag{4.17}$$

where $\Omega_\eta(y)$ is the image of $R_+ \times \partial X_\eta^0(y)$ under the change of variables in (4.15). It follows from (4.17) that

$$(I_{\partial X_\eta^0} f)(y) = \frac{1}{(2\pi)^n} \int_{\Omega_\eta(y)} e^{-ip\cdot y} f\hat{}(p) dp, \tag{4.18}$$

where $f\hat{}(p)$ is the Fourier transform of the function f. It is clear that if $\partial X_\eta^0(y) = \partial X$ then the operator $I_{\partial X_\eta^0}$ is the identity operator. If $\partial X_\eta^0(y) \neq \partial X$, then the operator $I_{\partial X_\eta^0}$ in (4.18) is *an operator of partial reconstruction*.

It is important to note that for each point $y \in X$, the region $\partial X_\eta^0(y)$ on the boundary ∂X can be explicitly constructed by ray tracing. Having found $\partial X_\eta^0(y)$, we can then determine explicitly the set $\Omega_\eta(y)$ in the Fourier domain, where the Fourier transform $f\hat{}(p)$ of the function f, which we would like to recover, is known. If only partial data are available this set $\Omega_\eta(y)$ in the Fourier domain determines the spatial resolution of the partial reconstruction (4.18) and controls what can be recovered in the migration schemes presented below.

The asymptotic solution of the integral equation in (4.1) is constructed by making use of the following theorem.

Theorem 1: The Fourier integral operator F in (4.9) is a pseudodifferential operator and can be represented as a sum

$$F = I_{\partial X_\eta^0} + T_1 + T_2 + \cdots, \tag{4.19}$$

where $I_{\partial X_\eta^0}$ denotes the operator described in (4.17) and the operators $T_1, T_2, \ldots$ belong to increasingly smooth classes of pseudodifferential operators.

The definition of classes of pseudodifferential operators can be found in Appendix B. For further references see Ref. 21, or any other reference where Fourier integral operators and pseudodifferential operators are studied.

It follows from (4.13) and Theorem 1 that by making use only of the first term in (4.19),

$$R^* \mathscr{F}^+ W \approx I_{\partial X_\eta^0}, \tag{4.20}$$

we obtain an approximate reconstruction algorithm. The expansion in (4.19) also explains the precise meaning of the approximation in (4.20). Since we neglect all terms in the expansion which appear to be smoothing operators, the approximation in (4.20) reconstructs only (the location of) the discontinuities of the function f (or the places, where the gradient of f is large). In this sense the formula in (4.20) provides an algorithm for imaging the discontinuities. This is, of course, what is sought in geophysics and many other applications where the discontinuities of parameters describing the medium are of interest. In Sec. VI we describe the algorithm contained in (4.20) in greater detail.

The remainder of this section contains an outline of the proof of Theorem 1. The material presented in Secs. V and VI is independent of the details of the proof.

The proof follows along the same lines as the arguments presented in Refs. 14 and 16. Consider the set

$$C_\Phi = \{(k,\xi,x,y) \in R_+ \times \partial X^0_\eta(y) \times X \times X : \Phi(x,y,\xi) = 0,$$
$$\nabla_\xi \Phi(x,y,\xi) = 0\}. \tag{4.21}$$

This set is of fundamental importance in the theory of Fourier integral operators since its structure determines the properties of the operator. The definition (4.21) is not standard (see Ref. 21, for example), however the change of variables (4.15) transforms it into the classical one.

Using the assumption that the function $\xi = \xi(\omega)$ is a diffeomorphism it can be shown that

$$C_\Phi = \{(k,\xi,x,x) : k \in R_+,\ \xi \in \partial X^0_\eta(y),\ x \in X\}, \tag{4.22}$$

so that the projection of C_Φ on $X \times X$ is the diagonal. This implies that the operator in (4.9) is a pseudodifferential operator as defined in Appendix B.

Let us consider $\chi_\delta(x,y) \in C^\infty(X \times X)$, $0 \leqslant \chi_\delta(x,y) \leqslant 1$, such that

$$\chi_\delta(x,y) = 1, \text{ if } |x-y| < \delta/2,$$
$$\chi_\delta(x,y) = 0, \text{ if } |x-y| > \delta,$$

where $\delta > 0$ is an arbitrary small parameter. Instead of the FIO (4.9) we can study the following operator (we keep the same notation):

$$(Ff)(y) = \frac{1}{(2\pi)^n} \int_0^\infty \int_{\partial X} \int_X e^{ik\Phi(x,y,\xi)} A(x,y,\xi)$$
$$\times \chi_\delta(x,y) f(x) dx\, d\xi\, k^{n-1}\, dk. \tag{4.23}$$

This operator differs from the operator in (4.9) by a regularizing operator (see Appendix B). The regularizing operator does not change the asymptotics and can be neglected since it is "infinitely smooth."

If δ is sufficiently small and $|x-y| < \delta$ we can write the phase function $\Phi(x,y,\xi)$ as

$$\Phi(x,y,\xi) = \nabla_y \phi(y,\xi,\eta)\cdot(x-y) + H(x,y,\xi), \tag{4.24}$$

where $H(x,y,\xi) = O(|x-y|^2)$, and the amplitude

$$A(x,y,\xi) = h(y,\xi)\chi(y,\xi) + \tilde{A}(x,y,\xi), \tag{4.25}$$

where $\tilde{A}(x,y,\xi) = O(|x-y|)$. Making the change of variables (4.15) and using (4.16), (4.23) becomes

$$(Ff)(y) = \frac{1}{(2\pi)^n} \int_{\Omega_\eta(y)} \int_X e^{ip\cdot(x-y) + iH(x,y,p)}$$
$$\times (1 + \tilde{A}(x,y,p)) f(x) dx\, dp, \tag{4.26}$$

where $H(x,y,p) = k(p) H(x,y,\xi(p))$ and $\tilde{A}(x,y,p) = \tilde{A}(x,y,\xi(p))$. The functions $k(p)$ and $\xi(p)$ can be determined from the change of variables in (4.15). Note, that as functions of p, $H(x,y,p)$ and $\tilde{A}(x,y,p)$ are homogeneous of degree 1 and 0, respectively. Consider now the operator

$$(F_s f)(y) = \frac{1}{(2\pi)^n} \int_{\Omega_\eta(y)} \int_X e^{ip\cdot(x-y) + isH(x,y,p)}$$
$$\times (1 + s\tilde{A}(x,y,p)) f(x) dx\, dp. \tag{4.27}$$

Since $F = F_1$ we can use the Taylor expansion of $F_s f$ as a function of s to express F as

$$F = \sum_{m=0}^{N} \frac{1}{m!} \left(\frac{d}{ds}\right)^m F_s \bigg|_{s=0} + \int_0^1 \frac{(1-s)^N}{N!} \left(\frac{d}{ds}\right)^{N+1} F_s\, ds. \tag{4.28}$$

It was shown[14,16] that the expansion in (4.28) of the operator F of the form in (4.27) consists of increasingly smooth pseudodifferential operators. Comparing the first term in the expansion (4.28) with (4.17) we obtain the expansion in (4.19). One can compute and use more terms in the expansion (4.19). For example, the operator T_1 was computed in Refs. 14–16.

V. ASYMPTOTIC SOLUTION OF THE LINEARIZED INVERSE PROBLEM WHEN THE RECEIVER POSITIONS DEPEND ON THE SOURCE POSITIONS

In this section we construct an asymptotic solution of the integral equation in (3.1) given the function $v(k,\xi,\eta)$, where the receiver positions $\xi = \xi(\eta)$ depend on the source positions η. The arguments in this case are analogous to those for the fixed source position, and are presented briefly for this reason. Again, consider the integral equation in (3.1), which we now write as

$$v(k,\eta) = (Wf)(k,\xi(\eta),\eta), \tag{5.1}$$

where

$$(Wf)(k,\xi(\eta),\eta) = (-ik)^{n-1} \int_X f(x) a(x,\xi(\eta),\eta)$$
$$\times e^{ik\phi(x,\xi(\eta),\eta)}\, dx. \tag{5.2}$$

The phase function ϕ is described in (3.2).

For functions $u(t,\eta) \in C^\infty(R_+ \times \partial X)$ we define the dual transform R^* as

$$(R^*u)(y) = \int_{\partial X} u(t,\eta)\Big|_{t=\phi(y,\xi(\eta),\eta)} b(y,\eta) d\eta, \tag{5.3}$$

where the weight function $b(y,\eta)$ is a smooth, non-negative function on $X \times \partial X$, $b(y,\eta) \in C^\infty(X \times \partial X)$ and is chosen to be

$$b(y,\eta) = [h(y,\eta)/a(y,\xi(\eta),\eta)]\chi(y,\eta). \tag{5.4}$$

Here, $h(y,\eta)$ is the determinant

$$h(y,\eta) = \begin{bmatrix} \phi_{y_1} & \phi_{y_2} & \cdots & \phi_{y_n} \\ \phi_{y_1\eta_1} & \phi_{y_2\eta_1} & \cdots & \phi_{y_n\eta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{y_1\eta_{n-1}} & \phi_{y_2\eta_{n-1}} & \cdots & \phi_{y_n\eta_{n-1}} \end{bmatrix}, \tag{5.5}$$

and $\chi(y,\eta)$ is a cutoff function described below. Note, that the function $h(y,\eta)$ differs from the one in (4.5) in the previous section. To compute the determinant (5.5) we again use Lemma A in Appendix A. In the two-dimensional case we find that

$$h(y,\eta) d\eta = \left(\hat{n}^2 \frac{d\xi}{d\eta} + \tilde{n}^2 + \hat{n}\tilde{n}\left(1 + \frac{d\xi}{d\eta}\right)\cos\psi\right) d\omega, \tag{5.6}$$

where

$$\cos\psi(y,\eta) = [\nabla_y \hat{\phi}(y,\xi(\eta))\cdot\nabla_y \tilde{\phi}(y,\eta)]/\hat{n}(y)\tilde{n}(y), \tag{5.7}$$

and $d\omega$ is the standard angle measure on the unit circle.

For simplicity let us assume that the function $\xi(\eta)$ is such that the function $h(y,\eta)$ in (5.5) is strictly positive for all $y \in X$ and $\eta \in \partial X$. In this case the infinitely differentiable cutoff function—$\chi(y,\eta)$—is introduced only to isolate a region of the boundary $\partial\tilde{X}(y)$, where we do not know the function $v(k,\eta)$. The cutoff function $\chi(y,\eta) = 0$ for points η in $\partial\tilde{X}(y)$; it has values $0 \leqslant \chi(y,\xi) \leqslant 1$ in an ϵ-neighborhood of $\partial\tilde{X}(y)$, where ϵ is arbitrarily small, and it is set equal to 1 elsewhere. Let $\partial X^0(y)$ denote the complement of $\partial\tilde{X}(y)$, i.e.,

$$\partial X^0(y) = \partial X \setminus \partial\tilde{X}(y).$$

Again, consider the Fourier integral operator F

$$(Ff)(y) = \frac{1}{(2\pi)^n} \int_0^\infty \int_{\partial X} \int_X e^{ik\Phi(x,y,\eta)} \times A(x,y,\eta) f(x) dx\, d\eta\, k^{n-1}\, dk, \quad (5.8)$$

where Φ is defined in terms of the function ϕ in (3.2) as

$$\Phi(x,y,\eta) = \phi(x,\xi(\eta),\eta) - \phi(y,\xi(\eta),\eta), \quad (5.9)$$

and

$$A(x,y,\eta) = \frac{a(x,\xi(\eta),\eta)}{a(y,\xi(\eta),\eta)} h(y,\eta)\chi(y,\eta). \quad (5.10)$$

Using the operator $\mathscr{F}^+$ defined in (4.12), we have

$$F = R^* \mathscr{F}^+ W. \quad (5.11)$$

The analysis of the operator F is conducted analogously to the one in the previous section. In this case the change of the variables of integration in (5.8) from k,η to p is as follows:

$$p = k\nabla_y \phi(y,\xi(\eta),\eta), \quad (5.12)$$

and

$$dp = k^{n-1} h(y,\eta) d\eta\, dk. \quad (5.13)$$

The asymptotic solution of the integral equation in (5.1) is constructed by making use of the following theorem.

Theorem 2: The Fourier integral operator in (5.8) is a pseudodifferential operator and can be represented as a sum

$$F = I_{\partial X^0} + T_1 + T_2 + \cdots, \quad (5.14)$$

where $I_{\partial X^0}$ denotes the operator

$$(I_{\partial X^0} f)(y) = \frac{1}{(2\pi)^n} \int_{\Omega(y)} e^{-ip\cdot y} f\hat{\ }(p) dp, \quad (5.15)$$

where $\Omega(y)$ is the image of $R_+ \times \partial X^0(y)$ under the change of variables in (5.12). The operators $T_1, T_2, \ldots$ belong to increasingly smooth class of pseudodifferential operators (see Appendix B).

The first term of the expansion in (5.14) yields the approximate reconstruction algorithm

$$R^* \mathscr{F}^+ W \approx I_{\partial X^0}, \quad (5.16)$$

where the generalized backprojection operator R^* is given by (5.3).

In the following section we show that for constant background and coincident sources and receivers the approximation (5.16) reduces to algorithms described in the literature.

VI. THE ASYMPTOTIC SOLUTIONS AND MIGRATION SCHEMES

This section contains a brief description of migration schemes which follow from our results. As we shall see, the measured scattered data are such that the migration schemes amount to the generalized backprojections (except when the Hilbert transform has to be applied first in spaces of even dimensions).

Let us recall that the goal is to estimate the unknown function $f(x)$ in (2.12) or (3.1) from observations of the (singly) scattered field. We assume that the scattered field $u = u^{sc}$ is given in the time domain, so that

$$u(t,\xi,\eta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} v(k,\xi,\eta) e^{-ikt}\, dk, \quad (6.1)$$

where $v(k,\xi,\eta)$ is described in (3.1). In many cases of practical interest the actual measurements are measurements of the total field, and the incident field must, by some means, be removed. However, in this paper we assume that the (singly) scattered field is given in the time domain to start with.

Comparing the definition of the operator $\mathscr{F}^+$ in (4.12) with the Fourier transform in (6.1) and denoting the real part of the operator (4.12) as $w(t,\xi,\eta) \equiv \mathrm{Re}(\mathscr{F}^+ v)(t,\xi,\eta)$ we obtain

$$w(t,\xi,\eta) = [(-1)^{(n-1)/2}/2(2\pi)^{n-1}]\, u(t,\xi,\eta), \quad (6.2)$$

in spaces of odd dimensions $n = 2m+1$, $m = 1,2,\ldots$, and

$$w(t,\xi,\eta) = \frac{(-1)^{n/2}}{2(2\pi)^{n-1}} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u(t',\xi,\eta)}{t-t'} dt', \quad (6.3)$$

in spaces of even dimensions $n = 2m$, $m = 1,2,\ldots$. Thus, one has to apply the Hilbert transform (6.3) to the scattered field in spaces of even dimensions to obtain $\mathrm{Re}(\mathscr{F}^+ v)(t,\xi,\eta)$.

It follows from (6.2) and (6.3) that the only remaining step in algorithms (4.20) and (5.16) is to construct the generalized backprojection operator (GBO) in (4.3) or (5.3), depending on the source-receiver configuration. Let us consider the case when the receiver position ξ depends on the source position η. The construction of the GBO for fixed source position in (4.3) is completely analogous.

To compute the GBO (5.3) we have to compute both the phase function and the weight function. Such a computation is equivalent to the construction of two Green's functions in ray approximation. Indeed, the computation of the functions $\phi^{\mathrm{in}} = \tilde{\phi}$ and $\phi^{\mathrm{out}} = \hat{\phi}$ in (3.2) and the factors A^{in} and A^{out} in (3.4)—which are necessary to construct the weight function in (5.3) [or (4.3)]—amounts to the computation of the ray approximation of two Green's functions along the two rays connecting the point of interest in the medium with the source and with the receiver. The additional obliquity factor in the weight function can be easily computed as it follows from (5.6) and (5.7) [or (4.6) and (4.7)]. This factor depends on the angle between the rays connecting the point of interest in the medium with the source and receiver.

Once both the phase function and the weight function are computed the GBO is applied in the time domain, either to the singly scattered field itself [in spaces of odd dimensions (6.2)] or to the Hilbert transform of the singly scattered field [in spaces of even dimensions (6.3)], as required by the approximate formulas in (4.20) or (5.16). In this way we obtain the reconstruction f_{mig} of the function f from (6.2), (6.3) and (5.16) as

$$f_{\mathrm{mig}} \approx \mathrm{Re}\, I_{\partial X^0} f. \quad (6.4)$$

If $\partial X^0 = \partial X$ then the operator $I_{\partial X^0}$ is the identity operator. The function f in (2.12) [or (3.1)] is assumed to be real, and, therefore, it follows from (6.4) that $f_{\mathrm{mig}}(x) \approx f(x)$ in the region X. The symbol $\approx$ expresses the fact that we *image the (location of) discontinuities* of the function $f(x)$ in the region X, since smooth terms in the asymptotic expansion in (5.14) are neglected in the approximation.

In most practical situations we have data only for limited view angles, and, therefore, $\partial X^0 \neq \partial X$. Thus, we obtain a partial reconstruction since we can recover the Fourier transform $f\hat{\ }$ [see (5.15)] only on a part of the Fourier space. The assumption that f is real implies that $\overline{f\hat{\ }(p)} = f\hat{\ }(-p)$, where the bar denotes the complex conjugate. In particular, this relation shows that the Fourier space is covered twice if $\partial X^0 = \partial X$. Given the source-receiver configuration of a particular experiment, we can determine the domain in the Fourier space where the function $f^\wedge$ is known. This domain controls the spatial resolution of the reconstruction. In examples where the domain X is a half-space (given later in this section) we only have partial coverage since observation points are restricted to the boundary of the half-space. However, assuming infinite aperture we still obtain the function $f^\wedge$ over the whole Fourier space by continuing $f^\wedge$ with the help of the identity $\overline{f^\wedge}(p) = f^\wedge(-p)$.

We shall discuss the implications of our results in exploration geophysics and the comparison with existing migration schemes in greater detail elsewhere. Here we note only, that, in general, the construction of a GBO requires ray tracing and computation of solutions of the transport equation. However, in the case of constant background one can obtain analytical expressions for the GBO. Let us illustrate this with a few examples where an explicit construction of the GBO is available, and show that at least for these examples some of the migration schemes appearing in the literature are given by a generalized backprojection operator. We consider the case with a constant index of refraction and set

$$\hat{n} = \tilde{n} = 1,$$

without loss of generality.

Example 1: Let the domain X be the half-space $x_n \geqslant 0$ of the n-dimensional space $(x_1, x_2, \ldots, x_{n-1}, x_n)$ and suppose measurements are performed everywhere on the boundary $\partial X = \{(x_1, x_2, \ldots, x_{n-1}, 0)\}$ of the half-space X. Let $\xi = \eta = (\eta_1, \eta_2, \ldots, \eta_{n-1}, 0)$, so that we have coincident sources and receivers. The phase functions $\hat{\phi}$ and $\tilde{\phi}$ are

$$\hat{\phi} = \tilde{\phi} = |x - \eta|,$$

so that ϕ in (3.2) is

$$\phi(x,\eta) = 2|x - \eta|. \tag{6.5}$$

To compute the determinant h in (5.5) we make use of the identities

$$\phi_{x_i \eta_j} = -(4/\phi)\delta_{ij} - \phi_{x_i}\phi_{\eta_j}/\phi,$$

$$\phi_{x_n \eta_j} = -\phi_{x_n}\phi_{\eta_j}/\phi,$$

where δ_{ij} is the Kronecker symbol and $i,j = 1,2,\ldots,n-1$, and obtain

$$h(x,\eta) = 4^{n-1}\phi_{x_n}/\phi^{n-1}, \tag{6.6}$$

where

$$\phi_{x_n}(x,\eta) = 4x_n/\phi(x,\eta). \tag{6.7}$$

It follows from (2.4), (3.1), and (3.4) that

$$a(x,\eta,\eta) = 1/4(\pi\phi)^{n-1}, \tag{6.8}$$

where ϕ is given by (6.5). Using (6.6)–(6.8) the weight function b in (5.4) can be written as

$$b(x,\eta) = C_n(x_n/|x - \eta|),$$

where

$$C_n = 2^{2n+1}\pi^{n-1}.$$

We set

$$x_n/|x - \eta| = \bar{l}\cdot\bar{l}_0,$$

where $\bar{l}$ and $\bar{l}_0$ are unit vectors pointing in the direction of x_n axis and in the direction of the line connecting points x and η. The generalized backprojection operator R^* thus is

$$(R^*w)(x) = C_n \int_{\partial X} w(2|x - \eta|, \eta)\bar{l}\cdot\bar{l}_0\, d\eta, \tag{6.9}$$

where w is given in (6.2) or (6.3) depending on the dimension n. Here we integrate over all source-receiver positions on the boundary ∂X. The GBO in (6.9) can be considered as a migration scheme within our definition. This operator was obtained in Ref. 17 for the case $n = 3$ by a different approach.

Some of the migration schemes that have appeared in the literature also have a form of the GBO. Reference 3 is an example. However, the particular weight functions used are generally different from the one presented here.

Example 2: This example deals with the case where the source and receiver positions are confined to a sphere—the surface of the n-dimensional ball of the radius ρ. We can write $\xi = \eta = \rho\nu$, where ν is a unit vector indicating the position of the coincident source and receiver on the sphere. Thus, we have

$$\hat{\phi} = \tilde{\phi} = |x - \rho\nu|,$$

and

$$\phi(x,\nu) = 2|x - \rho\nu|. \tag{6.10}$$

The computation of the determinant in (5.5) yields

$$h(x,\nu) = \rho^{n-1}2^{2n}(\rho - x\cdot\nu)/\phi^n.$$

Since the weight function $a(x,\xi(\eta),\eta)$ in (5.2) can be written for this example as

$$a(x,\nu,\nu) = 1/4(\pi\phi)^{n-1},$$

we obtain the weight function

$$b(x,\nu) = C_n[\rho^{n-1}(\rho - x\cdot\nu)/|x - \rho\nu|],$$

and the GBO

$$(R^*w)(x) = C_n \int_{|\nu|=1} w(2|x - \rho\nu|, \nu) \times \frac{\rho^{n-1}(\rho - x\cdot\nu)}{|x - \rho\nu|}\, d\nu, \tag{6.11}$$

where w is given in (6.2) or (6.3) depending on the dimension n. The integration in (6.11) is over the unit sphere and $d\nu$ is the standard solid angle differential form. The GBO in (6.11) differs from the one constructed in Ref. 17 for the case $n = 3$. The approximation in (6.11) remains valid even if the point x is not close to the center of the ball.

Example 3: Finally, we consider the case where source and receiver positions are confined to the boundary of a half-space as in our first example. However, we assume now that sources and receivers are separated by a fixed distance $2d$.

Let η denote the coordinate of the midpoint between the source and receiver, so that we can write

$$\hat{\phi} = |x - \eta - d|, \tag{6.12}$$

$$\tilde{\phi} = |x - \eta + d|, \tag{6.13}$$

and

$$\phi(x,\eta) = |x - \eta - d| + |x - \eta + d|.$$

We consider the case of the dimension $n = 2$. Making use of the identities

$$\phi_{x_1} = \frac{x_1-\eta-d}{\hat{\phi}} + \frac{x_1-\eta+d}{\tilde{\phi}},$$

$$\phi_{x_2} = x_2\Big(\frac{1}{\hat{\phi}} + \frac{1}{\tilde{\phi}}\Big),$$

$$\phi_{x_1\eta} = -\Big(\frac{1}{\hat{\phi}} + \frac{1}{\tilde{\phi}}\Big) - \Big(\frac{\hat{\phi}_{x_1}\hat{\phi}_\eta}{\hat{\phi}} + \frac{\tilde{\phi}_{x_1}\tilde{\phi}_\eta}{\tilde{\phi}}\Big),$$

$$\phi_{x_2\eta} = -\Big(\frac{\hat{\phi}_{x_2}\hat{\phi}_\eta}{\hat{\phi}} + \frac{\tilde{\phi}_{x_2}\tilde{\phi}_\eta}{\tilde{\phi}}\Big),$$

we can, in this case, compute the determinant in (5.5) directly and obtain

$$h(x,\eta) = (\hat{\phi}_{x_2} + \tilde{\phi}_{x_2})\Big(\frac{1}{\hat{\phi}} + \frac{1}{\tilde{\phi}}\Big) + (\hat{\phi}_{x_2}\tilde{\phi}_{x_1} - \hat{\phi}_{x_1}\tilde{\phi}_{x_2})\Big(\frac{\tilde{\phi}_\eta}{\tilde{\phi}} - \frac{\hat{\phi}_\eta}{\hat{\phi}}\Big).$$

Substituting appropriate expressions for the derivatives of the phase functions we write h as

$$h(x,\eta) = \frac{x_2}{\hat{\phi}^2\tilde{\phi}^2}\Big[(\hat{\phi} + \tilde{\phi})^2 - \frac{2d}{\hat{\phi}\tilde{\phi}}(d^2 + x_2^2 - 2(x-\eta)^2 d)\Big].$$

Hence,

$$b(x,\eta) = \frac{8\pi x_2}{(\hat{\phi}\tilde{\phi})^{3/2}}\Big[(\hat{\phi} + \tilde{\phi})^2 - \frac{2d}{\hat{\phi}\tilde{\phi}}(d^2 + x_2^2 - 2(x-\eta)^2 d)\Big],$$

where $\hat{\phi}$ and $\tilde{\phi}$ are given in (6.12) and (6.13). The GBO in this case is

$$(R*w)(x) = \int_{-\infty}^{+\infty} w(2|x-\eta|,\eta)b(x,\eta)d\eta, \tag{6.14}$$

where w is given in (6.3) for $n = 2$. It is easy to see that if $d = 0$ then we obtain the GBO in (6.9).

ACKNOWLEDGMENTS

I would like to express my gratitude to Douglas Miller, numerous discussions with whom accelerated the appearance of this paper and contributed to the geometrical interpretation of results. Also, I would like to thank Michael Oristaglio for his comments and Marion Orton for making many helpful suggestions to improve the presentation and to bring it into its final form.

APPENDIX A

The following lemma holds in the cases of Riemannian and Finsler spaces. It was used in Ref. 20 to prove a uniqueness theorem of the inverse travel time problem in the case of the Riemannian metric. We present here an elementary proof for the Euclidean space.

Lemma A: Let the function $\hat{\phi}$ satisfy the eikonal equation

$$(\nabla_x\hat{\phi}(x,\xi))^2 = n^2(x), \tag{A1}$$

in the domain X with the boundary ∂X, where the parameter $\xi \in \partial X$. We assume that the boundary ∂X is diffeomorphic (with the preservation of orientation) to the unit sphere S_x^{n-1} centered at the origin of the tangent space at the point $x \in X$. (This unit sphere represents all directions at the point x.)

Let $\tilde{\phi}(x)$ have first partial derivatives and consider the determinant

$$J(x,\xi) = \begin{bmatrix} \tilde{\phi}_{x_1} & \tilde{\phi}_{x_2} & \cdots & \tilde{\phi}_{x_n} \\ \hat{\phi}_{x_1\xi_1} & \hat{\phi}_{x_2\xi_1} & \cdots & \hat{\phi}_{x_n\xi_1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\phi}_{x_1\xi_{n-1}} & \hat{\phi}_{x_2\xi_{n-1}} & \cdots & \hat{\phi}_{x_n\xi_{n-1}} \end{bmatrix}, \tag{A2}$$

where $\xi_1,\dots,\xi_{n-1}$ is a local system of coordinates on the boundary. Then

$$J(x,\xi)d\xi_1\cdots d\xi_{n-1} = n^{n-2}(\nabla_x\tilde{\phi}\cdot\nabla_x\hat{\phi})d\omega, \tag{A3}$$

where $d\omega$ is the standard measure on the sphere S_x^{n-1}, and $\xi = \xi(\omega)$, where $\omega \in S_x^{n-1}$.

Proof: Since $\hat{\phi}$ satisfies the eikonal equation in (A1) we can write

$$\begin{cases} \hat{\phi}_{x_1} = n\cos\chi_1 \\ \hat{\phi}_{x_2} = n\sin\chi_1\cos\chi_2 \\ \vdots \\ \hat{\phi}_{x_{n-1}} = n\sin\chi_1\sin\chi_2\cdots\cos\chi_{n-1} \\ \hat{\phi}_{x_n} = n\sin\chi_1\sin\chi_2\cdots\sin\chi_{n-1}, \end{cases} \tag{A4}$$

where $\chi_j = \chi_j(x,\xi_1,\dots,\xi_{n-1}), j = 1,\dots,n-1$, are angular coordinates on the unit sphere S_x^{n-1}. By substituting (A4) in the functional Jacobian in (A2) we obtain (A3). Let us carry out the calculation for the case $n = 3$. We have

$$\begin{aligned} \hat{\phi}_{x_1\xi_j} &= -n\sin\chi_1\frac{\partial\chi_1}{\partial\xi_j}, \\ \hat{\phi}_{x_2\xi_j} &= n\cos\chi_1\cos\chi_2\frac{\partial\chi_1}{\partial\xi_j} - n\sin\chi_1\sin\chi_2\frac{\partial\chi_2}{\partial\xi_j}, \\ \hat{\phi}_{x_3\xi_j} &= n\cos\chi_1\sin\chi_2\frac{\partial\chi_1}{\partial\xi_j} + n\sin\chi_1\cos\chi_2\frac{\partial\chi_2}{\partial\xi_j}, \end{aligned} \tag{A5}$$

where $j = 1,2$.

Let us compute cofactors of the first row of the functional determinant. Using (A5) we compute the cofactor for the element $\tilde{\phi}_{x_1}$. We obtain

$$\begin{aligned} \begin{vmatrix} \hat{\phi}_{x_2\xi_1} & \hat{\phi}_{x_3\xi_1} \\ \hat{\phi}_{x_2\xi_2} & \hat{\phi}_{x_3\xi_2} \end{vmatrix} &= n^2\Big[\frac{\partial\chi_1}{\partial\xi_1}\frac{\partial\chi_2}{\partial\xi_2}\sin\chi_1\cos\chi_1 - \frac{\partial\chi_1}{\partial\xi_2}\frac{\partial\chi_2}{\partial\xi_1}\sin\chi_1\cos\chi_1\Big] \\ &= n\frac{\partial(\chi_1,\chi_2)}{\partial(\xi_1,\xi_2)}\hat{\phi}_{x_1}\sin\chi_1. \end{aligned}$$

Analogous calculations can be performed for all cofactors, so that we obtain (A3), where

$$d\omega = \sin\chi_1\,d\chi_1\,d\chi_2.$$

APPENDIX B

We briefly present here some first definitions and properties of pseudodifferential operators. Consider the operator $a(x,D)$,

$$(a(x,D)f)(x) = \int_{R^n} a(x,p)f\hat{~}(p)e^{ip\cdot x}\,dp,$$

where $f\hat{}(p)$ denotes the Fourier transform of the function f. The function $a(x,p)$ is the symbol of the pseudodifferential operator $a(x,D)$.

Definition: Let Ω be an open subset of R^n and m be a real number. Let $S^m(\Omega)$ be the class of symbols and consist of infinitely differentiable functions $a(x,p)$, $a(x,p)\in C^\infty(\Omega\times R^n)$, such that to every compact $Q\subset\Omega$ and to every two multi-indices α,β there is a constant $C_Q(\alpha,\beta)$, such that

$$|\partial_p^\alpha\,\partial_x^\beta a(x,p)|\leqslant C_Q(\alpha,\beta)(1+|p|)^{m-|\alpha|}.$$

The pseudodifferential operator $a(x,D)$ is said to belong to the class $L^m(\Omega)$ if its symbol $a(x,p)$ belongs to $S^m(\Omega)$.

The following properties describe $a(x,D)$ as an operator.

If $a(x,p)\in S^m(\Omega)$ then $a(x,D)$ is a continuous operator

$$a(x,D):C_0^\infty(\Omega)\mapsto C^\infty(\Omega),$$

where $C_0^\infty(\Omega)$ denotes the class of infinitely differentiable functions with compact support in Ω. The operator $a(x,D)$ can be extended to a continuous map

$$a(x,D):\mathscr{E}'(\Omega)\mapsto\mathscr{D}'(\Omega),$$

where $\mathscr{D}'(\Omega)$ is the space of distributions on X [the dual of $C_0^\infty(\Omega)$] and $\mathscr{E}'(\Omega)$ is the space of distributions with compact support [the dual of $C^\infty(\Omega)$].

Definition: An operator is called to be regularizing if it maps

$$\mathscr{E}'(\Omega)\mapsto C^\infty(\Omega).$$

(This means that a regularizing operator transforms functions with singularities into infinitely smooth functions.)

Let $L^{-\infty}(\Omega)$ be the intersection of all $L^m(x)$, where m is real. One can prove that every operator from the class $L^{-\infty}(\Omega)$ is regularizing and every regularizing operator can be represented as an operator from the class $L^{-\infty}(\Omega)$.

The asymptotics in (4.19) and (5.14) have the following meaning in terms of classes of pseudodifferential operators: we can prove[14,16] that

$$T_j\in L^{-j}(\Omega),$$

for $j=1,2,\ldots$, and

$$(F-I_{\partial X_\eta^0}-T_1-T_2-\cdots-T_l)\in L^{-l-1}(\Omega).$$

for $l=0,1,2,\ldots$. In particular,

$$F-I_{\partial X_\eta^0}\in L^{-1}(\Omega).$$

This means that approximations in (4.20) and (5.16) allow reconstruction of discontinuities, since the discrepancy operator is a smoothing operator. It can be shown that an operator from the class $L^{-1}(\Omega)$ increases by 1 the number of derivatives of a function to which it is applied. In precise terms the following theorem holds.

Theorem: Let $a(x,D)$ be a pseudodifferential operator in Ω of the class $L^m(\Omega)$. Given any real number s the operator $a(x,D)$ can be extended as a continuous map

$$a(x,D):H^s_{\text{comp}}(\Omega)\mapsto H^{s-m}_{\text{loc}}(\Omega),$$

where $H^s_{\text{comp}}(\Omega)$ and $H^{s-m}_{\text{loc}}(\Omega)$ are the so-called Sobolev spaces of distributions.

The index s can be interpreted as a "number of derivatives." For detailed descriptions and proofs see Ref. 21 or any other reference on pseudodifferential operators.

[1] J. F. Claerbout, Geophys. **36**, 467 (1971).
[2] W. S. French, Geophys. **39**, 265 (1974).
[3] W. A. Schneider, Geophys. **43**, 49 (1978).
[4] R. M. Stolt, Geophys. **43**, 23 (1978).
[5] J. K. Cohen and N. Bleistein, Geophys. **44**, 1077 (1979).
[6] A. J. Berkhout, *Seismic Migration* (Elsevier, New York, 1980).
[7] D. E. Miller, "Integral transforms and the migration of multiple-offset borehole seismic profiles," Research Note, Schlumberger–Doll Research, 1983.
[8] B. Gelchinsky, "*Ray Asymptotic Migration* (*Basic Concepts*)," in *Expanded Abstracts of 53rd SEG Meeting* (Society of Exploration Geophysics, Las Vegas, 1983), pp. 385–387.
[9] J. A. Carter and L. N. Fraser, Geophys. **49**, 46 (1984).
[10] J. Radon, Ber. Verh. Sächs Akad. **69**, 262 (1917).
[11] I. M. Gelfand, M. I. Graev, and Z. Ya. Shapiro, Func. Anal. Appl. **3**, 24 (1969).
[12] V. Guillemin and S. Sternberg, "Geometrical Asymptotics," in *Mathematical Surveys*, Vol. 14 (AMS, Providence, RI, 1977).
[13] E. Quinto, Trans. Am. Math. Soc. **257**, 2 (1980).
[14] G. Beylkin, Ph.D. thesis, New York University, 1982.
[15] G. Beylkin, Proc. SPIE **413**, 32 (1983).
[16] G. Beylkin, Commun. Pure Appl. Math. **37**, 579 (1984).
[17] S. G. Norton and M. Linzer, IEEE Trans. Biomed. Eng. **BME-28**, 202 (1981).
[18] J. H. Rose, Wave Motion **6**, 149 (1984).
[19] J. B. Keller, Bull. Am. Math. Soc. **84**, 727 (1978).
[20] G. Beylkin (Beilkin), J. Sov. Math. **21**, 3 (1983).
[21] F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators* (Plenum, New York, 1980), Vols. 1 and 2.

Jet extension of a classical particle field: Principles of covariance and minimal coupling

Renato Grassini
Dipartimento di Matematica ed Applicazioni, Università di Napoli, via Mezzocannone 8, Napoli, Italy

(Received 25 April 1984, accepted for publication 20 July 1984)

Jet theory of a classical particle field is developed through a systematic transition from a local to a global formulation. Emphasis is laid on the role of gauge covariance and minimal coupling principles in producing a global geometrical framework which faithfully generalizes the one of gravitation theories.

I. INTRODUCTION

This paper exhibits a new approach to the theory of jet extension of a classical particle field with an internal symmetry or gauge group.[1-4]

A line of thought closely related in spirit to the constructive methods of differential geometry[5] is followed, which allows a systematic and meaningful transition from physicists' local formulation to the geometrical global setting of the theory.

The local theory, taking place in any open set of space-time, is revised in Sec. II, where the classical principles of covariance and minimal coupling (inherited from general relativity) are clearly and distinctly singled out.

The global theory is then built up in Sec. III, by patching together the foregoing local results all over the space-time.

So covariance principle is shown to correspond to the existence itself of a global jet extension of the field from the phase space where it lives[6] to the proper jet space, such extension being just characterized by a Yang–Mills field on phase space—nonrelated, in principle, to the gauge group or to the Higgs field consequently distinguished on phase space.

Then the minimal coupling principle is shown to correspond to a natural relation between the Yang–Mills field and the Higgs field, which can be expressed in geometrical terms as a compatibility condition (generalizing the well-known one of gravitation theories[7]) between a connection and a metric on phase space.

II. LOCAL THEORY

The local theory of a classical particle field describes jets of the field, i.e., possible values of the field and its derivatives, as elements of a local jet space

$$\overline{V} = V \oplus L(TU, V)$$

given by the Whitney sum, over an open set U of a space-time manifold[8] M, of a local phase space

$$V = U \times R^n$$

(trivial vector bundle with a n-dimensional real fiber R^n) and vector bundle

$$L(TU, V) = \bigcup_{x \in U} L(T_x U, V_x)$$

(set of all linear maps from fibers of tangent bundle TU into corresponding fibers of V).

In fact, the field is defined as a local vector-valued wave function on space-time or, equivalently, as a section[9]

$$\psi \in S(V)$$

and a local jet extension of the field

$$j\psi \in S(\overline{V})$$

is defined through a map

$$j: S(V) \to S(\overline{V}),$$

given by

$$j = id \oplus D, \tag{1}$$

where

$$D: S(V) \to S(L(TU, V))$$

is a differential operator on wave functions.[10]

If, in particular,

$$d: S(V) \to S(L(TU, V))$$

is the ordinary differentiation of wave functions,[11] then, denoted with $\mathrm{gl}(n, R)$ the Lie algebra of general linear group $\mathrm{GL}(n, R)$, we have the following lemma.

Lemma 1: A local jet extension (1) is characterized by a $\mathrm{gl}(n, R)$-valued one-form

$$\Gamma: U \to L(TU, \mathrm{gl}(n, R))$$

called the local Yang–Mills field, through

$$D = d + \Gamma. \tag{2}$$

Proof: Operator $D - d$ is linear (with respect to the ring of real functions on U).

Then, by putting [for any $x \in U$ and $\psi \in S(V)$]

$$\Gamma_x \cdot \psi(x) = (D - d)_x \psi,$$

one defines a linear map

$$\Gamma_x \in L(V_x, L(T_x U, V_x)) = L(T_x U, L(V_x, V_x)) = L(T_x U, \mathrm{gl}(n, R)),$$

such that

$$\Gamma: x \to \Gamma_x$$

is a $\mathrm{gl}(n, R)$-valued one-form satisfying condition(2). ■

On the other hand, a wave function ψ undergoes (without change of physical meaning) G-gauge transformations

$$\psi = \phi \cdot \psi',$$

 0022-2488/85/010109-03$02.50

i.e., vertical vector bundle automorphisms

$$\phi: V \to V$$

whose transition functions

$$\phi: U \to G$$

take their values in a given Lie group G acting on R^n as a closed subgroup of $GL(n, R)$.

The vertical action ϕ of G on V also extends to a vertical action on $\overline{V}$,

$$\bar{\phi}: \overline{V} \to \overline{V},$$

which allows one to define covariant transformations of a local jet extension $j\psi$, by

$$j\psi = \bar{\phi} \cdot j'\psi'.$$

This is just the content of the following.

Covariance principle: The local jet extension j follows, under a G-gauge transformation ϕ, the covariant law

$$j \cdot \phi = \bar{\phi} \cdot j'. \tag{3}$$

Then, if ad denotes the adjoint representation of $GL(n, R)$ in $gl(n, R)$ and θ is the canonical one-form on G, we have the following lemma.

Lemma 2: The local Yang–Mills field Γ corresponding to a covariant local jet extension j follows, under a G-gauge transformation ϕ, the pseudotensorial transformation law

$$\Gamma' = \mathrm{ad}(\phi^{-1})\Gamma + \phi^*\theta. \tag{4}$$

Proof: Owing to Eqs. (1) and (2), covariance condition (3) can be read

$$D' \cdot \phi^{-1} = \phi^{-1} \cdot D$$

or

$$d + \Gamma' = \phi^{-1} \cdot \Gamma \cdot \phi + \phi^{-1} \cdot d \cdot \phi.$$

The Leibniz rule implies

$$d \cdot \phi = d\phi + \phi \cdot d.$$

So we have

$$\Gamma' = \phi^{-1} \cdot \Gamma \cdot \phi + \phi^{-1} \cdot d\phi,$$

that is, Eq. (4). ■

It is plain that a drastic cancellation law for local Yang–Mills fields, which reduces local jet extensions to ordinary differentiation, is generally inconsistent with transformation law (4) (because of the presence in it of one-form $\phi^*\theta$, vanishing only for first kind G-gauge transformations $\phi = \text{const}$).

On the contrary, denoted by

$$g \subset gl(n, R)$$

the Lie algebra of gauge group G, the request on local Yang–Mills fields to take values in g is always consistent with transformation law (4) [where both $\mathrm{ad}(\phi^{-1})$, restricted to g, and $\phi^*\theta$ take values in g].

This is just the content of the following.

Minimal coupling principle: Local jet extensions are characterized by g-valued local Yang–Mills fields.

III. GLOBAL THEORY

An intrinsic formulation of the local jet theory of a classical particle field can be achieved by thinking of a local jet space $\overline{V}$ as a set $\overline{E}_U$ endowed with G-related gauges, i.e., bijections

$$\overline{\Phi}: \overline{V} \to \overline{E}_U,$$

related to each other by G-gauge transformations

$$\overline{\Phi}' = \overline{\Phi} \cdot \bar{\phi}. \tag{5}$$

Then a nontrivial globalization of the foregoing theory will be achieved by patching together such local spaces all over the space-time through G-gauge transformations, according to the following definition.

Definition: The jet space of a classical particle field is a set

$$\overline{E} = \cup \overline{E}_U$$

covered by the ranges of a maximal atlas

$$\overline{A}_g = \{\overline{\Phi}\}$$

of G-related gauges, over an open covering $\{U\}$ of space-time M.

The phase space of the field is then the set

$$E = \cup E_U$$

covered by the ranges of the atlas

$$A_G = \{\Phi\}$$

of restricted gauges $\Phi = \overline{\Phi}/V$.

The geometrical content of the definition is shown in the following theorem.

Theorem 1: Phase space E is a R^n-vector bundle over M, carrying a Higgs field

$$\alpha: VE \to GL(n, R)/G$$

on the $GL(n, R)$-principal fiber bundle VE of vertical linear frames of E or, equivalently, a generalized fiber metric

$$\mathrm{Ker}\, \alpha \subset VE \tag{6}$$

reduction of structure group $GL(n, R)$ to G, characterized by condition

$$S(\mathrm{Ker}\, \alpha) = A_G. \tag{7}$$

Proof: Differentiable structure, fibration, and local trivializations (linearly related to each other on fibers) are exhibited by local charts $\Phi \in A_G$.

Then let P be a G-principal bundle over M carrying a set of sections $\{s\}$ over $\{U\}$ whose transition functions are the same as those of A_G. It can be seen as a reduction of structure group $GL(n, R)$ down to G through a map $h: P \to VE$ given by the following action on sections $h \cdot s = \Phi$. Owing to maximality of A_G, we have

$$S(h(P)) = A_G.$$

Moreover, reduction P is isomorphic to reduction (6), where α is the Higgs field defined by requiring

$$\mathrm{Ker}\, \alpha = h(P).$$

Condition (7), which characterizes[12] α, then follows. ■

Theorem 2: Jet space $\overline{E}$ is the Whitney extension of phase space E,

$$\overline{E} = E \oplus L(TM, E).$$

Proof: Each gauge $\overline{\Phi} \in \overline{A}_G$ pushes forward the (vector bundle) Whitney structure of $\overline{V}$ onto $\overline{E}_U$,

$$\overline{E}_U = \Phi(V) \oplus \overline{\Phi}(L(TU, V))$$
$$= E_U \oplus L(TU, E_U)$$

and this structure on $\overline{E}_U$ is preserved by G-gauge transformations (5).

Hence the global result easily follows. ■

In the methodological context of the above definition, the particle field is to be meant as a whole collection $\{\psi\}$ of G-related wave functions—one for each gauge—which in turn define a unique global section of the phase space

$$\Psi \in S(E),$$

locally characterized by

$$\psi = \Phi^{-1} \cdot \Psi.$$

Likewise one can think of a jet extension of the field as a whole collection $\{j\}$ of properly related local jet extensions—one for each gauge. Indeed each of them can be carried through the proper gauge over jet space, by

$$J = \overline{\Phi} \cdot j \cdot \Phi^{-1}, \tag{8}$$

and then, if (on any $U \cap U' \neq \varnothing$) joining condition

$$\overline{\Phi}' \cdot j' \cdot \Phi'^{-1} = \overline{\Phi} \cdot j \cdot \Phi^{-1} \tag{9}$$

is fulfilled, maps (8) can be seen as restrictions of a unique global jet extension

$$J: S(E) \mapsto S(\overline{E}).$$

To this purpose, the key role is played by the covariance principle, whose meaning and geometrical content are shown in the following theorem.

Theorem 3: The covariance principle is the necesssary and sufficient condition for a collection $\{j\}$ of local jet extensions to define a global jet extension J.

Then any such extension is given by

$$J = id \oplus D^{\omega},$$

D^{ω} being the covariant derivative associated with a connection one-form ω on VE, called the Yang–Mills field, whose gauge pullbacks[13]

$$\Gamma = \Phi^* \omega \tag{10}$$

are the local Yang–Mills fields of $\{j\}$.

Proof: It is easily seen that, due to Eq. (5), covariance condition (3) on a collection $\{j\}$ does not differ from joining condition (9) which just ensures existence of J.

Then, as to the second statement, it can be deduced from Lemma 2 by recalling that (i) a connection one-form ω on VE is uniquely determined by local equations (10), $\{\Gamma\}$ being local Yang–Mills fields undergoing transformations (4); (ii) the covariant derivative D^{ω} is then defined by local equations

$$D^{\omega} = \overline{\Phi} \cdot D \cdot \Phi^{-1},$$

$\{D\}$ being the local differential operators defined by $\{\Gamma\}$; and (iii) the extension map $J = id \oplus D^{\omega}$ is consequently characterized by local equations (8), $\{j\}$ being the covariant local jet extensions defined by $\{D\}$. ■

The physical significance of a Yang–Mills field ω coupling with the particle field through a jet extension of the latter, rests on that it cannot be made to globally vanish, due to its nontensorial character under G-gauge transformations.

Conversely, possible gauge-tensorial components of ω are all to be thought of as physically insignificant fields which can be asked to vanish.

To this purpose the key role is played by the minimal coupling principle, whose meaning and geometrical content are shown in the following theorem.

Theorem 4: The minimal coupling principle is the necessary and sufficient condition for a collection $\{j\}$ of covariant local jet extensions to define a gauge-tensorial component free Yang–Mills field ω.

Then any such Yang–Mills field ω is compatible with Higgs field α.

Proof: We recall that a gauge-tensorial component of a Yang–Mills field ω is the projection ω_h of restriction $\omega/\mathrm{Ker}\,\alpha$ onto any ad(G)-invariant subspace h complementary of g in gl(n, R). It is easy, in fact, to recognize (through Lemma 2) a tensorial transformation law

$$\Gamma'_h = \mathrm{ad}(\phi^{-1})\Gamma_h$$

for the gauge pullbacks

$$\Gamma_h = \Phi^* \omega_h,$$

whereas a pseudotensorial transformation law (4) is maintained by the gauge pullbacks of the projection of $\omega/\mathrm{Ker}\,\alpha$ onto g.

So, the minimal coupling principle comes out to require $\omega_h = 0$ for all the above subspaces h or, equivalently, that $\omega/\mathrm{Ker}\,\alpha$ be a (g-valued) connnection one-form on Ker α, which is the definition of compatibility between ω and α.

ACKNOWLEDGMENT

This work was performed under the auspices of the Italian Research Council CNR.

[1]R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[2]W. Drechsler and H. E. Mayer, "Fibre Bundle Techniques in Gauge Theories," in *Lecture Notes in Physics*, Vol. 67 (Springer-Verlag, Berlin, 1977).

[3]A. Trautman, Czech. J. Phys. B **29**, 107 (1979).

[4]A. Trautman, "Fibre Bundles, Gauge Fields, and Gravitation," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).

[5]S. Eilenberg and H. Cartan, "Foundations of Fibre Bundles," in *International Symposium on Algebraic Topology* (UNAM, Mexico City, 1958).

[6]R. Grassini, "The Phase space of a Classical Particle Field: Spontaneous Symmetry Breaking and Higgs Metric," (to appear in Red. Circ. Mat. Pal.).

[7]A. Trautman, Rep. Math. Phys. (Torun) **10**, (1976).

[8]All manifolds and mappings will be assumed to be C^{∞}.

[9]$S(\cdot)$ will denote the set of all sections of any fiber bundle.

[10]D is additive and satisfies the Leibniz rule (with respect to the ring of real functions on U).

[11]If $\psi = (id_U, (\psi^i)_{i=1,\dots,n})$, then $d\psi = (id_U, (d\psi^i)_{i=1,\dots,n})$.

[12]See Ref. 6.

[13]Here a gauge $\Phi \in A_G$ is meant as a section $\Delta \in S(\mathrm{Ker}\,\alpha)$.

The action-angle variables for the massless relativistic string in 1 + 1 dimensions

B. Söderberg, B. Andersson, and G. Gustafson
Department of Theoretical Physics, University of Lund, S-223 62 Lund, Sweden

(Received 31 January 1984; accepted for publication 8 June 1984)

In this paper the Poisson bracket algebra for the open massless relativistic string in the one-space- and one-time-dimensional case is considered. In order to characterize the orbit of the system the directrix function, i.e., the orbit of one of the endpoints of the string, is used. It turns out that the Poisson bracket algebra is of a very simple form in terms of the parameters of the directrix function. We use these results to construct action-angle variables for the general motion of the string. The variables are different for different Lorentz frames, with a continuous dependence. The action-angle variables of the center-of-mass frame and of the light-cone frames are of particular interest with respect to the simplicity of the Poincaré generators and the physical interpretation. For the light-cone frame variables the equivalence to a set of indistinguishable oscillators is shown, for which an excitation corresponds to an instantaneous momentum transfer to an endpoint of the string.

I. INTRODUCTION

The massless relativistic string[1] has played a large role in the development of the dual models[2] and as a model system for the breaking of a color field into stable final state particles.[3]

In a well-known early treatment of the quantization of the string[4] only the transverse degrees of freedom are taken into account. It has been pointed out[5] that the string in one space dimension does have longitudinal degrees of freedom, which correspond to a set of massless particles connected by linear potentials. For the open-end string, the simplest nontrivial mode corresponds to two such particles, one at each endpoint of the string, while more complicated modes also contain massless momentum carriers (kinks) in the interior of the string. We will in this paper derive the very simple Poisson-bracket (PB) algebra for the longitudinal degrees of freedom as expressed by means of natural parameters of the directrix function. (The directrix function is the orbit of one of the endpoints of the string.) Using these results, we construct explicitly the action-angle variables (AAV's) of the system. We find that a complete set of AAV's can be obtained in a number of ways depending on the choice of Lorentz frame. We show some particular simplifications that occur in the CM frame and in the two light-cone frames. We also show the relation of the light-cone frame AAV's to a set of independent indistinguishable oscillators, and we demonstrate how to describe an excitation of the string system in terms of these variables.

The plan of the paper is as follows. In Sec. II we review some well-known properties of the one-dimensional string and derive the Poisson-bracket algebra for the parameters describing the directrix. The introduction of constraints and some details on the derivation are referred to Appendix A. In Sec. III we show how to construct the action-angle variables in an arbitrary Lorentz frame. In Secs. IV and V we specialize to the CM frame and the light-cone frames, respectively. In Sec. VI we describe the relationship to a system of indistinguishable oscillators, and show how to describe an excitation of the string in terms of the action-angle variables of the light-cone frames. Finally, in Sec. VII we investigate the Lorentz transformation properties of the action variables, and conclude by an analysis of the generators of the Poincaré group.

II. SOME BASIC FEATURES OF THE ONE-DIMENSIONAL STRING SYSTEM

In this section we will review some well-known properties of the one-dimensional string, and derive the Poisson bracket algebra for the directrix function. For a review of the classical motion of the string see Ref. 6.

The string is defined as a two-parameter timelike surface in Minkowski space (which in our case in general will have only one space dimension) with an action that is proportional to the invariant area of the surface. When parametrized in terms of the time t and the energy σ between a point on the string and one of the endpoints (corresponding to the temporal conformal gauge), the equation of motion of the space coordinate x is simply the wave equation

$$\partial_t^2 x - \partial_\sigma^2 x = 0. \tag{2.1}$$

Furthermore, we have the gauge constraints

$$\partial_t x \, \partial_\sigma x = 0, \tag{2.2a}$$

$$(\partial_t x)^2 + (\partial_\sigma x)^2 = 1. \tag{2.2b}$$

The solution to these equations for open string boundary conditions can be expressed as

$$x(t, \sigma) = \tfrac{1}{2} A(t+\sigma) + \tfrac{1}{2} A(t-\sigma), \tag{2.3}$$

where the *directrix function* $A(\xi)$ is an arbitrary function subject to the constraint

$$|A'(\xi)| \equiv |\partial_\xi A(\xi)| = 1 \tag{2.4}$$

and the periodicity condition

$$A(\xi) + 2P = A(\xi + 2E), \tag{2.5}$$

where E and P are the total energy and momentum of the string.

We note that by setting $\sigma = 0$ in Eq. (2.3), we obtain the path of one endpoint of the string:

 0022-2488/85/010112-12$02.50

$$x(t,0) = A(t). \tag{2.6}$$

Apart from the trivial case when the directrix is just a straight line with slope plus or minus unity, corresponding to a string behaving like a massless point particle, the generic directrix in one space dimension is a sawtooth-type function, consisting of straight-line pieces, each with a slope plus or minus unity. The number of pieces within one period is always even, and we define the *N-sector* as the set of all solutions having this number equal to $2N$. This defines a string that behaves like a chain of $N+1$ massless particles connected by N linear potentials according to the effective Hamiltonian

$$H_{\text{eff}} = \sum_{k=0}^{N} |P_k| + \sum_{k=1}^{N} |x_k - x_{k-1}|. \tag{2.7}$$

In Fig. 1 the motion of two systems with $N=1$ and $N=2$ are described by space-time diagrams as well as the positions of the particles at a few representative time slits.

We will now derive the physical Poisson brackets (PB's) for the directrix function, including the constraints in Eq. (2.4) by the methods introduced by Dirac.[7]

The momentum density π on the string is given by

$$\pi(t,\sigma) = \tfrac{1}{2} A'(t+\sigma) + \tfrac{1}{2} A'(t-\sigma). \tag{2.8}$$

The naive PB's (denoted by $\{\ ,\ \}_0$) would be

$$\begin{aligned} &\{x(t,\sigma), x(t,\sigma')\}_0 = \{\pi(t,\sigma), \pi(t,\sigma')\}_0 = 0,\\ &\{x(t,\sigma), \pi(t,\sigma')\}_0 = \delta(\sigma - \sigma'), \end{aligned} \tag{2.9a}$$

together with

$$\begin{aligned} &\{x(t,\sigma), E\}_0 = \pi(t,\sigma),\\ &\{x(t,\sigma), P\}_0 = 1,\\ &\{\pi(t,\sigma), E\}_0 = \partial_\sigma^2 x(t,\sigma),\\ &\{\pi(t,\sigma), P\}_0 = 0. \end{aligned} \tag{2.9b}$$

Using Eqs. (2.3) and (2.8), we can express this in terms of the directrix (for further details, cf. Appendix A):

$$\{A(\xi), A(\xi')\}_0 = -\tilde{\epsilon}(\xi - \xi'), \tag{2.10a}$$

$$\{A(\xi), E\}_0 = A'(\xi), \tag{2.10b}$$

$$\{A(\xi), P\}_0 = 1, \tag{2.10c}$$

where $\tilde{\epsilon}$ is the periodized sign function, defined by

$$\begin{aligned} &\tilde{\epsilon}(0) = 0,\\ &\tilde{\epsilon}(\xi) = 1, \quad 0 < \xi < 2E,\\ &\tilde{\epsilon}(\xi + 2E) = \tilde{\epsilon}(\xi) + 2. \end{aligned} \tag{2.11}$$

However, this algebra is not consistent with the constraints (2.4), since, e.g.,

$$\{A'(\xi)^2, A(\xi')\}_0 = -4A'(\xi)\,\sigma(\xi - \xi') \not\equiv 0. \tag{2.12}$$

We therefore choose to redefine the PB algebra using the Dirac method.[7] This can be described as a projection of the PB's, that makes the algebra consistent with all the constraints, as shown in Appendix A.

As a result, we obtain the physical PB's as

$$\{A(\xi_1), A(\xi_2)\} = \tilde{\epsilon}(\xi_1 - \xi_2)(A'(\xi_1)A'(\xi_2) - 1), \tag{2.13a}$$

$$\{A(\xi), E\} = A'(\xi), \tag{2.13b}$$

$$\{A(\xi), P\} = 1. \tag{2.13c}$$

It is desirable to express this PB algebra in terms of a set of independent variables. To this end we parametrize the directrix in terms of the coordinates $(\xi_k, A_k = A(\xi_k))$ of the turning points [cf. Fig. (2)], ordered according to the values of ξ_k, so that

$$\xi_{k-1} < \xi_k < \xi_{k+1}, \quad k \in \mathbf{Z}, \tag{2.14}$$

and use the convention that an even (odd) index should correspond to a minimum (maximum) of the directrix, so that

$$A_{2k} < A_{2k\pm 1}, \quad k \in \mathbf{Z}. \tag{2.15}$$

Obviously this does not fix the choice of ξ_0. This will be discussed further in the next section.

From the algebra given in Eq. (2.13), one can derive the following PB algebra for the coordinates (ξ_k, A_k) (cf. Appendix A)

$$\{\xi_k, \xi_l\} = \epsilon^P_{k-l}, \tag{2.16a}$$

$$\{A_k, A_l\} = -\epsilon^P_{k-l}, \tag{2.16b}$$

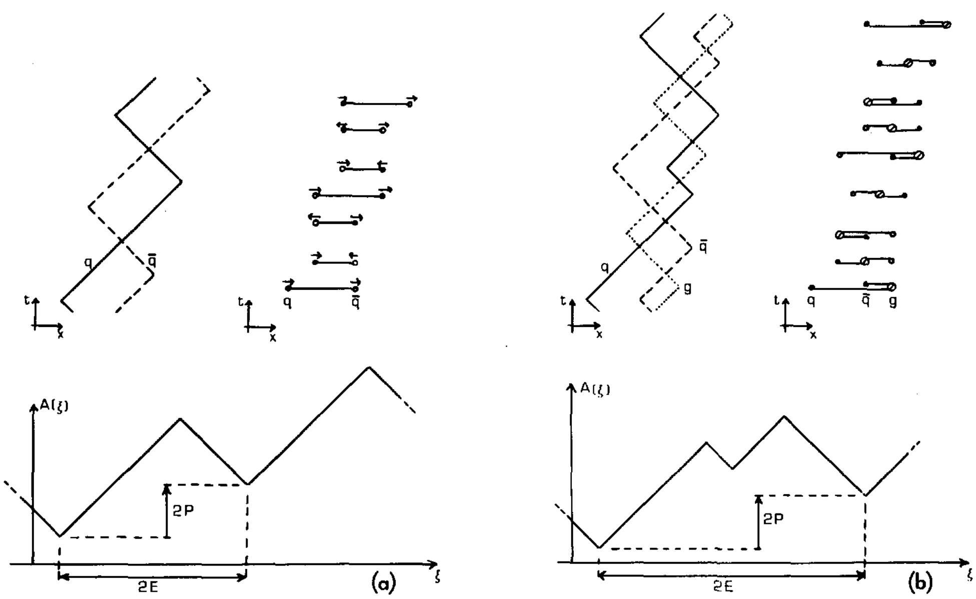


FIG. 1. (a) The motion for $N=1$ corresponding to two particles moving in a linear potential. (b) The motion for $N=2$ corresponding to three particles.

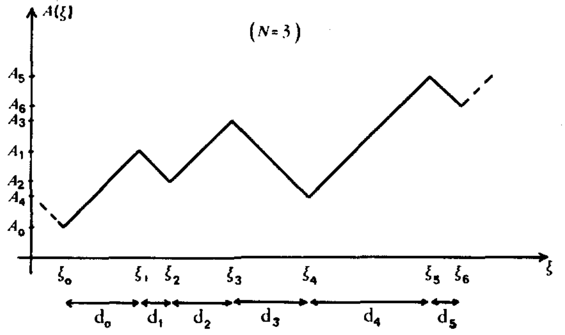


FIG. 2. The directrix parameters ξ_k, A_k, and d_k.

$$\{\xi_k, A_l\} = (-)^k \delta^P_{k-l}, \tag{2.16c}$$

where δ^P_k is the $2N$-periodical Kronecker symbol, and ϵ^P_k is the $2N$-periodical "sign" defined by [cf. Eq. (2.11)]

$$\begin{aligned} &\epsilon^P_k = 1, \quad 0<k<2N, \\ &\epsilon^P_0 = 0, \\ &\epsilon^P_{k+2N} = \epsilon^P_k + 2. \end{aligned} \tag{2.17}$$

These coordinates are not all independent, but we can define a complete set of independent variables by

$$d_k = \xi_{k+1} - \xi_k, \quad 0\leqslant k\leqslant 2N-1, \tag{2.18a}$$

$$X = \frac{1}{2N}\left(\frac{1}{2}A_0 + \sum_1^{2N-1} A_k + \frac{1}{2}A_{2N}\right), \tag{2.18b}$$

$$T = t - \frac{1}{2N}\left(\frac{1}{2}\xi_0 + \sum_1^{2N-1} \xi_k + \frac{1}{2}\xi_{2N}\right). \tag{2.18c}$$

The *directrix elements* d_k are seen to correspond to the lengths of the straight-line pieces of the directrix, while X and T measure the overall vertical and horizontal position, respectively. The explicit insertion of the time t in Eq. (2.18c) is to compensate for the time dependence involved in the transition from $x(t, \sigma)$ to $A(\xi)$ so that the equation of motion for any variable Q has the usual form

$$\dot{Q} = \{Q, E\}. \tag{2.19}$$

For this complete set, the PB algebra is given by

$$\{d_k, d_l\} = \delta^P_{k-l-1} - \delta^P_{k-l+1}, \tag{2.20a}$$

$$\{X, d_k\} = (-)^k/N, \tag{2.20b}$$

$$\{T, d_k\} = 1/N, \tag{2.20c}$$

$$\{X, T\} = 0. \tag{2.20d}$$

We note that the energy (Hamiltonian) and momentum are given by the sum rules

$$E = \frac{1}{2}\sum_0^{2N-1} d_k, \tag{2.21a}$$

$$P = \frac{1}{2}\sum_0^{2N-1} (-)^k d_k. \tag{2.21b}$$

We conclude by deriving the Lorentz transformation properties of the above directrix parameters. Since the directrix represents the motion of one endpoint [cf. Eq. (2.6)], $(\xi, A(\xi))$ must transform as a coordinate vector under a Lorentz transformation, and this must in particular be true for the coordinates (ξ_k, A_k) of the turning points. This implies that also

$$(t-T, X) \equiv \frac{1}{2N}\left(\frac{1}{2}(\xi_0, A_0) + \sum_1^{2N-1}(\xi_k, A_k) + \frac{1}{2}(\xi_{2N}, A_{2N})\right) \tag{2.22}$$

transforms as a coordinate vector. Furthermore, since

$$(\xi_{k+1}, A_{k+1}) - (\xi_k, A_k) = (d_k, (-)^k d_k), \tag{2.23}$$

we infer that the d_k's must transform as light-cone components of vectors. Specifically, for k even (odd), d_k transforms like $E+P$ $(E-P)$ in agreement with

$$E + P \equiv P_+ = \sum_{k \text{ even}} d_k, \tag{2.24a}$$

$$E - P \equiv P_- = \sum_{k \text{ odd}} d_k. \tag{2.24b}$$

This means that the effect of a Lorentz transformation on the d_k's is a rescaling of the even-numbered ones with a common factor e^y, and of the odd-numbered ones with the inverse factor e^{-y}, where y is the rapidity associated with the transformation.

III. THE CONSTRUCTION PROCEDURE FOR ACTION-ANGLE VARIABLES IN AN ARBITRARY LORENTZ FRAME

In this section we will, based upon the complete set of independent variables and the PB algebra (2.20) given in Sec. II, describe a general procedure to construct action-angle variables for the massless string in one space dimension.

This will be done in two steps. In the first step we construct a useful set of generalized coordinates and momenta, i.e., a set of canonically conjugate variables "diagonalizing" the PB algebra. In the second step we construct a canonical transformation to a new set of variables (P,Q) and (J_k, θ_k) $k = 1,...,N$ with the particular properties that (a) (P,Q) with P the total momentum, describes the translational degree of freedom for the total system; and (b) (J_k, θ_k) $k = 1,...,N$ are action-angle variables describing the periodic internal degrees of freedom. By action-angle variables[8] we mean the following.

(i) (J_k, θ_k) should be canonical variables, i.e.,

$$\{J_k, \theta_l\} = -\delta_{k,l}, \tag{3.1a}$$

$$\{\theta_k, \theta_l\} = \{J_k, J_l\} = 0; \tag{3.1b}$$

(ii) The action variables J_k are constants of motion, i.e., the Hamiltonian is functionally independent of the corresponding angle variables θ_k.

(iii) The action variables correspond to independent modes of the internal motion in the sense that there are no correlated restrictions on the values they can take on; in particular they are bounded independent of each other by

$$J_k \geqslant 0, \quad k = 1,...,N. \tag{3.2}$$

(iv) The directrix function describing the motion expressed as

$$A(\xi; \theta_k, J_k, Q, P) \tag{3.3}$$

is a continuous function of the action-angle variables and is periodic, with unit period, in each of the angle variables θ_k.

We start with the first step.

A. The construction of a set of canonical variables

There are evidently an infinite number of sets of canonical variables but for our purposes the following approach is useful. Consider the string in the $N = 1$ sector, i.e., when the directrix contains only two pieces. The directrix is then completely described by the variables (X, T, d_0, d_1) defined in Sec. II [Eq. (2.18)] and we may use the two canonical pairs $(X, P = (d_0 - d_1)/2)$ and $(T, E = (d_0 + d_1)/2)$.

For the case when the string is in a sector with $N > 1$ we will describe a general procedure to reduce the problem to $N = 1$. On the way we will at each step "downwards" construct one pair of canonical variables together with a "daughter directrix." For the first step we note that (T, E) is a useful pair, and we will now show that the remaining degrees of freedom in a natural way correspond to a reduced directrix function which is in the $(N-1)$ sector. To see that we make the following construction.

(i) Find the smallest of the $2N$ directrix elements d_{k_0}.

(ii) Subtract this element from all the elements d_k.

It is then evident that the element d_{k_0} itself will disappear and that its two neighbors d_{k_0-1} and d_{k_0+1} in that way will fuse into a single element. Consequently, the number of directrix elements is reduced to $2(N-1)$ and we regard them as the elements d'_k of the (first) daughter directrix (cf. Fig. 3).

The numbering, i.e., the labeling by an index k, of the elements d'_k of the daughter directrix is to some extent arbitrary (remember the corresponding problem for the original directrix elements as discussed in Sec. II). Our convention, which essentially follows from continuity arguments, is

$$\begin{aligned} d'_{k_0-2} &= d_{k_0-2} - d_{k_0}, \\ d'_{k_0-1} &= d_{k_0-1} - 2d_{k_0} + d_{k_0+1}, \\ d'_{k_0} &= d_{k_0+2} - d_{k_0}. \end{aligned} \tag{3.4}$$

It is further necessary to determine position variables T' and X' for the daughter directrix with properties similar to T and X. These variables are actually essentially unique in case we make the following requirements.

(i) The correct PB algebra with the d'_k variables as well as with each other.

(ii) In case two neighboring directrix elements have the same size and either can be chosen as the smallest one, the value of T' and X' should be independent of the choice.

Then we obtain

$$\begin{aligned} T' = \frac{1}{4N(N-1)} &\sum_{k=0}^{2N-1} (d_k - d_{k_0}) \\ &\times (2k - 2N + 1 - 2N\epsilon^P_{k-k_0}), \end{aligned} \tag{3.5a}$$

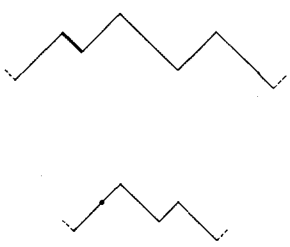


FIG. 3. The reduction of a directrix with $N = 3$ to a daughter directrix with $N = 2$.

$$\begin{aligned} X' = X - \frac{1}{4N(N-1)} &\sum_{k=0}^{2N-1} (-)^k (d_k - d_{k_0}) \\ &\times (2k - 2N + 1 - 2N\epsilon^P_{k-k_0}). \end{aligned} \tag{3.5b}$$

These variables are also closely related to the horizontal and vertical positions of the fusion point on the reduced directrix.

In Appendix B it is proven that the variables d'_k, T', and X' defined above do have a PB algebra characteristic of a real directrix as defined in Sec. II and that they all have vanishing PB's with T and E.

The daughter directrix and the pair

$$T_N \equiv T, \quad E_N \equiv E \tag{3.6}$$

correspond to the first step in the reduction. Obviously the method can be iterated and after $N-1$ steps, each time splitting off a canonical pair (T_k, E_k) we end up in the $N = 1$ sector and we have consequently in that way constructed a complete set of canonical variables: (T_k, E_k), $k = 1,\ldots,N$ together with (X_1, P_1).

Before we construct an action-angle variable set from the variables defined above, we make a number of observations.

(I) We note that in every step of the reduction, the energy is reduced while the momentum is unchanged so that

$$E_N \geqslant E_{N-1} \geqslant \cdots \geqslant E_1 \geqslant |P_1| \equiv |P|. \tag{3.7}$$

(II) It is further obvious that the procedure is frame dependent in the sense that "the smallest element" d_{k_0} in one frame is in general not the smallest one in another frame.

(III) The labeling ambiguity both in the original directrix and in the reduced ones corresponds to a certain ambiguity in value for the corresponding time variables T_k. If we relabel the original N sector directrix by moving the indices of the directrix elements d_k a complete period

$$d_k \to \tilde{d}_k = d_{k+2N}, \tag{3.8}$$

then the corresponding time variable T_N is shifted as

$$T_N \to \tilde{T}_N = T_N - 2E_N. \tag{3.9}$$

Similarly, a close analysis shows (for details cf. Appendix C) that similar relabelings of the reduced directrices always give shifts to the time variables T_k by amounts proportional to the corresponding energy variables E_k. At the same time the position variables X_k will acquire shifts proportional to P. Evidently, the state of motion as described by the directrix will remain unchanged by such operations.

B. The construction of action-angle variables

We define a set of new canonical variables from the set defined above:

$$\begin{aligned} &\varphi_N = T_N/2E_N, \quad && E_N^2, \\ &\vdots && \vdots \\ &\varphi_1 = T_1/2E_1, && E_1^2, \\ &\varphi_0 = X_1/2P, && P^2. \end{aligned} \tag{3.10}$$

According to the property (III) above a relabeling of the directrix elements will imply a shift of the φ_k's by an amount independent of the dynamical variables in the problem. This is reminiscent of the properties of the angle variables in action-angle theory,[8] i.e., when the system is moved in phase space a full period then the angle variables will acquire a simple numerical shift.

We note that our requirement (iii) for the action variables is not satisfied because the E_k^2 are not independent but satisfy the bounds

$$E_N^2 \geqslant E_{N-1}^2 \geqslant \cdots \geqslant E_1^2 \geqslant P^2 \geqslant 0. \tag{3.11}$$

This immediately suggests the transformation to a further set of canonical variables:

$$\begin{aligned}
\tilde{\theta}_N &= \varphi_N, & \tilde{J}_N &= E_N^2 - E_{N-1}^2,\\
\tilde{\theta}_{N-1} &= \varphi_N + \varphi_{N-1}, & \tilde{J}_{N-1} &= E_{N-1}^2 - E_{N-2}^2,\\
&\vdots & &\vdots\\
\tilde{\theta}_1 &= \varphi_N + \cdots + \varphi_1, & \tilde{J}_1 &= E_1^2 - P^2,\\
\tilde{\theta}_0 &= \varphi_N + \cdots + \varphi_0, & \tilde{J}_0 &= P^2.
\end{aligned} \tag{3.12}$$

A detailed analysis (cf. Appendix C) shows that a relabeling of the k-sector directrix elements

$$d_n^{(k)} \to \tilde{d}_n^{(k)} = d_{n-2}^{(k)} \tag{3.13}$$

will induce a shift in the corresponding angle variable $\tilde{\theta}_k$ by

$$\tilde{\theta}_k \to \tilde{\theta}_k + 1/k. \tag{3.14}$$

Obviously, such a relabeling will not change the configuration. Hence, the configuration is periodic in $\tilde{\theta}_k$ with period $1/k$. Thus, in order to have unit period, the final angle variables must be

$$\theta_k = k\tilde{\theta}_k, \quad k = 1,\ldots,N. \tag{3.15}$$

The corresponding action variables are then after this rescaling

$$J_k = (1/k)\tilde{J}_k, \quad k = 1,\ldots,N. \tag{3.16}$$

For the remaining degree of freedom, we transform to the more natural variables

$$Q \equiv 2P\tilde{\theta}_0, \quad P. \tag{3.17}$$

Summarizing, we have the final variables

$$\begin{aligned}
\theta_N &= N\varphi_N, & J_N &= (1/N)(E_N^2 - E_{N-1}^2),\\
\theta_{N-1} &= (N-1)(\varphi_N + \varphi_{N-1}), & J_{N-1} &= [1/(N-1)]\\
& & &\quad \times (E_{N-1}^2 - E_{N-2}^2),\\
&\vdots & &\vdots\\
\theta_1 &= \varphi_N + \cdots + \varphi_1, & J_1 &= E_1^2 - P^2,\\
Q &= X_1 + 2P(\varphi_N + \cdots + \varphi_1), & &P.
\end{aligned} \tag{3.18}$$

The Hamiltonian is given by

$$H \equiv E_N = \left(P^2 + \sum_1^N kJ_k\right)^{1/2}. \tag{3.19}$$

Thus, the total mass square is simply

$$M^2 = \sum_{k=1}^N kJ_k. \tag{3.20}$$

The construction of the set of action-angle variables described above implies a connection between the different sectors that makes it possible to introduce a *unified phase space* for all the sectors. In fact the following statement is true.

If an N-sector directrix is regarded as an $(N+1)$-sector directrix with one vanishing directrix element, then the construction of action-angle variables above will lead to the set (P,Q), (J_k, θ_k) $k = 1,\ldots,N$ with the same values as if they had been computed "directly" in the N-sector. The difference is that we obtain an extra "sleeping" degree of freedom corresponding to the pair (J_{N+1}, θ_{N+1}) with

$$J_{N+1} = 0 \tag{3.21}$$

and with θ_{N+1} related to the position of the vanishing directrix element.

Thus the difference between the different sectors is quantitative rather than qualitative and it is both consistent and useful to introduce a unified phase space for the string in one space dimension. Thus we describe the string in terms of the (infinite) set of variables

$$\begin{aligned}
&(Q, P),\\
&(\theta_k, J_k), \quad k = 1,2,\ldots,
\end{aligned} \tag{3.22}$$

with the Hamiltonian

$$H = \left(P^2 + \sum_1^\infty kJ_k\right)^{1/2}. \tag{3.23}$$

The "old" N-sector in this unified phase-space will correspond to the part where

$$J_k = 0, \quad k > N. \tag{3.24}$$

We also note that in case two directrix elements are equal and either may be chosen as the smallest one, the value of one of the angle variables will in general depend on the choice. However, in this case the corresponding action variable is zero (cf. Appendix B). This indicates that whenever a certain mode is excited and thus has a nonvanishing action variable, the corresponding angle variable is unambiguously determined (modulo 1) by our construction method.

With respect to the identification of the different modes, it is obvious that a pure N-mode corresponds to a directrix with $2N$ directrix elements, all of them of equal size. In that case the size of the action variable J_k defines the directrix element size and all the remaining action variables, as well as the total momentum, vanish. This corresponds to a string which is folded $(N-1)$ times. In the usual (massless) particle interpretation language the directrix generates a string with two particles at the (free) endpoints and $(N-1)$ kink particles. The motion consists of a symmetric oscillation with the particles moving along the light cones, all the evenly labeled ones along one of the light cones and all the oddly labeled ones along the opposite light cone.

The relation between energy and frequency and a physical interpretation of the excitation of the string are discussed in Sec. VI.

IV. THE CM FRAME ACTION-ANGLE VARIABLES

The procedure given in Sec. III is evidently applicable to the description of the string in an arbitrary Lorentz frame, and as we have already noted the action-angle variables constructed in that way will in general be different for different frames. Actually, since the elements of the daughter directrices are linear combinations of original directrix elements with, in general, different transformation properties, the transformation properties of the action-angle variables will be rather involved. This problem is further elaborated in Sec. VII. We note that the CM frame in some sense may be considered to be a "natural" frame for describing the internal degrees of freedom. It has the particular property that the transformation from an arbitrary frame to the CM frame depends on the dynamical variables, i.e., on the total momentum and energy, and therefore this case must be treated with some care.

To deal with this problem we start by considering the directrix in an N-sector mode in a fixed frame to be called the lab frame. In the usual way the directrix is parametrized by the independent variables T, X, and d_k, $k=1,...,2N$ described in Sec. II [Eq. (2.18)].

If the directrix is transformed to the CM frame, one degree of freedom, i.e., the total momentum, is lost because $P_{\rm CM}$ is identically zero in the CM frame. This degree of freedom must then be represented by the lab frame total momentum P together with a suitable conjugate variable.

We note that since $(\xi, A(\xi))$ transforms under Lorentz transformations as a coordinate vector, the directrix function will transform as

$$A^{\rm CM}\left(\frac{E}{M}\xi-\frac{P}{M}A(\xi)\right)=\frac{E}{M}A(\xi)-\frac{P}{M}\xi \tag{4.1}$$

and the directrix elements as

$$d_k^{\rm CM}=([E-(-)^kP]/[E+(-)^kP])^{1/2}d_k. \tag{4.2}$$

There is essentially only a single variable X_c which on the one hand is conjugate to P and on the other hand has vanishing PB's to all of the $d_k^{\rm CM}$:

$$X_c=X+\frac{P}{E}T+\frac{1}{4NE}\times\sum_{k,l=0}^{2N-1}(-)^k d_k\,d_l(k-l-N\epsilon(k-l)) \tag{4.3}$$

Here ϵ is the ordinary sign function defined to vanish for zero argument. Although the expression in Eq. (4.3) is seemingly a complicated one, X_c is actually identical to the string center of energy coordinate

$$X_c=\frac{1}{E}\int_0^E x(t,\sigma)\,d\sigma. \tag{4.4}$$

We also need a variable $T^{\rm CM}$ corresponding to the time variable T for the CM directrix. The requirement that it should commute with X_c will essentially fix it to

$$T^{\rm CM}=\frac{E}{M}T+\frac{P}{M}X-\frac{P}{M}X_c. \tag{4.5}$$

There is no difficulty to see that $T^{\rm CM}$ fulfills [cf. Eq. (2.18c)]

$$T^{\rm CM}=\frac{E}{M}t-\frac{P}{M}X_c-\frac{1}{2N}\left(M+\sum_{k=0}^{2N-1}\xi_k^{\rm CM}\right)$$
$$=t_c-\frac{1}{2N}\left(\frac{1}{2}\xi_0^{\rm CM}+\sum_1^{2N-1}\xi_k^{\rm CM}+\frac{1}{2}\xi_{2N}^{\rm CM}\right), \tag{4.6}$$

with t_c the time in the CM frame as measured in the point X_c.

In that way we have split off the overall translational degree of freedom represented by the canonical pair (X_c, P) from the internal degrees of freedom described by means of the variables

$$T^{\rm CM},\quad d_k^{\rm CM},\quad k=0,...,2N. \tag{4.7}$$

We note the following constraints on the variables $d_k^{\rm CM}$ corresponding to the total energy momentum in the CM frame:

$$\sum_0^{2N-1} d_k^{\rm CM}=2M, \tag{4.8a}$$

$$\sum_0^{2N-1}(-)^k d_k^{\rm CM}=0. \tag{4.8b}$$

By means of the internal variables in Eq. (4.7) we may then carry out the construction of the CM frame action-angle variables using the reduction procedure discussed in Sec. III.

The construction of the variables $(T_k^{\rm CM}, E_k^{\rm CM}\equiv M_k)$ $k=1,N$, the set $(\varphi_k^{\rm CM}=T_k^{\rm CM}/2M_k, M_k^2)$ $k=1,...,N$, and, finally, the action-angle variables,

$$\begin{aligned}
\theta_N^{\rm CM}&=N\varphi_N^{\rm CM}, & J_N^{\rm CM}&=(1/N)(M_N^2-M_{N-1}^2),\\
&\vdots & &\vdots\\
\theta_2^{\rm CM}&=2(\varphi_N^{\rm CM}+\cdots+\varphi_2^{\rm CM}), & J_2^{\rm CM}&=\tfrac{1}{2}(M_2^2-M_1^2),\\
\theta_1^{\rm CM}&=\varphi_N^{\rm CM}+\cdots+\varphi_1^{\rm CM}, & J_1^{\rm CM}&=M_1^2,
\end{aligned} \tag{4.9}$$

proceeds in the same way and the internal action-angle set in Eq.(4.9) together with the momentum P and cms coordinate X_c constitute a complete set of canonical variables.

V. THE LIGHT-CONE ACTION-ANGLE VARIABLES

The reduction process described in Secs. III and IV to construct action-angle variables will as always for collective variables be of a nonlocal nature, i.e., the variables are not easily associated with the local properties of the space-time orbit of the system. There is in this regard a particular simplicity in connection with the light-cone frame variables (to be further elaborated in the next section) which motivates a separate treatment. It will be sufficient to treat one of the frames in detail, e.g., the forward case—the backward case is completely analogous.

The forward light-cone variables can be obtained in the limit of a very large boost in the forward direction from the lab frame in such a way that the total momentum of the system becomes large and positive. For an N-sector directrix the lab-frame directrix elements (d_k) $k=1,...,N$ will transform as

$$d_k\to d_k'=\begin{cases}\epsilon^{-1}d_k, & k \text{ even},\\ \epsilon\, d_k, & k \text{ odd},\end{cases} \tag{5.1}$$

with ϵ a parameter describing the transformation. Eventually the limit $\epsilon\to 0$ will be taken.

The light-cone versions of the time and position variables T_N and X_N will be chosen as

$$T_N^{\rm LC}=(1/2\epsilon)(T_N-X_N)+(\epsilon/2)(T_N+X_N), \tag{5.2a}$$

$$X_N^{\rm LC}=-(1/2\epsilon)(T_N-X_N)+(\epsilon/2)(T_N+X_N). \tag{5.2b}$$

If the parameter ϵ is chosen small enough, it is evident that in *every* step of the reduction algorithm an odd-labeled element will be the smallest.

In the limit $\epsilon\to 0$ all the odd elements become very small and give negligible contributions to the variables T_N, which thus will depend on only the even elements, that are of order $1/\epsilon$. For the scaled variables $\varphi_n = T_n/2E_n$ the dependence on ϵ disappears, because also E_n are of order $1/\epsilon$. However, the differences between the variables E_n^2 are determined by the smallest elements used in the reduction algorithm, which always are linear combinations of the original odd elements. These therefore determine the variables

$$J_n = (1/n)(E_n^2 - E_{n-1}^2). \tag{5.3}$$

Our conclusion is thus that the *angle variables* $\theta_n^{\rm LC}$ *are determined primarily by the even elements, whereas the action variables* $J_n^{\rm LC}$ *are determined essentially by the odd elements.* In the backward light-cone frame the situation will obviously be the reversed.

In addition to the action-angle variables, which describe the internal motion, we can for the translational degree of freedom use the variables

$$Q_+^0 = \lim_{\epsilon\to 0}\left(\frac{1}{2\epsilon}X_1^{\rm LC} + \frac{P_+}{2\epsilon^2}\theta_1^{\rm LC}\right), \quad P_+ = E + P. \tag{5.4}$$

The light-cone frame action-angle variables can also be obtained in a more direct way, which we describe below to illuminate their simple structure.

We introduce light-cone coordinates in the lab frame (ξ, A)-plane, and define

$$\xi_k^{\pm} = \tfrac{1}{2}(\xi_k \pm A_k), \tag{5.5}$$

where (ξ_k, A_k) are the coordinates of the corners in the directrix. Because the directrix elements lie along the light cones we have

$$\xi_{2k}^+ = \xi_{2k-1}^+, \quad \xi_{2k}^- = \xi_{2k+1}^-. \tag{5.6}$$

The lengths of the directrix elements d_k are then given by

$$d_k = \begin{cases} \xi_{k+1}^- - \xi_{k-1}^-, & k \text{ odd},\\ \xi_{k+1}^+ - \xi_{k-1}^+, & k \text{ even}.\end{cases} \tag{5.7}$$

From the PB algebra for the directrix parameters given in Eqs. (2.16) we conclude

$$\{d_k, \xi_l^-\} = \delta_{k,l}^P, \quad k \text{ even}, \tag{5.8a}$$

$$\{d_k, \xi_l^+\} = \delta_{k,l}^P, \quad k \text{ odd}. \tag{5.8b}$$

The directrix is fully determined if we fix a starting point (ξ_0, A_0) and either the positions in the $(+)$ direction and lengths in the $(-)$ direction of the odd elements $(\xi_k^+, d_k;\ k$ odd) or the positions in the $(-)$ direction and the lengths in the $(+)$ direction of the even elements $(\xi_k^-, d_k,\ k$ even).

From Eq. (5.8) we find that either set of parameters may be used as canonical variables. If we add the time and rescale with P_+ we arrive at the following set of canonical variables for the first case:

$$\left.\begin{aligned} \alpha_k &= (1/{\rm P}_+)((t/2) - \xi_{2k-1}^+),\\ I_k &= P_+\, d_{2k-1},\end{aligned}\right\} k = 1,\ldots,N. \tag{5.9}$$

Together with P_+ and its conjugate coordinate

$$Q_+ = \frac{1}{2}t - \frac{1}{2}(\xi_0^- + \xi_{2N}^-) - \frac{1}{P_+}\sum_1^N I_k\,\alpha_k \tag{5.10}$$

they constitute a complete set.

At this point we make the following observations.

(I) The insertion of the explicit time dependence in connection with the definition of the α-variables is needed in order to fulfill the (naive) equations of motion with the lab-frame energy E as the Hamiltonian [cf. Eq. (2.18c)].

(II) The variable Q_+ in Eq. (5.10) is essentially equal to the variable Q_+^0 introduced in Eq. (5.4):

$$Q_+^0 = Q_+ + P_+\cdot \text{const}. \tag{5.11}$$

(III) The periodicity of the directrix implies

$$\xi_{k+2N}^{\pm} = \xi_k^{\pm} + P_{\pm}. \tag{5.12}$$

Thus, the rescaling with P_+ is done in such a way that a change in starting point (ξ_0, A_0) by one period implies a change in α_k by one unit, while I_k and Q_+ remain unchanged:

$$\alpha_k \to \alpha_k + 1, \quad k = 1,\ldots,N, \tag{5.13a}$$

$$Q_+ \to Q_+, \tag{5.13b}$$

$$I_k \to I_k, \quad k = 1,\ldots,N. \tag{5.13c}$$

(IV) Due to the obvious inequalities

$$\xi_1^+ \leqslant \xi_3^+ \leqslant \cdots \leqslant \xi_{2N-1}^+, \tag{5.14}$$

we have

$$\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_N, \tag{5.15}$$

i.e., the canonical pairs (α_k, I_k) are labeled according to an ordering with respect to the numerical values of the α's.

It is evidently possible to introduce another ordering principle, e.g., according to the magnitudes of the momentum variables I_k.

In that way we obtain from the set (α_k, I_k) $k = 1,\ldots,N$ another set (β_k, K_k) $k = 1,\ldots,N$ with

$$K_1 \leqslant K_2 \leqslant \cdots \leqslant K_N. \tag{5.16}$$

By a comparison with the definitions of the action-angle variables in the light-cone frame we expect that the set

$$\begin{array}{ll} \beta_1 + \cdots + \beta_N, & K_1,\\ \beta_2 + \cdots + \beta_N, & K_2 - K_1,\\ \vdots & \vdots\\ \beta_N, & K_N - K_{N-1},\end{array} \tag{5.17}$$

should be equivalent to our "old" set of action-angle variables $(\theta_k^{\rm LC}, J_k^{\rm LC})$, $k = 1,\ldots,N$. The action variables are identical and except for nondynamical additive numerical constants also the angle variables can be shown to be the same. Thus in the (forward) light-cone frame, the action angle variables seem to have an especially simple structure. The action variables are simply differences between the mass variables K_k, which correspond to the lengths of the odd directrix elements, rescaled with P_+. The angle variables are simple linear combinations of the variables β_k, which are the positions in the ξ^+-direction of the same elements, rescaled with P_+^{-1}. The total mass squared is evidently

$$M^2 = \sum_1^N K_k = \sum_1^N I_k = \sum_1^N kJ_k. \tag{5.18}$$

When we consider the set (K_k) or the set (J_k) we are concerned with the ordering in size of these mass variables. When we instead consider the set (I_k), we are actually ordering the same mass variables in accordance with positions along the positive light cone, i.e., according to the numerical values of the position variables α_k in Eq. (5.15).

VI. THE STRING SYSTEM AS A SET OF OSCILLATORS

In this section we will show that the string system is equivalent to a system of indistinguishable oscillators. We show in some detail the relationship between the dynamical variables of such an oscillator system and the light-cone frame action-angle variables of the string. We end by exhibiting the effect of exciting the string system by giving an impulse to one of its endpoints as expressed in terms of the dynamical variables of the oscillator system.

We start by considering a set of N oscillators with equal frequency ω_0. The oscillators will be labeled by indices k in an arbitrary way and the system is then described by action-angle variables (I_k, α_k) $k = 1,...,N$ which are all independent.

An excitation of the k th oscillator described by n_k will be such that

$$I_k = 2\pi n_k, \tag{6.1}$$

and the total excitation energy is given by

$$H = \omega_0 \sum_1^N n_k. \tag{6.2}$$

In a quantum description of the oscillator system all the excitations n_k correspond to integers and the excitation energy J_k is then the energy above the ground state energy. We will call the oscillators indistinguishable if any pair of action-angle variables can be exchanged without change in the state of motion of the system. Thus, e.g., the states $(\alpha_1, I_1; \alpha_2, I_2)$ and $(\alpha_2, I_2; \alpha_1, I_1)$ correspond to the same motion of the system. We note that we may under those circumstances find a unique way to represent a state by labeling the oscillators according to the values of α_k:

$$\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_N \Rightarrow I_k \text{ independent.} \tag{6.3}$$

Alternatively, we can reorder the oscillators according to the magnitude of the *action variables* and represent the same state by a set (β_k, K_k), $k = 1,...,N$, where

$$\begin{aligned} &K_1 = I_{k_1} = \max(I_k), && \beta_1 = \alpha_{k_1}, \\ &K_2 = I_{k_2} = (\text{next largest } I_k), && \beta_2 = \alpha_{k_2}, \\ &\text{etc.} \end{aligned} \tag{6.4}$$

Then we have

$$K_1 \geqslant K_2 \geqslant \cdots \geqslant K_N \geqslant 0 \Rightarrow \beta_k \text{ independent.} \tag{6.5}$$

We note that in the second case the requirement (iii) in Sec. III for independence is not fulfilled but this is easily remedied. We can replace the K's by their differences and define the variables

$$\begin{aligned} &\theta_1 = \beta_1, && J_1 = K_1 - K_2, \\ &\theta_2 = \beta_1 + \beta_2, && J_2 = K_2 - K_3, \\ &\vdots && \vdots \\ &\theta_N = \beta_1 + \cdots + \beta_N, && J_N = K_N. \end{aligned} \tag{6.6}$$

These variables are all independent and the energy is given by

$$H = \frac{\omega_0}{2\pi} \sum_{k=1}^N kJ_k = \omega_0 \sum_{k=1}^N km_k, \tag{6.7}$$

where $m_1 = n_{k_1} - n_{k_2}$, etc., i.e., $J_k = \omega_0 m_k$.

Thus we conclude that a set of N indistinguishable oscillators with frequencies ω_0 is equivalent to a set of N distinguishable oscillators with frequencies $k\omega_0$ ($k = 1,...,N$). We also observe that the action-angle variables for the string in the light-cone frame satisfy the same relations with the energy H above replaced by M^2.

Because of this difference, the frequencies of the string motion are not fixed as for an ordinary harmonic oscillator. The frequency of the k th mode is instead given by

$$\omega_k = 2\pi\dot{\theta}_k = 2\pi \frac{\partial H}{\partial J_k} = \frac{\pi k}{H}. \tag{6.8}$$

For a pure N-mode we have in the cms $H = M = \sqrt{NJ_N}$ and the frequency becomes

$$\omega_N^{\text{pure}} = \pi N / \sqrt{NJ_N} = \pi \sqrt{N/J_N}. \tag{6.9}$$

In this picture with the string system described as a set of N indistinguishable oscillators, it is easy to visualize an excitation of the system. Here the oscillator variables I_k, α_k are the lengths in the negative light-cone direction, and the positions in the positive light-cone direction, of the odd k directrix elements d_k [cf. Eq. (5.9)], scaled by the factors P_+ and $1/P_+$, respectively.

An excitation of the k th oscillator to a higher level corresponds to increasing the element d_{2k-1} by an amount δ. For the string motion this corresponds to giving the endpoint particle (the "quark") a kick with a lightlike momentum $(\Delta E, \Delta P) = (\delta, -\delta)$ in the negative direction at a time when the quark is moving in the same direction. On the other hand, if the quark would be moving in the opposite (positive) direction at the time of the impulse transfer, it cannot stay on its (zero) mass shell unless its original energy and momentum continue as a kink on the string.

This corresponds to the formation of a new tooth on the directrix, i.e., a transition from the N-sector to the $(N+1)$-sector. The new directrix element is determined by the impulse δ, and the corresponding position ξ^+ is related to the time of the interaction. Thus in this case a new oscillator is excited, the corresponding angle variable given by ξ^+/P_+, where ξ^+ is related to the space-time position of the quark at the time it is kicked, while the corresponding action variable is given by $2\delta \cdot P_+$.

In both cases, the change in α_k, I_k is simple and it is obviously easy to calculate the corresponding change in the variables (θ_k, J_k). We conclude that the action-angle variables of the forward light-cone system are well suited for giving a simple description of excitations of the above kind. We note that if the quark instead is kicked in the positive direction, $(\Delta E, \Delta P) = (\delta, \delta)$, the excitation is most easily described in the backward light-cone frame.

VII. THE LORENTZ TRANSFORMATION OF THE ACTION VARIABLES

In this section we demonstrate in a simple example how the action variables vary with the choice of Lorentz frame.

We consider for simplicity a configuration in the sector $N = 2$ that in some frame has the following directrix elements:

$$(d_0, d_1, d_2, d_3) = (4,2,1,3). \tag{7.1}$$

When a boost is applied, they will transform to

$$(d_k(z)) = (4z, 2/z, z, 3/z). \tag{7.2}$$

The parameter z is related to the boost velocity v by

$$z = ((1+v)/(1-v))^{1/2} > 0. \tag{7.3}$$

In the boosted frame, the energy will be

$$E_2(z) = \tfrac{5}{2}z + 5/2z. \tag{7.4}$$

The smallest element is

$$d_{k_0}(z) = \begin{cases} d_2(z) = z, & \text{for } z \leqslant \sqrt{2}, \\ d_1(z) = 2/z, & \text{for } z \geqslant \sqrt{2}. \end{cases} \tag{7.5}$$

Thus the energy of the reduced directrix becomes

$$E_1(z) = \begin{cases} z/2 + 5/2z, & z \leqslant \sqrt{2}, \\ 5z/2 - 3/2z, & z \geqslant \sqrt{2}. \end{cases} \tag{7.6}$$

The momentum of the reduced directrix, which of course is the same as that of the original directrix, will be

$$P(z) = 5z/2 - 5/2z. \tag{7.7}$$

It is straightforward to compute the action variables, and the result is for $z \leqslant \sqrt{2}$,

$$J_2(z) \equiv \tfrac{1}{2}(E_2^2 - E_1^2) = 3z^2 + 5, \tag{7.8a}$$

$$J_1(z) \equiv E_1^2 - P^2 = -6z^2 + 15; \tag{7.8b}$$

and, for $z \geqslant \sqrt{2}$,

$$J_2(z) = 10 + 2/z^2, \tag{7.9a}$$

$$J_1(z) = 5 - 4/z^2. \tag{7.9b}$$

This example illustrates the generic type of frame dependence for the action variables. They are continuous functions of z, that piecewise have the typical form

$$J_k(z) = A_k z^2 + B_k + C_k(1/z^2) \tag{7.10}$$

with A_k, B_k, and C_k in general different in different regions of the positive z-axis. Since all J_k's are positive and

$$\sum_k kJ_k = M^2, \tag{7.11}$$

every J_k must be bounded. We must then have $A_k = 0$ in the region closest to $z = \infty$, and $C_k = 0$ in the region closest to $z = 0$. This is obviously the case in our example. It is obvious that the Lorentz transformation properties of the action variables are not very simple. This is related to the fact that the angular momentum tensor does not have a simple representation in terms of our variables in a general frame.

There are however a few exceptions. One is for the CM-frame variables as defined in Sec. IV. In order to prove this, we note the general definition of angular momentum for an extended system

$$J^{\mu\nu} = \int (x^\mu \, dp^\nu - x^\nu \, dp^\mu). \tag{7.12}$$

In the case of one space dimension the angular momentum tensor has only one independent component $J^{10} = -J^{01}$. For the string case it is given by

$$\begin{aligned} J^{10} &= \int_0 d\sigma\Big(x(\sigma, t) - t\,\frac{dp}{d\sigma}\Big) \\ &= \int_0^E x(t, \sigma)\, d\sigma - Pt. \end{aligned} \tag{7.13}$$

Comparing with Eq. (4.4), we can rewrite this as

$$\begin{aligned} J^{10} &= EX_c - Pt \\ &= X_c\Big(P^2 + \sum_k kJ_k^{\rm CM}\Big)^{1/2} - Pt. \end{aligned} \tag{7.14}$$

A simple check shows that we have the correct PB algebra for the Poincaré-group generators E, P, and J^{10}:

$$\{ J^{10}, E \} = P, \tag{7.15a}$$

$$\{ J^{10}, P \} = E, \tag{7.15b}$$

$$\{ E, P \} = 0. \tag{7.15c}$$

The other exceptions are for the light-cone frame variables, as defined in Sec. V. The variable Q_+ defined there as conjugate to P_+ [Eq. (5.10)] is actually closely related to the center-of-energy coordinate X_c:

$$Q_+ = EX_c/P_+. \tag{7.16}$$

Thus, in terms of the forward light-cone frame variables, the boost generator J^{10} is given by

$$\begin{aligned} J^{10} &= P_+Q_+ - Pt = P_+Q_+ \\ &\quad - \frac{t}{2}\Big(P_+ - \frac{\Sigma_1^N J_k^{\rm LC}}{P_+}\Big). \end{aligned} \tag{7.17}$$

We note that the boost generator also can be expressed in a simple way in terms of the original directrix parameters as

$$\begin{aligned} J^{10} &= EA_0 - P\xi_0 \\ &\quad + \frac{1}{2} \sum_{\substack{k=0 \\ k \text{ odd}}}^{2N-1} \sum_{\substack{l=0 \\ l \text{ even}}}^{2N-1} d_k \, d_l \, \epsilon(k-l). \end{aligned} \tag{7.18}$$

We conclude that whenever one is interested in a simple representation of the Poincaré group generators one should use the CM-frame variables or the light-cone-frame variables.

APPENDIX A: THE DIRAC BRACKETS

Consider a Hamiltonian system with a set of canonical variables (q_i, p_i), and a set of constraints on them

$$c_k(q_i, p_i) = 0, \quad k = 1,\dots,2N. \tag{A1}$$

(It is necessary to have an even number, see below.)

We will require the PB between a constraint variable c_k and any other quantity to vanish. In particular, we will require for arbitrary k and i

$$\begin{aligned} \{ c_k, q_i \} &= 0, \\ \{ c_k, p_i \} &= 0. \end{aligned} \tag{A2}$$

Then, clearly, the naive PB's, i.e., those defined by

$$\{ f, g \} = \sum_k \frac{\partial f}{\partial q_k} \cdot \frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k} \cdot \frac{\partial g}{\partial q_k},$$

will not be acceptable, since

$$\begin{aligned} \{ c_k, q_i \}_0 &= -\frac{\partial c_k}{\partial p_i}, \\ \{ c_k, p_i \}_0 &= \frac{\partial c_k}{\partial q_i}, \end{aligned} \tag{A3}$$

which not all could vanish.

Dirac[7] has suggested the following projection method for adjusting the naive PB's in order to obtain Eq. (A2).

Define the (antisymmetric) matrix

$$M_{kl} = \{c_k, c_l\}_0 \tag{A4}$$

and invert it (possible for an antisymmetric matrix only if it has even dimension):

$$W_{kl} = (M^{-1})_{kl}. \tag{A5}$$

The physical PB (sometimes called the Dirac bracket) between two arbitrary variables, A and B, will then be defined as

$$\{A, B\} = \{A, B\}_0 - \sum_{kl} \{A, c_k\}_0 W_{kl} \{c_l, B\}_0. \tag{A6}$$

We wish to apply this procedure in terms of the string directrix $A_i(\xi)$ (for an arbitrary number of dimensions). The naive PB's are

$$\begin{aligned} &\{x_i(t,\sigma), \pi_j(t,\sigma')\}_0 = \delta_{ij}\delta(\sigma-\sigma'), \\ &\{x_i(t,\sigma), x_j(t,\sigma')\}_0 = \{\pi_i(t,\sigma), \pi_j(t,\sigma')\}_0 = 0, \end{aligned} \tag{A7}$$

where $x_i(t,\sigma)$ is the position and $\pi_j(t,\sigma)$ the momentum density at the point on the string defined by σ at the time t. Using

$$\begin{aligned} x_i(t,\sigma) &= \tfrac{1}{2}(A_i(t+\sigma) + A_i(t-\sigma)), \\ \pi_i(t,\sigma) &= \tfrac{1}{2}(A'_i(t+\sigma) + A'_i(t-\sigma)), \end{aligned} \tag{A8}$$

and choosing $t=0$, we get the naive PB's in terms of the directrix

$$\{A_i(\xi), A_j(\xi')\}_0 = -\delta_{ij}\epsilon(\xi-\xi'), \quad -E \leqslant \xi, \xi' < E, \tag{A9}$$

where $\epsilon(\xi-\xi')$ is the ordinary sign function, and E is the total energy of the system.

The constraints are given by

$$c(\xi) \equiv \sum_i A'_i(\xi)^2 - 1 = 0, \quad -E \leqslant \xi < E. \tag{A10}$$

From Eq. (A7) it is straightforward to derive

$$\begin{aligned} &\{A_i(\xi), c(\xi')\}_0 = 4A'_i(\xi)\,\delta(\xi-\xi'), \\ &\{c(\xi), A_j(\xi')\}_0 = -4A'_j(\xi)\,\delta(\xi-\xi'), \end{aligned} \tag{A11}$$

$$M(\xi,\xi') \equiv \{c(\xi), c(\xi')\}_0 = 8\delta'(\xi-\xi'), \tag{A12}$$

$M(\xi,\xi')$ is the generalization of the matrix M_{kl} to a continuous system. It is easily inverted, giving

$$W(\xi,\xi') = \tfrac{1}{16}\epsilon(\xi-\xi'). \tag{A13}$$

Using Eqs. (A11) and (A13) we get the physical PB's for the directrix

$$\{A_i(\xi), A_j(\xi')\} = \epsilon(\xi-\xi')(A'_i(\xi)A'_j(\xi') - \delta_{ij}), \quad -E \leqslant \xi, \xi' < E. \tag{A14}$$

In addition to $A_i(\xi)$, $-E \leqslant \xi < E$, we need the total energy (E) and momentum (P_i) to get a complete description of the system. Since they both have vanishing naive PB's with the constraints $c(\xi)$, the resulting PB's will be unchanged:

$$\{A_i(\xi), E\} = A'_i(\xi), \quad \{A_i(\xi), P_j\} = \delta_{ij}, \quad \{E, P_j\} = 0. \tag{A15}$$

Using the periodicity condition for the directrix

$$A_i(\xi + 2E) = A_i(\xi) + 2P_i, \tag{A16}$$

we may continue the PB definition for the directrix outside the interval $(-E, E)$, giving

$$\{A_i(\xi), A_j(\xi')\} = \tilde{\epsilon}(\xi-\xi')(A'_i(\xi)A'_j(\xi') - \delta_{ij}), \tag{A17}$$

where $\tilde{\epsilon}(\xi)$ is the periodized sign function defined in Eq. (2.11).

APPENDIX B: JUSTIFICATION FOR THE REDUCTION ALGORITHM

In this appendix we justify the introduction of the reduced-directrix parameters as generally defined in Sec. III [Eqs.(3.4) and (3.5)]. Assuming the original directrix to be given in the N-sector with parameters T, X, d_k we will show for the (first) reduced-directrix parameters T', X', d'_k that (i) they all commute (i.e., have vanishing PB's) with E and T, and (ii) they have an internal PB algebra, characteristic of the parameters of a real directrix in the $(N-1)$-sector. This justifies the first step in the reduction scheme, and thus, by induction, the complete algorithm.

We start by noting that, since

$$\begin{aligned} &\{E, d_k\} = 0, \quad \{E, X\} = 0, \\ &\{T, d_k\} = 1/N, \quad \{T, X\} = 0, \end{aligned} \tag{B1}$$

the first statement (i) follows from the fact that the reduced parameters can all be expressed as linear combinations of differences of the original directrix elements, except X' that also includes X.

For the second statement (ii), we note that the PB's between the original directrix elements are such that only nearest neighbors have nonvanishing PB's. Thus, when we subtract d_{k_0} from all elements, the PB between two elements, neither of which is nearest neighbor to d_{k_0}, will not change.

The only element of the reduced directrix that is related to the neighbors of d_{k_0} is

$$d'_{k_0-1} = d_{k_0-1} - 2d_{k_0} + d_{k_0+1}, \tag{B2}$$

which is easily seen to have the correct PB with the other reduced elements, and accordingly we have proved that the internal PB-algebra of the reduced elements d'_k comes out correct. To complete the proof, we must analyze the PB's involving X' and T'.

From the definitions [Eqs. (3.5a) and (3.5b)] of T' and $X - X'$ in terms of linear combinations of the elements d_k, it is straightforward to derive their PB's with d_k. The result is

$$\{T', d_k\} = \frac{1}{N(N-1)} + \frac{1}{N-1}\left[(N-k_0-1)\,\delta^P_{k-k_0+1} - \delta^P_{k-k_0} + (k_0-N)\,\delta^P_{k-k_0-1}\right], \tag{B3a}$$

$$\{X - X', d_k\} = \frac{-(-)^k}{N(N-1)} + \frac{(-)^{k_0}}{N-1}\left[-\frac{1}{2}\delta^P_{k-k_0+1} + \delta^P_{k-k_0} - \frac{1}{2}\delta^P_{k-k_0-1}\right]. \tag{B3b}$$

Using Eq. (2.16b) we obtain for X' the PB's

$$\{X', d_k\} = \frac{(-)^k}{N-1} + \frac{(-)^{k_0}}{N-1} \times \left[\frac{1}{2} \delta^P_{k-k_0+1} - \delta^P_{k-k_0} + \frac{1}{2} \delta^P_{k-k_0-1} \right]. \quad \text{(B4)}$$

Subtracting the PB's with d_{k_0}, we get

$$\{T', d_k - d_{k_0}\} = \frac{1}{N-1} + \frac{1}{N-1} \Big[(N - k_0 - 1)\delta^P_{k-k_0+1} - \delta^P_{k-k_0} + (k_0 - N)\, \delta^P_{k-k_0-1} \Big], \quad \text{(B5a)}$$

$$\{X', d_k - d_{k_0}\} = \frac{(-)^k}{N-1} + \frac{(-)^{k_0}}{N-1} \left[\frac{1}{2} \delta^P_{k-k_0+1} - \delta^P_{k-k_0} + \frac{1}{2} \delta^P_{k-k_0-1} \right]. \quad \text{(B5b)}$$

One readily realizes that this leads to the correct PB's with the elements of the reduced directrix

$$\{T', d'_k\} = 1/(N-1), \quad \text{(B6a)}$$

$$\{X', d'_k\} = (-)^k/(N-1). \quad \text{(B6b)}$$

Using Eq. (B5b) and the definition of T', one finally derives

$$\{X', T'\} = 0. \quad \text{(B7)}$$

We conclude this appendix by showing an important continuity property for X' and T'. Assume that either of two neighboring elements, d_{k_0} and d_{k_0+1} of the original N-sector directrix can be taken as the smallest element.

In case d_{k_0} is chosen, we get

$$T' = \frac{1}{4N(N-1)} \sum_{k=0}^{2N-1} (d_k - d_{k_0}) \times [2k - 2N + 1 - 2N\epsilon^P_{k-k_0}], \quad \text{(B8a)}$$

$$X - X' = \frac{1}{4N(N-1)} \sum_{k=0}^{2N-1} (-)^k (d_k - d_{k_0}) \times [2k - 2N + 1 - 2N\epsilon^P_{k-k_0}], \quad \text{(B8b)}$$

whereas if d_{k_0+1} is chosen, we obtain

$$T' = \frac{1}{4N(N-1)} \sum_{k=0}^{2N-1} (d_k - d_{k_0+1}) \times [2k - 2N + 1 - 2N\epsilon^P_{k-k_0-1}], \quad \text{(B9a)}$$

$$X - X' = \frac{1}{4N(N-1)} \sum_{k=0}^{2N-1} (-)^k (d_k - d_{k_0+1}) \times [2k - 2N + 1 - 2N\epsilon^P_{k-k_0-1}]. \quad \text{(B9b)}$$

Using that $d_{k_0} = d_{k_0+1}$ we obtain the difference between the two definitions

$$\Delta T' = \frac{2N}{4N(N-1)} \sum_{k=0}^{2N-1} (d_k - d_{k_0}) \times (\epsilon^P_{k-k_0-1} - \epsilon^P_{k-k_0}), \quad \text{(B10a)}$$

$$\Delta (X - X') = \frac{2N}{4N(N-1)} \sum_{k=0}^{2N-1} (-)^k (d_k - d_{k_0}) \times (\epsilon^P_{k-k_0-1} - \epsilon^P_{k-k_0}). \quad \text{(B10b)}$$

This can be rewritten as

$$\Delta T' = \frac{-1}{2(N-1)} \sum_{k=0}^{2N-1} (d_k - d_{k_0}) \times (\delta^P_{k-k_0-1} + \delta^P_{k-k_0}) \equiv 0, \quad \text{(B11a)}$$

$$\Delta (X - X') = \frac{-1}{2(N-1)} \sum_{k=0}^{2N-1} (-)^k (d_k - d_{k_0}) \times (\delta^P_{k-k_0-1} + \delta^P_{k-k_0}) \equiv 0. \quad \text{(B11b)}$$

Thus, since X does not depend on the choice of k_0, we have shown that X' and T' are the same in both cases.

In this case, if d_{k_0} is chosen as the smallest element, the element d_{k_0+1} disappears from the daughter directrix; it is absorbed in d'_{k_0-1}. However if the two small elements are not nearest neighbors, then if one of them is chosen, in the next step of the reduction algorithm the other one is reduced to zero and will be the smallest one. The corresponding action-variable J_{N-1} will then also be zero. In this case the two choices of k_0 will in general give different values for X' and T', and thus a nonuniqueness of the angle variable θ_{N-1}. The other angle variables will, however, be unaffected.

By a comparison with the harmonic oscillators, where the vanishing of the action variable makes the angle variable completely nonsignificant, one realizes that this is a quite normal behavior.

APPENDIX C: DERIVATION OF THE PERIODICITY PROPERTIES OF THE ANGLE VARIABLES

In this appendix, we justify our choice of angle variables by proving that they have the correct periodicity properties, i.e., that the configuration of the system, represented by the directrix, is periodic with unit period in each of the angle variables.

The elements of a given directrix can be labeled in different ways. Different labelings give different values for our variables corresponding to the same directrix, i.e., the same state of motion. We will show below that various relabelings will imply shifts of the angle variables by integer numbers. On the other hand, because the directrix is a continuous function of the angles (see Appendix B), a unit change of the angle variables by a continuous variation corresponds to a change of the directrix coming back to the same state, implying the periodic property.

The first kind of relabeling we consider is a shift in the labeling of the elements of the original directrix. Since we still want ξ_0 to correspond to a local minimum of the directrix, the smallest possible relabeling shift is by two units:

$$\xi_k \to \xi'_k = \xi_{k-2}, \qquad d_k \to d'_k = d_{k-2}. \quad \text{(C1)}$$

Then, because of their definition [Eqs. (2.18b) and (2.18c)], T_N and X_N will be shifted according to

$$T_N \to T'_N = T_N + (1/N)\, 2E_N, \quad \text{(C2a)}$$

$$X_N \to X'_N = X_N - (1/N)\, 2P. \quad \text{(C2b)}$$

Furthermore, since the labeling of the daughter directrix is related to the labeling of the original directrix in accordance with Eq. (3.4), it is seen from the definition [Eq. (3.5a)] that also the other T_k's will acquire shifts:

$$T_k \to T'_k = T_k + [1/k(k+1)]\,2E_k, \quad k<N. \tag{C3}$$

From the definition [Eq. (3.5b)] there will be a corresponding shift in the variables $X_{k+1} - X_k$:

$$\begin{aligned} X_{k+1} - X_k &\to X'_{k+1} - X'_k \\ &= X_{k+1} - X_k + [1/k(k+1)]\,2P. \end{aligned} \tag{C4}$$

The only X_k we are interested in is X_1, which will acquire the following shift:

$$X_1 \to X'_1 = X_1 - 2P. \tag{C5}$$

For the angle variables, as defined in Eqs. (3.18), we thus get the following shifts:

$$\theta_k \to \theta'_k = \theta_k + 1, \quad k = 1,...,N. \tag{C6}$$

All the other variables are left invariant by this operation.

The second kind of relabeling is related to the possibility of choosing another copy of the smallest element in one of the directrices, say the one belonging to the n-sector $(n>1)$:

$$k_0^{(n)} \to k_0^{(n)'} = k_0^{(n)} + 2n. \tag{C7}$$

From the definitions (3.5), and the properties of the periodical sign function [Eq. (2.11)], this implies the following shifts in T_{n-1} and $X_n - X_{n-1}$:

$$T_{n-1} \to T'_{n-1} = T_{n-1} + [1/(n-1)]\,2E_{n-1}, \tag{C8}$$

$$\begin{aligned} X_n - X_{n-1} &\to X'_n - X'_{n-1} \\ &= X_n - X_{n-1} + [1/(n-1)]\,2P. \end{aligned} \tag{C9}$$

Furthermore, a change of $k_0^{(n)}$ implies a shift in the numbering of all the lower-sector directrices [cf. Eq. (3.4)]. Accordingly, the T_k's with $k<n-1$ will require shifts:

$$T_k \to T'_k = T_k + [1/k(k+1)]\,2E_k, \quad k<n-1. \tag{C10}$$

Again, the same goes for $X_{k+1} - X_k$:

$$\begin{aligned} X_{k+1} - X_k &\to X'_{k+1} - X'_k \\ &= X_{k+1} - X_k + [1/k(k+1)]\,2P, \quad k<n-1. \end{aligned} \tag{C11}$$

For X_1, this implies the shift

$$X_1 \to X'_1 = X_1 - 2P. \tag{C12}$$

For the angle variables, we can derive the following shifts:

$$\theta_k \to \theta'_k = \begin{cases} \theta_k, & k \geqslant n, \\ \theta_k + 1, & k = 1,...,n-1. \end{cases} \tag{C13}$$

Again, all the other variables are left invariant by the operation. Thus, in both kinds of relabeling, only the angle variables are affected.

We note that the shifts derived from the relabelings discussed above constitute a complete basis of lattice vectors in $\bar\theta$-space with $\bar\theta$ the vector $\bar\theta = (\theta_1,...,\theta_N)$.

The lattice vectors are evidently

$$\begin{aligned} \bar\Delta_N &= (1,1,...,1,1), \\ \bar\Delta_{N-1} &= (1,1,...,1,0), \\ &\vdots \\ \bar\Delta_1 &= (1,0,...,0,0), \end{aligned} \tag{C14}$$

with the case given in Eq. (C13) corresponding to $\bar\Delta_n$, $n = 1,...,N-1$ and the case in Eq. (C6) to $\bar\Delta_N$. In case we only want to change the angular variable θ_k leaving the rest unaffected, we may use the shifts

$$\begin{aligned} \bar\Delta_N - \bar\Delta_{N-1} &= (0,...,0,1), \\ \bar\Delta_{N-1} - \bar\Delta_{N-2} &= (0,...,0,1,0), \\ &\vdots \\ \bar\Delta_2 - \bar\Delta_1 &= (0,1,0,...,0), \\ \bar\Delta_1 &= (1,0,...,0). \end{aligned} \tag{C15}$$

Since these lattice vectors correspond to different relabelings of a given directrix and its daughters, we have obviously shown that the configuration is periodic in each of the θ_k's with unit period.

[1]Y. Nambu, Lectures at the Copenhagen Symposium, 1970.

[2]S. Mandelstam, Phys. Rep. C **13**, 259 (1974).

[3]B. Andersson, G. Gustafson, and T. Sjöstrand, Z. Phys. C **6**, 235 (1980).

[4]P. Goddard, J. Goldstone, C. Rebbi, and C. Thorn, Nucl. Phys. B **56**, 109 (1972).

[5]A. Patrascioiu, Nucl. Phys. B **81**, 525 (1974). W. A. Bardeen, I. Bars, A. J. Hanson, and R. D. Peccei, Phys. Rev. D **13**, 2364 (1976).

[6]X. Artru, Phys. Rep. **97**, 147 (1983).

[7]P. A. M. Dirac, Proc. R. Soc. London Ser. A **246**, 333 (1958).

[8]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).

Mass eigenfunction expansions for the relativistic Kepler problem and arbitrary static magnetic field in relativistic quantum theory

Levere C. Hostler
Physics Department, Wilkes College, Wilkes-Barre, Pennsylvania 18766

(Received 6 February 1984; accepted for publication 20 July 1984)

We investigate the existence of orthogonality and completeness relations for the eigenvalue problem associated with the differential operator $\Lambda = -\Pi_\mu \Pi_\mu - ie\boldsymbol{\sigma}\cdot(\mathbf{E}+i\mathbf{B})$, $\Pi_\mu = -i\partial_\mu - eA_\mu$. The operator Λ acts on 2×1 Pauli-type spinor fields defined over all Minkowski space, and may be interpreted as the square of the mass of a charged Dirac particle moving in an external c-number electromagnetic field. We show that Λ is self-adjoint with respect to the not positive-definite inner product $(\phi_b;\phi_a) = \int d^4x\, \bar\phi_b \phi_a$, where $\bar\phi_b$ is defined as $\bar\phi_b = \phi_b^\dagger(-i\overleftarrow{\Pi}_4 - \boldsymbol{\sigma}\cdot\overleftarrow{\mathbf{\Pi}})$. A proof is provided for the Coulomb case that the mass eigenfunctions form a complete set in spite of the indefinite metric in Hilbert space. The mass eigenfunction expansion of the propagator is worked out explicitly for the Kepler case. This mass eigenfunction expansion is expected to be quite useful for bound state calculations in quantum electrodynamics, since it involves the covariant denominators $(m')^2 - (m)^2$.

I. INTRODUCTION

We investigate completeness and orthogonality relations for the eigenvalue problem associated with the differential operator

$$\Lambda = -\Pi_\mu \Pi_\mu - ie\boldsymbol{\sigma}\cdot(\mathbf{E}+i\mathbf{B}), \quad \Pi_\mu = -i\partial_\mu - eA_\mu, \tag{1.1}$$

acting on 2×1 Pauli-type spinor fields defined over all Minkowski space, concentrating mainly on the case in which A_μ is the four-potential of a static Coulomb field. We can view[1,2] Λ as the square of the mass of a charged Dirac particle interacting with the field A_μ.

In Sec. II A we show that Λ is self-adjoint with respect to the Lorentz invariant inner product

$$(\phi_b;\phi_a) = \int d^4x\, \bar\phi_b \phi_a \tag{1.2}$$

in Hilbert space, where the dual state $\bar\phi$ is defined as

$$\bar\phi \equiv \phi^\dagger(-i\overleftarrow{\Pi}_4 - \boldsymbol{\sigma}\cdot\overleftarrow{\mathbf{\Pi}}). \tag{1.3}$$

It is expected that a complete orthonormal basis of eigenfunctions of Λ will exist for an arbitrary external potential. However, the lack of positive definiteness of the inner product (1.2) prevents the straightforward application of the spectral theorem, which would otherwise guarantee this. That an orthonormal basis of eigenfunctions nevertheless exists is trivial for the special case of an arbitrary static magnetic field. The proof for a Coulomb field is given in Sec. II B 1. The proof involves relating the given eigenvalue problem with an indefinite metric to an associated eigenvalue problem with a positive-definite metric, for which the full power of the spectral theorem is at our disposal. An application is presented in Sec. II B 2; where a mass eigenfunction expansion of the Coulomb propagator is derived.

II. MASS EIGENFUNCTION EXPANSIONS

A. Self-adjointness of Λ with respect to the indefinite metric

To organize the proof of self-adjointness of Λ with respect to the indefinite metric (1.2) we introduce operators A and B through the defining equations

$$A = -i\Pi_4 + \boldsymbol{\sigma}\cdot\mathbf{\Pi} \tag{2.1}$$

and

$$B = -i\Pi_4 - \boldsymbol{\sigma}\cdot\mathbf{\Pi}. \tag{2.2}$$

All relevant operations can be performed in terms of these two operators. We note the operator identity

$$\Lambda = AB \tag{2.3}$$

and the relation

$$\bar\phi = \phi^\dagger \overleftarrow{B}, \tag{2.4}$$

which is just Eq. (1.3) in the new notation.

If Λ is to be self-adjoint with respect to the metric (1.2) we must have the identity $(\Lambda\phi_b;\phi_a) = (\phi_b;\Lambda\phi_a)$ for any two spinor fields $\phi_{b,a}$. Writing this identity out in terms of integrals, and using the defining equation (2.4) of the dual, we find the requisite condition in the form:

$$\int d^4x (\overrightarrow{A}\,\overrightarrow{B}\phi_b)^\dagger \overleftarrow{B}\phi_a = \int d^4x\, \phi_b^\dagger \overleftarrow{B}(\overrightarrow{A}\overrightarrow{B}\phi_a). \tag{2.5}$$

That the condition (2.5) holds for any two spinor fields $\phi_{b,a}$ is an immediate consequence of the self-adjointness of A and B with respect to the simple metric $\int d^4x\, \phi_b^\dagger \phi_a$. This establishes the self-adjointness of Λ with respect to the metric (1.2).

B. The relativistic Kepler problem

1. General theory

When written out explicitly for the case of a Coulomb potential, the operator Λ takes the form

$$\Lambda = \left(i\frac{\partial}{\partial t}\right)^2 + 2i\frac{\partial}{\partial t}\frac{Z\alpha}{r} + \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} - \frac{K^2 - K - (Z\alpha)^2 - iZ\alpha\sigma\cdot\hat{\mathbf{r}}}{r^2}, \tag{2.6}$$

$$K \equiv \boldsymbol{\sigma}\cdot\mathbf{L} + 1.$$

The identity

 0022-2488/85/010124-04$02.50

$$K^2 - K - (Z\alpha)^2 - iZ\alpha\sigma\cdot\hat{r} = S\{-\tfrac{1}{4}(1-(2\gamma+1)^2)\}S^{-1}, \tag{2.7}$$

in which

$$\gamma+\tfrac{1}{2} = (K^2-(Z\alpha)^2)^{1/2} - \tfrac{1}{2}\epsilon(K),$$
$$\epsilon(K) = \begin{cases} +1, & K>0, \\ -1, & K<0, \end{cases} \tag{2.8}$$

and

$$S = e^{i\sigma\cdot\hat{r}\theta/2}, \quad \theta = \tanh^{-1}(Z\alpha/K), \tag{2.9}$$

parallels earlier results of Biedenharn[3] and of Martin and Glauber,[4] who dealt with the conventional Dirac equation.

By use of the identity (2.7), the expression (2.6) for Λ can be written

$$\Lambda = SHS^{-1}, \tag{2.10}$$

where

$$H = \left(i\frac{\partial}{\partial t}\right)^2 + 2i\frac{\partial}{\partial t}\frac{Z\alpha}{r} + \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1-(2\gamma+1)^2}{4r^2}. \tag{2.11}$$

According to Eq. (2.10) there will be a one-to-one correspondence between solutions $\phi_{\Lambda'}$ of the eigenvalue equation $\Lambda\phi_{\Lambda'} = \Lambda'\phi_{\Lambda'}$ and solutions $\chi_{\Lambda'}$ of the eigenvalue equation $H\chi_{\Lambda} = \Lambda'\chi_{\Lambda'}$ with $\phi_{\Lambda'} = S\chi_{\Lambda'}$. But H is self-adjoint with respect to the simple positive definite inner product $\int d^4x\, \chi^\dagger_{\Lambda''}\chi_{\Lambda'}$ in Hilbert space. By the spectral theorem, the functions $\chi_{\Lambda'}$ can be chosen to form an orthonormal basis. Now let ϕ be an arbitrary spinor field, and let us expand $S^{-1}\phi$ as a linear superposition of the eigenfunctions $\chi_{\Lambda'}: S^{-1}\phi = \Sigma_{\Lambda'} c_{\Lambda'}\chi_{\Lambda'}$. Substituting $\chi_{\Lambda'} = S^{-1}\phi_{\Lambda'}$ and removing the common factor S^{-1} on both sides of the equation gives $\phi = \Sigma_{\Lambda'} c_{\Lambda'}\phi_{\Lambda'}$. Since ϕ is arbitrary, we have here a proof of completeness of the eigenfunctions $\phi_{\Lambda'}$.

Next we investigate the relationship between the dual states in the two eigenvalue problems. From the relation $H\chi_{\Lambda'} = \Lambda'\chi_{\Lambda'}$ we find first $\chi^\dagger_{\Lambda'}H = \Lambda'\chi^\dagger_{\Lambda'}$, then $\chi^\dagger_{\Lambda'}S^{-1}SHS^{-1} = \Lambda'\chi^\dagger_{\Lambda'}S^{-1}$, and then

$$\chi^\dagger_{\Lambda'}S^{-1}\Lambda = \Lambda'\chi^\dagger_{\Lambda'}S^{-1}. \tag{2.12}$$

On the other hand, taking the dual with respect to the indefinite metric of the eigenvalue equation $\Lambda\phi_{\Lambda'} = \Lambda'\phi_{\Lambda'}$ we get $(\overleftrightarrow{\Lambda}\,\overrightarrow{B}\phi_{\Lambda'})^\dagger\overleftarrow{B} = \Lambda'\phi^\dagger_{\Lambda'}\overleftarrow{B}$ or

$$\bar{\phi}_{\Lambda'}\Lambda = \Lambda'\bar{\phi}_{\Lambda'}. \tag{2.13}$$

By comparing the two equations (2.12) and (2.13); we see that a given $\bar{\phi}_{\Lambda'} = (\phi_{\Lambda'})^\dagger\overleftarrow{B}$ is proportional to $\chi^\dagger_{\Lambda'}S^{-1}$ if there is no degeneracy: $\bar{\phi}_{\Lambda'} = \mu_{\Lambda'}\chi^\dagger_{\Lambda'}S^{-1}$, for some proportionality constant $\mu_{\Lambda'}$. For the general case of degeneracy we have (a and b are degeneracy quantum numbers)

$$\bar{\phi}_{\Lambda'a} = \sum_b c_{ab}\chi^\dagger_{\Lambda'b}S^{-1}. \tag{2.14a}$$

We record here also the relation

$$\phi_{\Lambda'a} = S\chi_{\Lambda'a} \tag{2.14b}$$

defining the correspondence between the eigenfunctions of the two eigenvalue problems.

Next we investigate the constants c_{ab}. From Eq. (2.14a), and using the orthogonality and completeness relations for the functions $\chi_{\Lambda'}$, we find $c_{ab} = \int d^4x\, \chi^\dagger_{\Lambda'a}SBS\chi_{\Lambda'b}$. Since the operator SBS is self-adjoint with respect to the simple inner product $\int d^4x\, \chi^\dagger_{\Lambda'a}\chi_{\Lambda'b}$ it follows that the constants c_{ab} form a self-adjoint matrix C. Let U_{ab} be matrix elements of a unitary matrix U diagonalizing $C: (U^{-1}CU)_{ab} = \mu_{\Lambda'a}\delta_{ab}$. If we define new functions $\chi^T_{\Lambda'a} = \Sigma_b\chi_{\Lambda'b}U_{ba}$, and

$$\phi^T_{\Lambda'a} = S\chi^T_{\Lambda'a}; \tag{2.15a}$$

then we find

$$\bar{\phi}^T_{\Lambda'a} = \mu_{\Lambda'a}\chi^{T\dagger}_{\Lambda'a}S^{-1} \tag{2.15b}$$

also in the degenerate case. If our original states $\chi_{\Lambda'a}$ were normalized according to $\int d^4x\, \chi^\dagger_{\Lambda''a}\chi_{\Lambda'b} = \delta_{\Lambda''\Lambda'}\delta_{ab}$ then the functions $\chi^T_{\Lambda'a}$ likewise have this normalization. The normalization of the $\phi^T_{\Lambda'a}$ is [from Eqs. (2.15b) and (2.14b)]

$$\int d^4x\, \bar{\phi}^T_{\Lambda''b}\phi^T_{\Lambda'a} = \mu_{\Lambda'a}\delta_{\Lambda''\Lambda'}\delta_{ab}. \tag{2.16}$$

Of course the new functions ϕ^T can be rescaled so as to have a norm of magnitude unity:

$$\int d^4x\, \bar{\phi}^T_{\Lambda''b}\phi^T_{\Lambda'a} = \epsilon_{\Lambda'a}\delta_{\Lambda''\Lambda'}\delta_{ab},$$
$$\epsilon_{\Lambda'a} = \begin{cases} +1, & \mu_{\Lambda'a}>0, \\ -1, & \mu_{\Lambda'a}<0. \end{cases} \tag{2.17}$$

If our eigenfunctions are normalized according to Eq. (2.17), then the completeness relation for the eigenfunctions reads $\delta^4(2,1) = \Sigma_{\Lambda'a}\epsilon_{\Lambda'a}\phi(2)_{\Lambda'a}\bar{\phi}(1)_{\Lambda'a}$ or, in abstract operator notation

$$1 = \sum_{\Lambda'a}\epsilon_{\Lambda'a}|\phi_{\Lambda'a}\rangle\langle\overline{\phi_{\Lambda'a}}|, \tag{2.18}$$

where

$$\langle\overline{\phi_{\Lambda'a}}| \equiv \langle\phi_{\Lambda'a}|B.$$

2. Coulomb propagator

We have by now developed all the machinery necessary to write down a formal mass eigenfunction expansion of the Coulomb propagator. This propagator can be defined through the abstract operator statement

$$g = 1/(\Lambda - m^2). \tag{2.19}$$

The mass eigenfunction expansion of g emerges if one uses the completeness relation (2.18) in a usual way:

$$g = \sum_{\Lambda'a}\epsilon_{\Lambda'a}\frac{|\phi_{\Lambda'a}\rangle\langle\overline{\phi_{\Lambda'a}}|}{\Lambda' - m^2}. \tag{2.20}$$

The expression (2.20) remains mathematically well-defined when the parameter m takes any complex value not in the eigenvalue spectrum of Λ. To obtain the usual Feynman propagator, however, one specializes m to be the physical particle mass, less an infinitesimal imaginary part. Because of the covariant denominators in Eq. (2.20), the mass eigenfunction expansion of the propagator is expected to be quite useful in quantum electrodynamics calculations.

For practical applications we still require explicit expressions for the states $|\phi_{\Lambda'a}\rangle$ and $\langle\overline{\phi_{\Lambda'a}}|$. Because of the degeneracy of the hydrogen spectrum, the requisite states for Eq. (2.20) must be computed using the diagonalization procedure of Sec. II B 1. Acordingly, these expressions are a bit complicated, and are presented as reference material in the Appendix. Here we derive an alternate "short form" of the mass eigenfunction expansion, which seems to serve just as well as the "standard form" (2.20).

The short form exploits the completeness relation

$$1=\sum_{A'a}|\chi_{A'a}\rangle\langle\chi_{A'a}| \tag{2.21}$$

for the eigenfunctions of H. If we multiply on the left by S and on the right by S^{-1}, we obtain the identity

$$1=\sum_{A'a}S|\chi_{A'a}\rangle\langle\chi_{A'a}|S^{-1}. \tag{2.22}$$

The virtue of the identity (2.22) is that the states $S|\chi_{A'a}\rangle$ are right eigenfunctions of A, $AS|\chi_{A'a}\rangle = A'S|\chi_{A'a}\rangle$; and the states $\langle\chi_{A'a}|S^{-1}$ are left eigenfunctions of A, $\langle\chi_{A'a}|S^{-1}A = A'\langle\chi_{A'a}|S^{-1}$. The expansion (2.22) differs from that in Eq. (2.18) in that in Eq. (2.22) the bra $\langle\chi_{A'a}|S^{-1}$ is in general not the dual of the ket $S|\chi_{A'a}\rangle$. Nevertheless, the expansion (2.22) can still be used as (2.18) was, and will again yield a mass eigenfunction expansion: namely,

$$g=\sum_{A'a}\frac{S|\chi_{A'a}\rangle\langle\chi_{A'a}|S^{-1}}{A'-m^2}. \tag{2.23}$$

The expansion (2.23) has the advantage that the expressions for the functions $|\chi_{A'a}\rangle$ are relatively simple. These expressions are written in the form $\chi=(dE/2\pi)^{1/2}\times\exp(-iEt)R(r)Y_{lJM}(\hat{\mathbf{r}})$, where the radial functions $R(r)$ are found to be[5]

Continuum states

$$R=\left(\frac{dk}{2\pi}\right)^{1/2}|\Gamma(1+\gamma-i\nu)|e^{\pi\nu/2}r^{-1}\mathcal{M}_{i\nu;\gamma+(1/2)}(-2ikr),$$
$$\nu=EZ\alpha/k,\quad -\infty<E<+\infty,\quad 0<k<\infty,$$
$$\Lambda=E^2-k^2<E^2; \tag{2.24}$$

Bound states

$$R=\left\{\frac{(2\eta)^3}{2(\gamma+n)}\frac{(n-1)!}{(2\gamma+n)!}\right\}^{1/2}(2\eta r)^\gamma e^{-\eta r}L_{n-1}^{2\gamma+1}(2\eta r), \tag{2.25}$$

$$0<E<\infty,\quad n=1,2,3,...,$$

$$\eta=\frac{EZ\alpha}{(\gamma+n)},\quad \Lambda=E^2\left\{1+\frac{(Z\alpha)^2}{(\gamma+n)^2}\right\}>E^2. \tag{2.26}$$

By describing the continuum states above in a discrete approximation, we are able to write down a uniform normalization condition

$$\int d^4x\,\chi_A^\dagger\chi_B=\delta_{AB}, \tag{2.27}$$

in which A and B are short-hand notations for the whole set of quantum numbers E, k, l, J, M (continuum states) or E, n, l, J, M (bound states), as appropriate.

APPENDIX: DETAILS OF MASS EIGENFUNCTIONS

Here we obtain the functions ϕ needed for the mass eigenfunction expansion in standard form [Eq. (2.20)]. For this purpose we require the functions χ^T and $\phi^T=S\chi^T$ for which $\bar{\phi}^T=\mu\chi^{T\dagger}S^{-1}$. We start by forming $\bar{\phi}_A S=(\overline{S\chi_A})S=\chi_A^\dagger SBS$; since this is expected to be a simple linear superposition of degenerate levels $\chi_A^\dagger$ in accordance with Eq. (2.14a). The operator SBS has been worked out in Ref. 2, Appendix A, and has been found to be

$$SBS=\frac{-E|K|}{(K^2-(Z\alpha)^2)^{1/2}}-i\left(\frac{\partial}{\partial r}+\frac{1+\epsilon(K)(K^2-(Z\alpha)^2)^{1/2}}{r}-EZ\alpha\frac{\epsilon(K)}{(K^2-(Z\alpha)^2)^{1/2}}\right)\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}. \tag{A1}$$

Corresponding to Eq. (2.14a), we find[6]

Continuum states

$$\begin{bmatrix}\bar{\phi}_{J-(1/2)}\\ \bar{\phi}_{J+(1/2)}\end{bmatrix}=\begin{bmatrix}E|K|/\gamma & k|\gamma-i\nu|/\gamma\\ k|\gamma-i\nu|/\gamma & E|K|/\gamma\end{bmatrix}\begin{bmatrix}\chi^\dagger_{J-(1/2)}S^{-1}\\ \chi^\dagger_{J+(1/2)}S^{-1}\end{bmatrix}; \tag{A2}$$

Bound states

$$\begin{bmatrix}\bar{\phi}_{nJ-(1/2)}\\ \bar{\phi}_{n-1\,J+(1/2)}\end{bmatrix}=\begin{bmatrix}\dfrac{E|K|}{\gamma} & \dfrac{i\eta((n-1)(n-1+2\gamma))^{1/2}}{\gamma}\\ \dfrac{-i\eta((n-1)(n-1+2\gamma))^{1/2}}{\gamma} & \dfrac{E|K|}{\gamma}\end{bmatrix}\begin{bmatrix}\chi^\dagger_{nJ-(1/2)}S^{-1}\\ \chi^\dagger_{n-1\,J+(1/2)}S^{-1}\end{bmatrix}. \tag{A3}$$

In Eqs. (A2) and (A3) $\gamma=(K^2-(Z\alpha)^2)^{1/2}$. The functions χ appearing here are the functions of Eqs. (2.24) and (2.25), but only as many quantum numbers are shown explicitly as are needed to indicate which state is intended. Next we form linear combinations of the states $\chi_{J-(1/2)}$ and $\chi_{J+(1/2)}$ that serve to diagonalize the coefficient matrices in Eqs. (A2) and (A3). These linear combinations are

Continuum states

$$\chi_1^T=(\chi_{J-(1/2)}-\chi_{J+(1/2)})/(2)^{1/2},$$
$$\chi_2^T=(\chi_{J-(1/2)}+\chi_{J+(1/2)})/(2)^{1/2}; \tag{A4}$$

Bound states

$$\chi_1^T=(\chi_{nJ-(1/2)}-i\chi_{n-1J+(1/2)})/(2)^{1/2},$$
$$\chi_2^T=(-i\chi_{nJ-(1/2)}+\chi_{n-1J+(1/2)})/(2)^{1/2}. \tag{A5}$$

The mass eigenfunctions $\phi^T=S\chi^T/(|\mu|)^{1/2}$ will have the property that $\bar{\phi}^T=\mu\chi^{T\dagger}S^{-1}/(|\mu|)^{1/2}$, in which the constant μ has the value for

Continuum states

$$\mu_1=(E|K|-k|\gamma-i\nu|)/\gamma,$$
$$\mu_2=(E|K|+k|\gamma-i\nu|)/\gamma, \tag{A6}$$

Bound states

$$\mu_1 = (E|K| + \eta((n-1)(n-1+2\gamma))^{1/2})/\gamma,$$
$$\mu_2 = (E|K| - \eta((n-1)(n-1+2\gamma))^{1/2})/\gamma. \qquad \text{(A7)}$$

The normalization of the mass eigenfunctions has the requisite form [corresponding to Eq. (2.17)]

$$\int d^4x\, \bar{\phi}_A^T \phi_B^T = \epsilon_A \delta_{AB}, \qquad \text{(A8)}$$

in which $\epsilon_A = \pm 1 = \mu_A/|\mu_A|$. The functions ϕ_A^T defined here are the functions that we are looking for. All states have positive norm except for certain virtual states, the continuum states of type ϕ_1^T. These virtual states are off-shell intermediate states in the sense of Feynman–Dyson perturbation theory.

[1]L. Hostler, J. Math. Phys. **23**, 1179 (1982).
[2]L. Hostler, J. Math. Phys. **24**, 2366 (1983).
[3]L. C. Biedenharn, Phys. Rev. **126**, 845 (1962).
[4]P. C. Martin and R. J. Glauber, Phys. Rev. **109**, 1307 (1958).

[5]The functions M are Whittaker functions, and the functions $L_{n-1}^{2\gamma+1}$, are Laguerre polynomials. As in Refs. 1 and 2 we adopt the definitions of H. Bucholz, *Die Konfluente Hypergeometrische Function* (Springer-Verlag, Berlin, 1953).

[6]The calculation of the dual for bound states goes essentially as in Ref. 2, Appendix A. The calculation of the dual for continuum states is quite similar, but requires the identity [Ref. 5, p. 82, Eq. (42b)]

$$(1 \pm \mu)\frac{d}{dz}\mathscr{M}_{\varkappa;(\mu/2)} = \left[\frac{(1 \pm \mu)^2}{2z} - \varkappa\right]\mathscr{M}_{\varkappa;(\mu/2)} - \begin{cases} [\varkappa^2 - ((1+\mu)/2)^2]\ \mathscr{M}_{\varkappa;1+(\mu/2)} & \text{(upper sign)} \\ \mathscr{M}_{\varkappa;-1+(\mu/2)} & \text{(lower sign).} \end{cases}$$

A general approach to the systematic derivation of SO(3) shift operator relations. I. Theory

H. De Meyer[a)] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit-Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 27 February 1984; accepted for publication 25 May 1984)

A new technique is established for the construction of relations between products consisting of two, three, or more SO(3) shift operators, within the framework of the Lie algebras of SO(3) tensor operators constructed out of an underlying single-boson structure.

I. INTRODUCTION

The classification and analysis of irreducible unitary representations of various groups possessing an SO(3) subgroup is often a problem of physical interest. Indeed, in many systems endowed with a group symmetry, particularly the ones encountered in the field of nuclear physics, the group representation states should be labeled such that they are eigenstates of an angular momentum operator L^2 and a projection operator l_0. In order to carry out a systematic search for a complete set of intercommuting operators constructed out of the enveloping algebra of the group generators, which then should provide the so-called missing labels, SO(3) shift operators have been proven extremely useful.[1–5] Algebraically they can be defined by means of their commutation properties with the operators L^2 and l_0, which express that they raise or lower by one or more units the value of l, where $l(l+1)$ is the eigenvalue of the SO(3) Casimir L^2, and that they leave m, the eigenvalue of l_0 unchanged.[1] Taking into account that the algebra of group generators decomposes into its SO(3) subalgebra and a set of SO(3) tensor representations, the simplest kind of shift operators are the ones linear with respect to the tensor components and of appropriate degree with respect to the SO(3) generators $l_0, l_\pm$. By clarifying the strong relationship between matrix elements of these shift operators on the one hand and reduced matrix elements of the SO(3) tensor operators on the other hand, Hughes and Yadegar[6] succeeded to establish a very general formula which enables the explicit construction of these shift operators. This construction avoids every reference to the commutator properties of the tensor operators, the only algebraical element involved being Wigner's $3j$-symbol. Amongst the set of shift operators, the SO(3) scalars, in other words the ones which leave l unchanged, are the most interesting since they provide a natural basis for the construction of the missing label generating operators. Hence, one needs to derive the eigenvalues of these scalar shift operators.

To that aim one has to establish a sufficient number of independent relations between products of shift operators. So far, the only technique adopted consists in writing the products of shift operators in a so-called standard form with respect to the group generators.[4] Even for low-dimensional groups it turns out that this procedure involves tedious and time consuming calculations.

In the present paper we discuss a new technique to derive such relations between shift operator products without making use of the explicit forms of the shift operators in terms of generators.

II. SO(3) TENSOR AND SHIFT OPERATORS

Let $T(j,\mu)$, $\mu=-j,\ldots,j$ and j integral be a $(2j+1)$-dimensional tensor representation of SO(3) whose commutators with the SO(3) generators $l_0, l_\pm$ are

$$[l_\pm, T(j,\mu)] = [(j\mp\mu)(j\pm\mu+1)]^{1/2}T(j,\mu\pm 1), \tag{2.1}$$

$$[l_0, T(j,\mu)] = \mu T(j,\mu).$$

It is well known[7,8] that such SO(3) tensor operator can be realized as a linear operator acting in a $(2b+1)$-dimensional space with angular momentum basis $\{|bm\rangle, m=-b,\ldots,b\}$ as follows:

$$T(j,\mu)|bm\rangle = \sum_{m'}(-1)^{b-m'}(2j+1)^{1/2} \times\begin{pmatrix} b & j & b\\ -m' & \mu & m\end{pmatrix}|bm'\rangle. \tag{2.2}$$

In (2.2) the only restriction on b so far is $2b\geqslant j$. Moreover, it is a well-established property that the generators of any simple Lie algebra can be realized as a set of tensor operators of the type (2.1) with respect to a particular SO(3) subalgebra. The SO(3) subalgebra generators $l_0, l_\pm$ are themselves proportional to the components of a tensor operator $T(1,\mu)$ of rank one.

It is evident that in general the tensor operators which span a particular Lie algebra, cannot all be realized in the form (2.2) with a single fixed b-value. Only the Lie algebra's having at least one representation which decomposes with respect to SO(3) into a single SO(3) irrep b, and for which the Kronecker product with itself contains the adjoint representation of the algebra, will be considered further. Moreover, we restrict ourselves to integral b-values. Inspection of branching rule tables[9] makes it clear that therefore the Lie algebra's A_{2n}, B_n, and G_2 fall into the scope of the present study.

Also the generators satisfy[7]

$$[T(j_1,\mu_1),T(j_2,\mu_2)] = \sum_{j_3\mu_3}(-1)^{\mu_3}[(-1)^{j_1+j_2+j_3}-1] \times[(2j_1+1)(2j_2+1)(2j_3+1)]^{1/2}$$

[a)] Senior Research Associate (N.F.W.O.) Belgium.

 0022-2488/85/010128-05$02.50

$$\times \begin{Bmatrix} j_1 & j_2 & j_3 \\ b & b & b \end{Bmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & -\mu_3 \end{pmatrix} \times T(j_3, \mu_3) . \tag{2.3}$$

As a consequence of (2.3) the tensor operator $T(1,\mu)$ is related by the formula

$$T(1,\mu) = [6/b(2b+1)(2b+2)]^{1/2}\, l_\mu \quad (\mu = 0, \pm 1), \tag{2.4}$$

to the spherical components $l_0, l_{\pm 1} = \mp l_\pm/\sqrt{2}$ of the angular momentum operator. Furthermore, we shall suppose the $T(j,\mu)$ act upon states $|\gamma,l,m\rangle$, where $l(l+1)$ and m are the eigenvalues of, respectively, the SO(3) Casimir $L^2 = l_+ l_- + l_0^2 - l_0$ and l_0, and γ denotes an additional collection of labels needed to completely specify the states. According to the Wigner–Eckart theorem[7,8]

$$\langle \gamma' l' m' | T(j,\mu) | \gamma l m \rangle = (-1)^{l'-m'} \begin{pmatrix} l' & j & l \\ -m' & \mu & m \end{pmatrix} \langle \gamma' l' \| T(j) \| \gamma l \rangle , \tag{2.5}$$

the matrix elements of $T(j,\mu)$ are expressed in terms of reduced matrix elements. For further use we introduce the shorthand notation

$$\langle \gamma' l' \| j \| \gamma l \rangle \equiv \langle \gamma' l' \| T(j) \| \gamma l \rangle . \tag{2.6}$$

Some years ago Hughes and Yadegar[6] introduced so-called SO(3) shift operators O_l^k which are constructed in terms of the tensor operator components $T(j,\mu)$ and the SO(3) generators, and which are defined on account of their action upon $|\gamma l m\rangle$ states, i.e.,

$$O_l^k |\gamma l m\rangle = \sum_{\gamma'} \gamma' l + k m | O_l^k | \gamma l m \rangle |\gamma' + l + k m\rangle , \tag{2.7}$$

showing that they shift l by an amount k, and leave m unchanged. By a general analysis, Hughes and Yadegar proved[6] that in case O_l^k is linear with respect to the tensor operator components, its nonzero matrix elements on the right-hand side of (2.7) are related to the reduced matrix elements of $T(j)$ in the following way:

$$\begin{aligned}
\langle \gamma' l + k m | O_l^k | \gamma l m \rangle &= \left[\frac{(l+m+k)!(l-m+k)!}{(l+m)!(l-m)!} \right]^{1/2} \times \langle \gamma' l + k m | A_l^k | \gamma l m \rangle \quad (k > 0), \\
\langle \gamma' l - k m | O_l^{-k} | \gamma l m \rangle &= (-1)^k \left[\frac{(l+m)!(l-m)!}{(l+m-k)!(l-m-k)!} \right]^{1/2} \times \langle \gamma' l - k m | A_l^{-k} | \gamma l m \rangle \quad (k > 0), \\
\langle \gamma' l + k m | A_l^k | \gamma l m \rangle &= \left[\frac{(j+k)!(j-k)!(2l+j+k+1)!}{(2j)!(2l-j+k)!(2l+2k+1)^2} \right]^{1/2} \times \langle \gamma' l + k \| j \| \gamma l \rangle \quad (k = -j, -j+1, \ldots, j) .
\end{aligned} \tag{2.8}$$

The explicit expressions for these shift operators O_l^k in terms of the algebra generators which are obtained by a general formula, permit the derivation of the matrix elements (2.8) in a step-by-step procedure. The essential point however is that a set of relations between products of shift operators must be constructed first. Up to now the only way to obtain them is to rely on the shift operator expressions in terms of the generators and to exploit the commutation properties amongst the latter. This method has been discussed in detail elsewhere.[2–5]

As has been mentioned in the Introduction, the aim of the present paper is to establish these relations in a direct way without referring to the generator forms of the shift operators.

III. RELATIONS BETWEEN SHIFT OPERATORS QUADRATIC IN THE $T(j,\mu)$

Let us consider the reduced matrix elements between SO(3) representation states of the tensor product of two tensor operators (2.1) and let us develop these in the usual way in terms of reduced matrix elements of the tensors separately[10]:

$$\begin{aligned}
\langle \gamma' l' \| j_1 j_2 ; k \| \gamma l \rangle &\equiv \langle \gamma' l' \| (T(j_1) \times T(j_2))^k \| \gamma l \rangle \\
&= \sum_{\gamma'' l''} (-1)^{l+l'+k} (2k+1)^{1/2} \begin{Bmatrix} j_1 & j_2 & k \\ l & l' & l'' \end{Bmatrix} \langle \gamma' l' \| j_1 \| \gamma'' l'' \rangle \langle \gamma'' l'' \| j_2 \| \gamma l \rangle .
\end{aligned} \tag{3.1}$$

Let us, moreover, also consider the coupled commutator

$$\begin{aligned}
[T(j_1), T(j_2)]_\mu^k &= \sum_{\mu_1 \mu_2} \langle j_1 \mu_1 j_2 \mu_2 | k \mu \rangle [T(j_1, \mu_1), T(j_2, \mu_2)] \\
&= (T(j_1) \times T(j_2))_\mu^k - (-1)^{k+j_1+j_2} (T(j_2) \times T(j_1))_\mu^k .
\end{aligned} \tag{3.2}$$

Due to (2.3) it follows that

$$[T(j_1), T(j_2)]_\mu^k = (-1)^{j_2 - j_1} [(-1)^{j_1+j_2+k} - 1] [(2j_1+1)(2j_2+1)]^{1/2} \begin{Bmatrix} j_1 & j_2 & k \\ b & b & b \end{Bmatrix} T(k,\mu) . \tag{3.3}$$

Combining (3.1) — (3.3) one obtains a set of equations relating a sum of quadratic products of reduced matrix elements on the one hand to a single reduced matrix element on the other hand, i.e.,

$$\sum_{\gamma''l''} (-1)^{l+l''+k}(2k+1)^{1/2}\Big[\begin{Bmatrix} j_1 & j_2 & k \\ l & l' & l'' \end{Bmatrix} \langle\gamma'l'\|j_1\|\gamma''l''\rangle\langle\gamma''l''\|j_2\|\gamma l\rangle$$
$$-(-1)^{j_1+j_2+k}\begin{Bmatrix} j_2 & j_1 & k \\ l & l' & l'' \end{Bmatrix} \langle\gamma'l'\|j_2\|\gamma''l''\rangle\langle\gamma''l''\|j_1\|\gamma l\rangle\Big]$$
$$=(-1)^{j_2-j_1}[(-1)^{j_1+j_2+k}-1][(2j_1+1)(2j_2+1)]^{1/2}\begin{Bmatrix} j_1 & j_2 & k \\ b & b & b \end{Bmatrix}\langle\gamma'l'\|k\|\gamma l\rangle . \tag{3.4}$$

Equation (3.4) complemented with the definitions (3.1) for the reduced matrix elements $\langle\gamma'l'\|j_1j_2;k\|\gamma l\rangle$ and $\langle\gamma'l'\|j_2j_1;k\|\gamma l\rangle$ form a complete system of linear equations with respect to

$$\sum_{\gamma''}\langle\gamma'l'\|j_1\|\gamma''l''\rangle\langle\gamma''l''\|j_2\|\gamma l\rangle$$

and

$$\sum_{\gamma''}\langle\gamma'l'\|j_2\|\gamma''l''\rangle\langle\gamma''l''\|j_1\|\gamma l\rangle$$

for all possible l''-values.

If we restrict our attention to the case $j_1=j_2=j$ which is frequently encountered in practice, the system of the above-mentioned equations can be solved in closed form on account of the orthogonality relation for $6j$-symbols,[10] yielding

$$\sum_{\gamma''}\langle\gamma'l'\|j\|\gamma''l''\rangle\langle\gamma''l''\|j\|\gamma l\rangle$$
$$=\frac{1}{2}(-1)^{l+l'}(2l''+1)\sum_k(2k+1)^{1/2}\begin{Bmatrix} j & j & k \\ l & l' & l'' \end{Bmatrix}\Big[(1-(-1)^k)(2j+1)$$
$$\times\begin{Bmatrix} j & j & k \\ b & b & b \end{Bmatrix}\langle\gamma'l'\|k\|\gamma l\rangle+(-1)^k(1+(-1)^k)\langle\gamma'l'\|jj;k\|\gamma l\rangle\Big] . \tag{3.5}$$

Let us associate according to the relationships (2.7) and (2.8), to a reduced matrix element $\langle\gamma'l'\|j\|\gamma l\rangle$ the shift operator $O_l^{l'-l}$ and to the reduced matrix element $\langle\gamma'l'\|jj;k\|\gamma l\rangle$ the shift operator $Q_l^{k;l'-l}$. Note that the latter type of shift operators is quadratic in the $T(j)$ tensor components. In terms of these shifts operators (3.5) is rewritten, if $l'\leqslant l$, as

$$f(l,l',l'',m)O_{l''}^{l'-l''}O_l^{l''-l}=\frac{1}{2}(-1)^{l+l'}\sum_k\frac{[(2k+1)!]^{1/2}}{(2j)!}$$
$$\times\frac{(j+l'-l'')!(j-l'+l'')!(j+l''-l)!(j-l''+l)!}{\Delta(jjk)\Delta(jl'l'')\Delta(ljl'')\Delta(ll'k)(l+l'+k+1)!}\begin{Bmatrix} j & j & k \\ l & l' & l'' \end{Bmatrix}$$
$$\times\Big[(1-(-1)^k)(2j+1)\begin{Bmatrix} j & j & k \\ b & b & b \end{Bmatrix}O_l^{l'-l}+(1+(-1)^k)Q_l^{k;l'-l}\Big]\quad(l'\leqslant l), \tag{3.6}$$

where

$$f(l,l',l'',m)=1\quad\text{if}\quad l'\leqslant l''\leqslant l,$$
$$=(-1)^{l+l''}\frac{(l+m)!(l-m)!}{(l''+m)!(l''-m)!}\quad\text{if}\quad l'\leqslant l\leqslant l'',$$
$$=(-1)^{l'+l''}\frac{(l''+m)!(l''-m)!}{(l'+m)!(l'-m)!}\quad\text{if}\quad l''\leqslant l'\leqslant l, \tag{3.7}$$

and

$$\Delta(abc)=[(a+b-c)!(a-b+c)!(-a+b+c)!/(a+b+c+1)!]^{1/2}. \tag{3.8}$$

For $l'>l$ similar results hold on account of the formal identities $O_l^{-p}=O_{-l-1}^{p}$ $(p>0)$ and $Q_l^{k;-p}=Q_{-l-1}^{k;p}$ $(p>0)$. Shift operator relations of the type (3.6) have been studied in the past for various algebras, such as SU(3),[1-3] SO(5),[4,5] SO(7),[3-5] G_2,[3-5] etc. Nevertheless, since only very recently for the first time SO(3) shift operators of higher degree than one in the tensor components have been explicitly introduced[3] in the SU(3) enveloping algebra, most of the relations encountered in the literature do not involve $Q_l^{k;l'-l}$ operators, except for $Q_l^{0;0}$ which is clearly related to the second-order Casimir of the considered algebra.

Finally it is worthwhile to notice that the expressions $\Delta(abc)$ defined in (3.8), which are incorporated in the relations (3.6), exactly cancel the Δ-type factors which occur in Racah's formula for the $6j$-symbol.[10,11] It follows that (3.6) does not contain any irrational factor which is l-, l'-, or l''-dependent.

IV. RELATIONS BETWEEN SHIFT OPERATORS OF HIGHER DEGREE IN THE $T(j,\mu)$

The method of constructing relations between products of shift operators or equivalently of reduced matrix elements as described in the previous section can be straightforwardly extended at the level of any higher degree in the tensor components

$T(j,\mu)$. As an example we shall present here the third-order case. In analogy with (3.1) we define

$$\langle\gamma l'\|j_1j_2;kj_3;K\|\gamma l\rangle\equiv\langle\gamma l'\|((T(j_1)\times T(j_2))^k\times T(j_3))^K\|\gamma l\rangle$$
$$=\sum_{\gamma''l''}(-1)^{l+l'+K}(2K+1)^{1/2}\begin{Bmatrix}k & j_3 & K\\ l & l' & l''\end{Bmatrix}\langle\gamma l'\|j_1j_2;k\|\gamma''l''\rangle\langle\gamma''l''\|j_3\|\gamma l\rangle\,. \tag{4.1}$$

Next, we consider a coupled commutator of the form

$$[(T(j_1)\times T(j_2))^k,T(j_3)]^K_M=\sum_{\mu\mu_3}\langle k\mu j_3\mu_3|KM\rangle\,[(T(j_1)\times T(j_2))^k_\mu,T(j_3\mu_3)]$$
$$=((T(j_1)\times T(j_2))^kT(j_3))^K_M-(-1)^{k+j_3+K}(T(j_3)\times(T(j_1)\times T(j_2))^k)^K_M\,, \tag{4.2}$$

which on account of (2.3) equals

$$[(T(j_1)\times T(j_2))^k,T(j_3)]^K_M=\sum_{k_3}[(2j_3+1)(2k+1)(2k_3+1)]^{1/2}\{(-1)^{j_1+K}[(-1)^{j_2+j_3+k_3}-1]$$
$$\times(2j_2+1)^{1/2}\begin{Bmatrix}j_2 & j_3 & k_3\\ b & b & b\end{Bmatrix}\begin{Bmatrix}k & j_3 & K\\ k_3 & j_1 & j_2\end{Bmatrix}(T(j_1)\times T(k_3))^K_M+(-1)^{j_1+j_2+k_3-k}$$
$$\times[(-1)^{j_1+j_3+k_3}-1]\,(2j_1+1)^{1/2}\begin{Bmatrix}j_1 & j_3 & k_3\\ b & b & b\end{Bmatrix}\begin{Bmatrix}k & j_3 & K\\ k_3 & j_2 & j_1\end{Bmatrix}(T(k_3)\times T(j_2))^K_M\Big\}\,. \tag{4.3}$$

The combination of (4.1)–(4.3) leads to the following set of equations between reduced matrix elements:

$$\sum_{\gamma''l''}(-1)^{l+l'+K}(2K+1)^{1/2}\Big[\begin{Bmatrix}k & j_3 & K\\ l & l' & l''\end{Bmatrix}\langle\gamma l'\|j_1j_2;k\|\gamma''l''\rangle\langle\gamma''l''\|j_3\|\gamma l\rangle$$
$$-(-1)^{k+j_3+K}\begin{Bmatrix}j_3 & k & K\\ l & l' & l''\end{Bmatrix}\langle\gamma l'\|j_3\|\gamma''l''\rangle\langle\gamma''l''\|j_1j_2;k\|\gamma l\rangle\Big]$$
$$=\sum_{k_3}[(2j_3+1)(2k+1)(2k_3+1)]^{1/2}\Big\{(-1)^{j_1+K}[(-1)^{j_2+j_3+k_3}-1](2j_2+1)^{1/2}$$
$$\times\begin{Bmatrix}j_2 & j_3 & k_3\\ b & b & b\end{Bmatrix}\begin{Bmatrix}k & j_3 & K\\ k_3 & j_1 & j_2\end{Bmatrix}\langle\gamma l'\|j_1\,k_3;K\|\gamma l\rangle+(-1)^{j_1+j_2+k_3-k}[(-1)^{j_1+j_3+k_3}-1]\,(2j_1+1)^{1/2}$$
$$\times\begin{Bmatrix}j_1 & j_3 & k_3\\ b & b & b\end{Bmatrix}\begin{Bmatrix}k & j_3 & K\\ k_3 & j_2 & j_1\end{Bmatrix}\langle\gamma l'\|\,k_3\,j_2;K\|\gamma l\rangle\Big\}\,. \tag{4.4}$$

Notice that when k_3 is equal to one, the occurring elements of the type (3.1) can be further evaluated by means of the relationship (2.4), i.e.,

$$\langle\gamma l'\|1j;k\|\gamma l\rangle=(-1)^{l+l'+k}[3(2k+1)l'(l'+1)(2l'+1)/b(b+1)(2b+1)]^{1/2}\begin{Bmatrix}1 & j & k\\ l & l' & l'\end{Bmatrix}\langle\gamma l'\|j\|\gamma l\rangle\,, \tag{4.5}$$

and

$$\langle\gamma l'\|j1;k\|\gamma l\rangle=(-1)^{l+l'+k}[3(2k+1)l(l+1)(2l+1)/b(b+1)(2b+1)]^{1/2}\begin{Bmatrix}j & 1 & k\\ l & l' & l\end{Bmatrix}\langle\gamma l'\|j\|\gamma l\rangle\,. \tag{4.6}$$

Equations (4.4) together with the definitions (4.1) allow us to express every product consisting of a reduced matrix element of the type (3.1) multiplied with a reduced matrix element of the type (2.6) or vice versa, in terms of elements of the types (4.1), (3.1), and (2.6).

Obviously, all reduced matrix elements of the type (4.1) are not independent. In order to point out how they are connected one can derive a set of additional equations containing the before-mentioned products of reduced matrix elements in the following way. Multiply (3.5) either on the left or on the right with an appropriate reduced matrix element of the type (2.6) such that two different expressions are obtained for the same combination of three matrix elements. Equating these expressions results in an additional relation.

Clearly, it is again possible to associate to the new type (4.1) of reduced matrix elements according to (2.7)–(2.8) shift operators which will be used in a forthcoming paper with $j_1=j_2=j_3=j$ and will be denoted $R_l^{k;K;l'-l}$. The translation of the formulas in this section into shift operator language is not of special interest here.

V. DISCUSSION

Relations between shift operator products or equivalently between products of reduced matrix elements, have been derived here by a newly introduced technique. They are commonly valid for all algebras of tensor operators constructed out of a single boson structure (same b-value). Closed analytic forms of the Racah-coefficients which emerge in these relations are either found in the literature for low values of the entries, or can be obtained by means of an appropriate computer code.

[1]J. W. B. Hughes, J. Phys. A: Math. Nucl. Gen. **6**, 48 (1973); **6**, 281 (1973).
[2]H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981); **22**, 2366 (1981).
[3]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1025 (1983); **22**, 2332 (1981); **23**, 1223 (1982); **24**, 1339 (1983); **24**, 1345 (1983).
[4]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1902 (1980); **21**, 1913 (1980); **21**, 1973 (1980); **21**, 2635 (1980).
[5]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980); **21**, 1913 (1980); **21**, 1973 (1980), **21**, 2635 (1980).
[6]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[7]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[8]L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory* (Addison-Wesley, Reading, MA, 1981).
[9]W. G. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[10]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[11]L. C. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965).

A general approach to the systematic derivation of SO(3) shift operator relations. II. Applications

G. Vanden Berghe and H. De Meyer[a)]

Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit-Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 27 February 1984; accepted for publication 25 May 1984)

The technique reported on in the preceding paper is applied to construct shift operator relations in the SU(3) and SO(5) Lie algebras. Comments are made concerning the integrity basis for SO(3) scalar operators appearing in their respective enveloping algebras.

I. INTRODUCTION

In a preceding paper[1] (to be referred to as I) we have shown how the relationship which exists between matrix elements of SO(3) shift operators and reduced matrix elements of SO(3) tensor operators, can be exploited to set up in a general way relations between products of these shift operators. Such relations are relevant for the derivation of spectra of SO(3) scalar operators which may be used for solving state labeling problems. These problems and more in particular the shift operator approach to them have been thoroughly investigated.

One of the first examples studied in this way is the SU(3) Lie algebra in an SO(3) basis.[2,3] The reported relations are of second and third degree in the shift operators and hence also of same degree in the tensor operator. Moreover, they are all of SO(3) scalar type, meaning that they leave, when acting upon SO(3) basis states, the angular momentum l invariant. The extension to nonscalar relations[4,5] and to shift operators of second degree in the tensor components[6] led to an elegant method for deriving the eigenvalues and eigenstates of the two independent scalar shift operators for various irreducible SU(3) representations.

With a view to obtaining an orthogonal solution to the state labeling problem of the nuclear quadrupole-phonon states, relations between SO(3) shift operator products of scalar and nonscalar type in the SO(5) enveloping algebra have been constructed.[7] These allowed to diagonalize the SO(3) scalar shift operator in large parts of the totally symmetric SO(5) irreps.[8] Analogous calculations have been done for the nuclear octupole-phonon states of which the labeling is associated to the totally symmetric irreps of SO(7)[7,9] and G_2.[10–12]

In the present paper we illustrate on the examples of SU(3) and SO(5) the general theory developed in I which permits to establish shift operator relations without referring to the shift operator structure in terms of the algebra generators. Relations up to third degree within the SO(3) tensor will be considered.

II. SU(3): QUADRATIC SHIFT OPERATOR RELATIONS

It is well known that the SU(3) Lie algebra decomposes into its principal SO(3) Lie subalgebra and a five-dimensional SO(3) tensor representation $T(2,\mu)$, $\mu = -2,\dots,2$. Obviously the tensor T is only determined upon an overall numerical factor. According to the prescription outlined in I we fix that factor by means of the boson realization [I (2.2)] where b equals 1, since the lowest-dimensional irrep of SU(3) reduces into the three-dimensional irrep of its principal SO(3).[13] Previously,[2–6] in applying shift operator techniques to the SU(3) state labeling problem another scale has been used. The tensor operators defined there were denoted by q_μ, $\mu = -2,\dots,2$ and are related to the present $T(2,\mu)$ by

$$q_\mu = \sqrt{6}T(2,\mu) \quad (\mu = -2,\dots,2). \tag{2.1}$$

In order to be consistent with the previous notations and to make the comparison with earlier results easier, we shall adapt our formulas in I such that all T's are replaced by q's and that all shift operators defined in [I (2.7)–I (2.8)] are associated to reduced matrix elements of those q's.

Clearly, since apart from the SO(3) generators only one tensor operator is involved we can immediately apply formulas I (3.6)–I (3.8) in order to obtain all the relations between quadratic products of shift operators O_l^q in terms of a set of SO(3) shift operators $Q_l^{k;p}$. The $6j$-symbols occurring in I (3.6) can be found in analytic form in Biedenharn and Van Dam.[14] In the order of increasing total shift value these relations read

$$O_{l-2}^{-2} O_l^{-2} = Q_l^{4;-4}, \tag{2.2}$$

$$O_{l-2}^{-1} O_l^{-2} = Q_l^{4;-3}, \tag{2.3}$$

$$O_{l-1}^{-2} O_l^{-1} = Q_l^{4;-3}, \tag{2.4}$$

$$O_{l-2}^{0} O_l^{-2} = Q_l^{4;-2} + [-4][-5]/\sqrt{21}Q_l^{2;-2}, \tag{2.5}$$

$$O_{l-1}^{-1} O_l^{-1} = Q_l^{4;-2} - \sqrt{3}[2][-4]/4\sqrt{7}Q_l^{2;-2}, \tag{2.6}$$

$$O_l^{-2} O_l^{0} = Q_l^{4;-2} + [2][3]/\sqrt{21}Q_l^{2;-2}, \tag{2.7}$$

$$\begin{aligned} &O_{l-2}^{+1} O_l^{-2}/(l+m-1)(l-m-1) \\ &\quad = -Q_l^{4;-1} - \sqrt{3}[-3][-4]/\sqrt{7}Q_l^{2;-1}, \end{aligned} \tag{2.8}$$

$$O_{l-1}^{0} O_l^{-1} = Q_l^{4;-1} - [-3][10]/2\sqrt{21}Q_l^{2;-1}, \tag{2.9}$$

$$O_l^{-1} O_l^{0} = Q_l^{4;-1} - [3][-10]/2\sqrt{21}Q_l^{2;-1}, \tag{2.10}$$

$$\begin{aligned} &O_{l+1}^{-2} O_l^{+1}/(l+m+1)(l-m+1) \\ &\quad = -Q_l^{4;-1} - \sqrt{3}[3][4]/\sqrt{7}Q_l^{2;-1}, \end{aligned} \tag{2.11}$$

$$\begin{aligned} &O_{l-2}^{+2} O_l^{-2}/(l+m)(l-m)(l+m-1)(l-m-1) \\ &\quad = Q_l^{4;0} + 2\sqrt{3}[-2][-3]/\sqrt{7}Q_l^{2;0} \\ &\qquad + 3[0][-1][-2][-3]/5 \\ &\qquad \times \{12I_2 - (l+1)(l-5)\}, \end{aligned} \tag{2.12}$$

[a)] Senior Research Associate N.F.W.O. Belgium.

 0022-2488/85/010133-06$02.50

$$O_{l-1}^{+1}O_l^{-1}/(l+m)(l-m)$$
$$= -Q_l^{4;0} - \sqrt{3}[-2][-10]/4\sqrt{7}Q_l^{2;0} + 3[0][-1][-2][2]/40 \times\{24I_2 - (2l^2-3l-15)\}, \qquad (2.13)$$

$$(O_l^0)^2 = Q_l^{4;0} - [-3][5]/\sqrt{21}Q_l^{2;0} + [-1] \times[0][2][3]/20 \times\{24I_2 - (2l^2+2l-15)\}, \qquad (2.14)$$

$$O_{l+1}^{-1}O_l^{+1}/(l+m+1)(l-m+1)$$
$$= -Q_l^{4;0} - \sqrt{3}[4][12]/4\sqrt{7}Q_l^{2;0} + 3[0][2][3][4]/40\{24I_2-(2l^2+7l-10)\}, \qquad (2.15)$$

$$O_{l+2}^{-2}O_l^{+2}/(l+m+1)(l-m+1)(l+m+2)(l-m+2)$$
$$= Q_l^{4;0} + 2\sqrt{3}[4][5]/\sqrt{7}Q_l^{2;0} + 3[2][3][4][5]/5\{12I_2 - l(l+6)\}, \qquad (2.16)$$

$$O_{l-1}^{+2}O_l^{-1}/(l+m)(l-m)$$
$$= -Q_l^{4;+1} - \sqrt{3}[-1][-2]/\sqrt{7}Q_l^{2;+1}, \qquad (2.17)$$

$$O_l^{+1}O_l^0 = Q_l^{4;+1} - [-1][12]/2\sqrt{21}Q_l^{2;+1}, \qquad (2.18)$$

$$O_{l+1}^0 O_l^{+1} = Q_l^{4;+1} - [5][-8]/2\sqrt{21}Q_l^{2;+1}, \qquad (2.19)$$

$$O_{l+2}^{-1}O_l^{+2}/(l+m+2)(l-m+2)$$
$$= -Q_l^{4;+1} - \sqrt{3}[5][6]/\sqrt{7}Q_l^{2;+1}, \qquad (2.20)$$

$$O_l^{+2}O_l^0 = Q_l^{4;+2} + [0][-1]/\sqrt{21}Q_l^{2;+2}, \qquad (2.21)$$

$$O_{l+1}^{+1}O_l^{+1} = Q_l^{4;+2} - \sqrt{3}[0][6]/4\sqrt{7}Q_l^{2;+2}, \qquad (2.22)$$

$$O_{l+2}^0 O_l^{+2} = Q_l^{4;+2} + [6][7]/\sqrt{21}Q_l^{2;+2}, \qquad (2.23)$$

$$O_{l+1}^{+2}O_l^{+1} = Q_l^{4;+3}, \qquad (2.24)$$

$$O_{l+2}^{+1}O_l^{+2} = Q_l^{4;+3}, \qquad (2.25)$$

$$O_{l+2}^{+2}O_l^{+2} = Q_l^{4;+4}. \qquad (2.26)$$

Herein $[a]$ stands for $(2l+a)$, and I_2 is the second-order SU(3) Casimir. In terms of the SU(3) generators, I_2 takes the form $I_2 = \{3L^2 - \Sigma_\mu(-1)^\mu q_\mu q_{-\mu}\}/36$, whereas for any SU(3) irrep (p, q) it produces the eigenvalue $\langle I_2\rangle = (p^2+q^2-pq+3p)/9$. It is easily verified that

$$Q_l^{0;0} = 3\{12I_2 - l(l+1)\}/\sqrt{5}. \qquad (2.27)$$

It should be noticed that Eqs. (2.15)–(2.26) follow from Eqs. (2.13)–(2.2) by replacing formally l by $-l-1$ since this operation turns O_l^p and $Q_l^{k;p}$ into O_l^{-p} and $Q_l^{k;-p}$, respectively. Moreover, Eq. (2.14) is invariant under this formal transformation.

The relations between quadratic products of shift operators of the type O_l^k which have been previously established can be retrieved from (2.2)–(2.26) by eliminating from the relations shifting l by a same amount, the Q-type operators. As an example one obtains from (2.3)–(2.4) that $O_{l-1}^{-2}O_l^{-1} - O_{l-2}^{-1}O_l^{-2} = 0$.

A new aspect of the present relations is the introduction of a set of shift operators quadratic with respect to the q's, of which all the matrix elements are related to the reduced matrix elements of only three tensor products $(q\times q)^k$ $(k=4,2,0)$. Up to now only for $k=2$ the associated shift operators have been studied already in some detail elsewhere.[6]

Finally attention should be drawn upon the fact that at the SO(3) scalar level [Eqs. (2.12)–(2.16)], all quadratic products are expressed in terms of three independent SO(3) scalar operators $Q_l^{4;0}$ or $(O_l^0)^2$; $Q_l^{2;0}$, and I_2 or $Q_l^{0;0}$.

III. SU(3): RELATIONS OF THIRD DEGREE IN THE TENSOR REPRESENTATION

In the present section we want to establish relations which express the SO(3) scalar operator products $O_l^pQ_l^{k;-p}$ and $Q_l^{k;p}O_l^{-p}$ for all allowed p- and k- values, in terms of a set of independent SO(3) scalars O_l^0, $Q_l^{2;0}$, I_2 if necessary complemented with newly defined scalars $R_l^{k;K;0}$ of third degree in the q's.

Obviously, one could consider first Eqs. I (4.4) together with definitions I (4.1), translated into shift operator notation. However, it turns out here that it is much more advantageous to derive third-order relations from Eqs. (2.5)–(2.23) directly by multiplying either on the left or on the right with an appropriate O_l^k-shift operator. Let us demonstrate this by the following illustrative examples.

Multiplying both sides of Eq. (2.5) on the left with O_{l-2}^{+2} and multiplying both sides of Eq. (2.21) on the right with O_l^{-2} after having first changed l formally into $l-2$, two expressions are obtained for $O_{l-2}^{+2}O_{l-2}^0O_l^{-2}$. Equating their right-hand sides immediately yields the relation

$$(Q_{l-2}^{4;+2}O_l^{-2} - O_{l-2}^{+2}Q_l^{4;-2}) + ([-4][-5]/\sqrt{21}) \times(Q_{l-2}^{2;+2}O_l^{-2} - O_{l-2}^{+2}Q_l^{2;-2}) = 0. \qquad (3.1)$$

Similarly, multiplying (2.20) and (2.6), respectively, on the left and on the right with O_{l+1}^{-1} and O_l^{+2}, having replaced in the latter first l by $l+2$, gives

$$Q_{l+2}^{4;-2}O_l^{+2} - (\sqrt{3}[6][0]/4\sqrt{7})Q_{l+2}^{2;-2}O_l^{+2}$$
$$= (l+m+2)(l-m+2)\{-O_{l+1}^{-1}Q_l^{4;+1} - (\sqrt{3}[5][6]/\sqrt{7})O_{l+1}^{-1}Q_l^{2;+1}\}. \qquad (3.2)$$

In an analogous way one obtains from Eqs. (2.22) and (2.11)

$$O_{l+2}^{-2}Q_l^{4;+2} - (\sqrt{3}[6][0]/4\sqrt{7})O_{l+2}^{-2}Q_l^{2;+2}$$
$$= (l+m+2)(l-m+2)\{-Q_{l+1}^{4;-1}O_l^{+1} - (\sqrt{3}[5][6]/\sqrt{7})Q_{l+1}^{2;-1}O_l^{+1}\}. \qquad (3.3)$$

Furthermore, the form of relation (3.1) suggests that it is worthwhile to consider instead of the operators $O_l^pQ_l^{k;-p}$ and $Q_l^{k;p}O_l^{-p}$ their sum and difference. In order to cancel the m-dependent factors we introduce the notations

$$S_\pm^{k;+p} = [Q_{l-p}^{k;+p}O_l^{-p} \pm O_{l-p}^{+p}Q_l^{k;-p}]/g(p,l,m), \qquad (3.4)$$

where

$$g(p,l,m) = (l+m)!(l-m)!/(l+m-p)!(l-m-p)! \quad \text{if } p>0,$$
$$= (l+m-p)!(l-m-p)!/(l+m)!(l-m)! \quad \text{if } p<0. \qquad (3.5)$$

Adding or subtracting then Eqs. (3.2)–(3.3) yields

$$S_{\pm}^{4;-2} \pm S_{\pm}^{4;-1} = (\sqrt{3}[0][6]/4\sqrt{7})S_{\pm}^{2;-2} \mp (\sqrt{3}[5][6]/\sqrt{7})S_{\pm}^{2;-1}. \quad (3.6)$$

The systematic multiplication on the left and on the right of Eqs. (2.5)–(2.23) leads to two separate systems of equations, one between S_- operators, the other between S_+ operators. The first system is complete in the sense that all the $S_-^{k;p}$ operators can be solved in terms of a single operator $S_-^{2;0}$ which due to (3.4) is the commutator $[Q_l^{2;0}, O_l^0]$, hence a combination of lower-order SO(3) scalar operators. The complete solution is

$$S_-^{4;+2} = ([-3][-4][-5]/2\sqrt{21})S_-^{2;0}, \quad (3.7)$$

$$S_-^{4;+1} = ([-3][-4][10]/8\sqrt{21})S_-^{2;0}, \quad (3.8)$$

$$S_-^{4;0} = ([-3][5]/\sqrt{21})S_-^{2;0}, \quad (3.9)$$

$$S_-^{4;-1} = -([5][6][-8]/8\sqrt{21})S_-^{2;0}, \quad (3.10)$$

$$S_-^{4;-2} = -([5][6][7]/2\sqrt{21})S_-^{2;0}, \quad (3.11)$$

$$S_-^{2;+2} = -([-3]/2)S_-^{2;0}, \quad (3.12)$$

$$S_-^{2;+1} = ([-4]/4)S_-^{2;0}, \quad (3.13)$$

$$S_-^{2;-1} = -([6]/4)S_-^{2;0}, \quad (3.14)$$

$$S_-^{2;-2} = ([5]/2)S_-^{2;0}. \quad (3.15)$$

From the definition (3.4) together with (3.5) it is easily proved that $S_{\pm}^{k;p}$ and $S_{\pm}^{k;-p}$ go over into each other by the formal replacement of l by $-l-1$, a property which for the S_- operators reveals itself in the expressions (3.7)–(3.15). It should also be noticed that Eq. (3.9) which is a quadratic relation between the SO(3) scalars O_l^0, $Q_l^{2;0}$, and $Q_l^{4;0}$ is an immediate consequence of the fact that $Q_l^{4;0}$ can be linearly expressed in terms of $(O_l^0)^2$, $Q_l^{2;0}$, I_2, and L^2 as given in (2.14).

Finally, Eqs. (3.12)–(3.15) have been published[6] elsewhere yet in a slightly different form.

The situation for the S_+-type operators is somewhat more complicated. Since from left and right side multiplication only seven independent linear equations amongst the ten operators $S_+^{4;p}$ and $S_+^{2;p}$ are derived we cannot express them all in terms of $S_+^{2;0}$ and $S_+^{4;0}$ alone.

This shows that considering only these equations, there is at the level of third degree in the SO(3) tensor components q_μ, besides the combinations $S_+^{2;0}$ and $S_+^{4;0}$ of lower-order SO(3) scalar operators, room for one other SO(3) scalar operator. Clearly the existence of such an operator is only ensured if no other independent linear equation amongst the S_+-operators can be established. If that is the case, that operator is not expressible as a polynomial in the lower-order SO(3) scalars and hence, it should belong to the integrity basis of SO(3) scalar operators in the SU(3) enveloping algebra.

Now, we know such an operator exists, namely I_3, the third-order SU(3) Casimir. Therefore, we can reverse the argument and predict that none of the third-order relations I (4.4) transformed into shift operator language, can provide us with additional independent equations. It has indeed been explicitly verified by us that this is true. The third-order shift operator which is closely related to I_3 has been denoted in I as $R_l^{2;0;0}$. From I (4.1) it follows that

$$R_l^{2;0;0} = \{2\sqrt{5}[0][-1][-2][1][2][3][4]\}^{-1}\{[2][3][4]S_+^{2;+2} + 4[-1][3][4]S_+^{2;+1} + 6[1][-2][4]S_+^{2;0} + 4[-1][-2][3]S_+^{2;-1} + [-1][-2][0]S_+^{2;-2}\}. \quad (3.16)$$

With the help of (3.16) together with the seven relations amongst S_+ operators mentioned before, we find

$$S_+^{2;+2} = S_+^{2;0} + (\sqrt{3}[-1][-3]/\sqrt{7})\{48I_2 - [0][-10]\} \times O_l^0 + (\sqrt{5}[0][-1]^2[-6]/3)R_l^{2;0;0}, \quad (3.17)$$

$$S_+^{2;+1} = -S_+^{2;0} - (\sqrt{3}[0]/2\sqrt{7}) \times \{24[-4]I_2 - [2](2l^2 - 7l + 9)\} \times O_l^0 + (\sqrt{5}[2][0]^2[-1]/6)R_l^{2;0;0}, \quad (3.18)$$

$$S_+^{4;+2} = S_+^{4;0} - ([2][3]/\sqrt{21})S_+^{2;+2} + (2\sqrt{3}[-2][-3]/ \times\sqrt{7})S_+^{2;0} + (3[0][-1][-2][-3]/10) \times \{48I_2 - [2][-10]\}O_l^0, \quad (3.19)$$

$$S_+^{4;+1} = -S_+^{4;0} + ([3][-10]/2\sqrt{21})S_+^{2;+1} - (\sqrt{3}[-2][-10]/4\sqrt{7})S_+^{2;0} + (3[0][-1][-2][2]/20) \times \{24I_2 - (2l^2 - 3l - 15)\}O_l^0, \quad (3.20)$$

whereas $S_+^{2;-2}$, $S_+^{2;-1}$, $S_+^{4;-2}$, and $S_+^{4;-1}$ again follows from (3.17)–(3.20) by the formal transformation $l \to -l-1$.

From the foregoing it is clear that the SO(3) scalar operators up to third degree in the tensor representation, which we can denote by $R_l^{k;K;0}$, $Q_l^{k;0}$, and O_l^0 can be all expressed in terms of the SO(3) scalars O_l^0, $Q_l^{2;0}$, $Q_l^{0;0}$ or I_2, $R_l^{2;0;0}$ or I_3 and L^2. Hence, these five operators belong to the integrity basis for SO(3) scalars in the SU(3) enveloping algebra. On the other hand, it is already known from a paper by Judd *et al.*[15] that the complete integrity basis should be a subset of $\{O_l^0, Q_l^{2;0}, I_2, I_3, L^2$, and $R_l^{k;3;0}$ $(k=2$ or $4)\}$, which is an integrity basis for a related problem arising in the theory of polynomial invariants. The above comments indicate that $R_l^{k;3;0}$ is not algebraically independent from the other ones. Indeed, taking into account the definition I (4.1) we arrive at

$$R_l^{2;3;0} = S_l^{2;0}/4\sqrt{14} - (54\sqrt{6}/35)(L^2 - 2)O_l^0, \quad (3.21)$$

$$R_l^{4;3;0} = -S_l^{2;0}/4\sqrt{35} - (18\sqrt{3}/7\sqrt{5})(L^2 - 2)O_l^0. \quad (3.22)$$

IV. SO(5): SHIFT OPERATOR RELATIONS

Since the SO(5) Lie algebra reduces into its principal SO(3) subalgebra and a seven-dimensional tensor $T(3,\mu)$,

$\mu = -3,...,3$, and since the five-dimensional SO(5) irrep reduces into a five-dimensional SO(3) irrep, we can apply the formulas of I with $b = 2$.[13] The SO(5) Lie algebra also possesses a four-dimensional irrep which reduces into the four-dimensional irrep of the principal SO(3) subalgebra, showing that we could equally well choose $b = \frac{3}{2}$, but this choice is not covered by the general theory developed in I, and would also not lead to a simplification of the mathematics. The tensor operator $T(3,\mu)$ $\mu = -3,...,3$, defined by means of I (2.1)–I (2.2), in the present case coincides with the q_μ, $\mu = -3,...,3$, operator previously introduced in the context of SO(5) state labeling problems.[7]

The quadratic shift operator relations which are the analogs of (2.2)–(2.26) again follow immediately from I (3.6)–I (3.8). However, not all $6j$-symbols occurring can be found in closed analytic form in the standard tables. In order to generate such forms we have developed a FORTRAN program.

The following relations with nonpositive total shift values which result from our calculations, are

$$O_{l-3}^{-3} O_l^{-3} = Q_l^{6;-6}, \tag{4.1}$$

$$O_{l-3}^{-2} O_l^{-3} = Q_l^{6;-5}, \tag{4.2}$$

$$O_{l-2}^{-3} O_l^{-2} = Q_l^{6;-5}, \tag{4.3}$$

$$O_{l-3}^{-1} O_l^{-3} = Q_l^{6;-4} + ([-8][-9]/\sqrt{55})Q_l^{4;-4}, \tag{4.4}$$

$$O_{l-2}^{-2} O_l^{-2} = Q_l^{6;-4} - (\sqrt{5}[2][-8]/6\sqrt{11})Q_l^{4;-4}, \tag{4.5}$$

$$O_{l-1}^{-3} O_l^{-1} = Q_l^{6;-4} + ([2][3]/\sqrt{55})Q_l^{4;-4}, \tag{4.6}$$

$$\begin{aligned} O_{l-3}^{0} O_l^{-3} = Q_l^{6;-3} &+ (3[-7][-8]/\sqrt{55})Q_l^{4;-3} \\ &+ ([-6][-7][-8]/20\sqrt{2})O_l^{-3}, \end{aligned} \tag{4.7}$$

$$\begin{aligned} O_{l-2}^{-1} O_l^{-2} = Q_l^{6;-3} &- (2[-7][14]/3\sqrt{55})Q_l^{4;-3} \\ &- ([2][-6][-7]/30\sqrt{2})O_l^{-3}, \end{aligned} \tag{4.8}$$

$$\begin{aligned} O_{l-1}^{-2} O_l^{-1} = Q_l^{6;-3} &- (2[3][-18]/3\sqrt{55})Q_l^{4;-3} \\ &+ ([2][3][-6]/30\sqrt{2})O_l^{-3}, \end{aligned} \tag{4.9}$$

$$\begin{aligned} O_l^{-3} O_l^{0} = Q_l^{6;-3} &+ (3[3][4]/\sqrt{55})Q_l^{4;-3} \\ &- ([2][3][4]/20\sqrt{2})O_l^{-3}, \end{aligned} \tag{4.10}$$

$$\begin{aligned} &O_{l-3}^{+1} O_l^{-3}/(l+m-2)(l-m-2) \\ &\quad = -Q_l^{6;-2} - (6[-6][-7]/\sqrt{55})Q_l^{4;-2} \\ &\qquad - ([-4][-5][-6][-7]/3\sqrt{14})Q_l^{2;-2} \\ &\qquad - ([-5][-6][-7]/5\sqrt{2})O_l^{-2}, \end{aligned} \tag{4.11}$$

$$\begin{aligned} O_{l-2}^{0} O_l^{-2} = Q_l^{6;-2} &+ ([-6][-40]/2\sqrt{55})Q_l^{4;-2} \\ &- ([-4][-5][-6][2]/6\sqrt{14})Q_l^{2;-2} \\ &- ([-5][-6][8]/20\sqrt{2})O_l^{-2}, \end{aligned} \tag{4.12}$$

$$\begin{aligned} O_{l-1}^{-1} O_l^{-1} = Q_l^{6;-2} &- (4(4l^2-4l-57)/3\sqrt{55})Q_l^{4;-2} \\ &+ (\sqrt{2}[-4][-5][2][3]/15\sqrt{7})Q_l^{2;-2} \\ &+ (\sqrt{2}[3][-5]/5)O_l^{-2}, \end{aligned} \tag{4.13}$$

$$\begin{aligned} O_l^{-2} O_l^{0} = Q_l^{6;-2} &+ ([4][38]/2\sqrt{55})Q_l^{4;-2} \\ &- ([2][3][4][-4]/6\sqrt{14})Q_l^{2;-2} \\ &+ ([3][4][-10]/20\sqrt{2})O_l^{-2}, \end{aligned} \tag{4.14}$$

$$\begin{aligned} &O_{l+1}^{-3} O_l^{+1}/(l+m+1)(l-m+1) \\ &\quad = -Q_l^{6;-2} - (6[4][5]/\sqrt{55})Q_l^{4;-2} \\ &\qquad - ([2][3][4][5]/3\sqrt{14})Q_l^{2;-2} \\ &\qquad + ([3][4][5]/5\sqrt{2})O_l^{-2}, \end{aligned} \tag{4.15}$$

$$\begin{aligned} &O_{l-3}^{+2} O_l^{-3}/(l+m-2)(l-m-2)(l+m-1)(l-m-1) \\ &\quad = Q_l^{6;-1} + (2\sqrt{5}[-5][-6]/\sqrt{11})Q_l^{4;-1} \\ &\qquad + (5[-3][-4][-5][-6]/3\sqrt{14})Q_l^{2;-1} \\ &\qquad + ([-4][-5][-6]/2\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.16}$$

$$\begin{aligned} &O_{l-2}^{+1} O_l^{-2}/(l+m-1)(l-m-1) \\ &\quad = -Q_l^{6;-1} - (4[-5](4l-23)/3\sqrt{55})Q_l^{4;-1} \\ &\qquad + ([-3][-4][-5][6]/3\sqrt{14})Q_l^{2;-1} \\ &\qquad + ([-4][-5]/\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.17}$$

$$\begin{aligned} O_{l-1}^{0} O_l^{-1} = Q_l^{6;-1} &- ([-6][17]/\sqrt{55})Q_l^{4;-1} \\ &+ ([-3][-4][3][22]/15\sqrt{14})Q_l^{2;-1} \\ &- ([-4](2l^2-5l-27)/10\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.18}$$

$$\begin{aligned} O_l^{-1} O_l^{0} = Q_l^{6;-1} &- ([6][-17]/\sqrt{55})Q_l^{4;-1} \\ &+ ([3][4][-3][-22]/15\sqrt{14})Q_l^{2;-1} \\ &+ ([4](2l^2+5l-27)/10\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.19}$$

$$\begin{aligned} &O_{l+1}^{-2} O_l^{+1}/(l+m+1)(l-m+1) \\ &\quad = -Q_l^{6;-1} - (4[5](4l+23)/3\sqrt{55})Q_l^{4;-1} \\ &\qquad + ([3][4][5][-6]/3\sqrt{14})Q_l^{2;-1} \\ &\qquad + ([4][5]/\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.20}$$

$$\begin{aligned} &O_{l+2}^{-3} O_l^{+2}/(l+m+2)(l-m+2)(l+m+1)(l-m+1) \\ &\quad = Q_l^{6;-1} + (2\sqrt{5}[5][6]/\sqrt{11})Q_l^{4;-1} \\ &\qquad + (5[3][4][5][6]/3\sqrt{14})Q_l^{2;-1} \\ &\qquad - ([4][5][6]/2\sqrt{2})O_l^{-1}, \end{aligned} \tag{4.21}$$

$$\begin{aligned} &O_{l-3}^{+3} O_l^{-3}/(l+m-2)(l-m-2)(l+m-1) \\ &\qquad \times (l-m-1)(l+m)(l-m) \\ &\quad = -Q_l^{6;0} - (3\sqrt{5}[-4][-5]/\sqrt{11})Q_l^{4;0} \\ &\qquad - (5[-2][-3][-4][-5]/\sqrt{14})Q_l^{2;0} \\ &\qquad - ([0][-1][-2][-3][-4][-5]/\sqrt{7})Q_l^{0;0} \\ &\qquad - ([-3][-4][-5]/\sqrt{2})O_l^{0} \\ &\qquad + (3[-1][-2][-3][-4][-5]/10)L^2, \end{aligned} \tag{4.22}$$

$$O_{l-2}^{+2}O_l^{-2}/(l+m-1)(l-m-1)(l+m)(l-m)$$
$$= Q_l^{6;0} + (\sqrt{5}[-4](7l-23)/3\sqrt{11})Q_l^{4;0}$$
$$-(5[-2][-3][-4]/\sqrt{14})Q_l^{2;0}$$
$$-([0][-1][-2][-3][-4][2]/6\sqrt{7})Q_l^{0;0}$$
$$+([-3][-4][-10]/6\sqrt{2})O_l^0$$
$$+([-1][-2][-3][-4][5]/60)L^2, \quad (4.23)$$
$$O_{l-1}^{+1}O_l^{-1}/(l+m)(l-m)$$
$$= -Q_l^{6;0} - (2(2l^2-75l+153)/3\sqrt{55})Q_l^{4;0}$$
$$+(\sqrt{2}[-2][-3](2l^2-5l-22)/5\sqrt{7})Q_l^{2;0}$$
$$-([0][-1][-2][-3][2][3]/15\sqrt{7})Q_l^{0;0}$$
$$+(\sqrt{2}[-3](2l^2+5l-27)/15)O_l^0$$
$$+([-1][-2][-3][3][12]/150)L^2, \quad (4.24)$$
$$(O_l^0)^2 = Q_l^{6;0} - (6(l^2+l-17)/\sqrt{55})Q_l^{4;0}$$
$$+([-2][4](4l^2+4l-33)/5\sqrt{14})Q_l^{2;0}$$
$$-([0][-1][-2][2][3][4]/20\sqrt{7})Q_l^{0;0}$$
$$+(3\sqrt{2}(2l^2+2l-9)/5)O_l^0$$
$$+(3[-1][-2][3][4]/50)L^2. \quad (4.25)$$

From Eqs. (4.1)–(4.24) an independent set of 24 relations with non-negative total shift values is obtained by the transformation $l \mapsto -l-1$, whereas Eq. (4.25) is invariant. Again $Q_l^{0;0}$ is related to the second-order Casimir, i.e.,

$$Q_l^{0;0} = -\{10I_2 + l(l+1)\}/10. \quad (4.26)$$

If we eliminate the $Q_l^{k;p}$ operators ($k = 2,4,6$) from the 49 relations we obtain results which have elsewhere already been sufficient for solving the SO(5) state labeling problem for symmetric representations in an SO(3) basis.[7,8] Finally, relation (4.25) reveals that all SO(3) scalar operators $O_{l-k}^k O_l^{-k}$ can be expressed in terms of four independent operators $Q_l^{4;0}$, $Q_l^{2;0}$, $Q_l^{0;0}$ or I_2 and O_l^0. This is in complete agreement with the results of Gaskell *et al.*[16] and confirms that the operators above together with L^2 belong to the operator integrity basis.

Exactly, as we did before for SU(3) we now construct relations connecting the scalars $O_l^p Q_l^{k;-p}$ and $Q_l^{k;p} O_l^{-p}$ by multiplying the Eqs. (4.7)–(4.25) and their l-transformed forms on the right or on the left with appropriate shift operators O_l^k, and by equating left-hand side operator combinations.

It is also again advantageous to introduce the S-operators as defined in (3.4)–(3.5), because then the equations divide into two disjoint classes, separating S_+ and S_- operators completely. The set of equations in the S_- operators is complete in the sense that these can be solved in terms of combinations of lower-order SO(3) scalars. Indeed, straightforward calculations yield as results

$$S_-^{2;+2} = -([-4]/3)S_-^{2;0}, \quad (4.27)$$
$$S_-^{2;+1} = ([-6]/6)S_-^{2;0}, \quad (4.28)$$
$$S_-^{4;+3} = ([-6]/4)S_-^{4;0} + ([-2][-6](8l-19)/\sqrt{770})S_-^{2;0}, \quad (4.29)$$
$$S_-^{4;+2} = -([-7]/6)S_-^{4;0}$$
$$-(2\sqrt{2}[-1](l^2-7l+11)/\sqrt{385})S_-^{2;0}, \quad (4.30)$$
$$S_-^{4;+1} = ([-12]/12)S_-^{4;0} - (\sqrt{2}[0](3l-8)/\sqrt{385})S_-^{2;0}, \quad (4.31)$$
$$S_-^{6;+3} = -(3[-7][-8]/\sqrt{55})S_-^{4;+3}, \quad (4.32)$$
$$S_-^{6;+2} = -([-6][-40]/2\sqrt{55})S_-^{4;+2}$$
$$+([-4][-5][-6][2]/6\sqrt{14})S_-^{2;+2}, \quad (4.33)$$
$$S_-^{6;+1} = ([-6][17]/\sqrt{55})S_-^{4;+1}$$
$$-([-3][-4][3][22]/15\sqrt{14})S_-^{2;+1}. \quad (4.34)$$

For $S_-^{k;-p}$ with $p>0$ similar expressions follow from the transformation $l \mapsto -l-1$.

The set of equations for the S_+-type operators can be solved along the same lines as discussed for the SU(3) case. It is striking that again the third-order relations which follow from I (4.4) are not independent from the ones obtained so far by left–right multiplications. By solving the S_+-type equations we can conclude that the $S_+^{k;p}$ operators can be expressed in terms of combinations of lower-order SO(3) scalars and of two new independent SO(3) scalar operators of the type $R_l^{k;K;0}$. From the paper of Gaskell *et al.*[16] we learn that only $R_l^{k;1;0}$, $R_l^{k;3;0} (k \neq 0)$, $R_l^{k;4;0}$, and $R_l^{k;6;0}$ must be considered. Now, from the formulas I (4.1) and I (4.4) it is easy to verify that the $R_l^{k;K;0}$ SO(3) scalar operator for K even can be expressed in terms of S_- operators and lower-order scalars. Hence, they are not algebraically independent, and consequently the two operators $R_l^{2;1;0}$ and $R_l^{2;3;0}$ have to be added to the integrity basis for operators. However, that integrity basis is not yet complete. We know that in fact scalar operators up to the 15th degree in the generators have to be investigated.[16] Although, this analysis is theoretically possible within the framework of the developed theory, too many practical calculational difficulties would arise.

V. DISCUSSION

We have given two applications here of the general theory established in I. Both examples have in common that besides the SO(3) generators only one tensor operator is involved. However, the theory allows the treatment of cases with more tensor operators, of which SO(7)⊃SO(3) is the simplest one. The relations quadratic in the tensor operators, previously derived by means of other techniques[7,9] are easily reproduced by applying I (3.4) in shift operator notation.

[1]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **26**, 128 (1985).
[2]J. W. B. Hughes, J. Phys. A: Math. Gen. **6**, 48 (1973).
[3]J. W. B. Hughes, J. Phys. A: Math. Gen. **6**, 281 (1973).
[4]H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).

[5]H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2366 (1981).
[6]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1025 (1983).
[7]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1902 (1980); **21**, 2632 (1980).
[8]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980); **21**, 1913 (1980); **21**, 2635 (1980).
[9]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1339 (1983); **24**, 1345 (1983).
[10]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1967 (1980).
[11]G. Vanden Berghe, H. De Meyer, and J. Van der Jeugt, J. Math. Phys. **22**, 2332 (1981).
[12]H. De Meyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **23**, 1223 (1982).
[13]B. G. Wybourne, *Symmetry Principles and Atomic Spectrocopy* (Wiley–Interscience, New York, 1970).
[14]L. C. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[15]B. R. Judd, W. Miller Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[16]R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).

On the wave operator in few-body quantum scattering with Coulomb-like interactions

Helmut Kröger
Département de Physique, Université Laval, Québec, P.Q. G1K 7P4, Canada

(Received 4 April 1984; accepted for publication 27 July 1984)

Functional analytical methods like strong approximation of operator-valued functions and Dunfords' calculus are applied to yield a contour integral representation with a finite contour of wave operators for nonrelativistic potential scattering with Coulomb and short-range interactions, which avoids high singularities of Coulomb-like Green's functions of stationary scattering theory. The Green's function is the unique solution of the resolvent equation, which has an A-proper kernel allowing projector methods for its solution.

I. INTRODUCTION

The nonrelativistic quantum mechanical scattering theory is mathematically well established if the interaction is of short range.[1] That is reflected by many powerful approximation methods for the calculation of the physically relevant transition matrix elements. The transition matrix can be obtained from integral equations with compact kernels in a Hilbert space.[2,3] However, one often encounters charged particles in a nuclear or atomic scattering process. That means a long-range Coulomb interaction is involved. The Coulomb interaction brings about some major changes, e.g., the elastic forward scattering cross section (Rutherford) is infinite. A conventional partial wave expansion of the transition matrix cannot be summed up. From the mathematical point of view, the Coulomb potential manifests itself in higher singularities appearing in Green's functions, transition matrix, and scattering matrix, than those occurring from nonpathological strong interactions. In a two-body system interacting only via the Coulomb potential, these quantities are known analytically.[4] However, in most cases of practical interest one is concerned with a Coulomb plus a short-range interaction (called Coulomb-like). If one writes down integral equations for transition amplitudes analogously as for short–range interactions, i.e., the Lippmann–Schwinger equation in the two-body system and the Faddeev equations in the three-body system, one encounters highly singular kernels which makes the equation difficult to solve.[5,6] One way to resolve this problem is the screening technique which approximates the long-range Coulomb potential by a short-range (screened) interaction. That makes it possible to apply the conventional apparatus of short-range scattering theory, but requires some additional phase factors in the amplitudes. This technique has been successfully applied in elastic $p + d$ scattering,[7] but there remains some controversy about its applicability to the break-up reaction $p + d \rightarrow p + p + n$.[8]

In Ref. 9 we have suggested an approach to avoid the difficulties arising in stationary scattering integral equations due to the Coulomb force. The wave operators, which contain all scattering information of the system, build the S-matrix. In that approach the wave operators are strongly approximated by analytical functions (exponential function) of bounded operators, the latter being approximations of the full and asymptotic Hamiltonian, respectively, in the sense of strong resolvent convergence. It has been shown in Ref. 10 that the bounded Hamiltonians can be chosen to be of finite rank in a suitable basis of expansion functions, which reduces the calculation of the wave operators to a diagonalization of finite, real, symmetric matrices.

In this paper we want to suggest an alternative which leads to integral equations with well-behaved kernels. The Dunford calculus is applied to represent the approximate wave operator by a contour integral with a finite contour. The integrand carries the resolvent or Green's function of the approximate full Hamiltonian, which obeys a resolvent equation, relating it to the resolvent of the approximate asymptotic Hamiltonian. Most of the paper deals with the study of properties of this equation. We show that it has a unique solution, and it can be obtained from summing up the iterated equation (Neumann series) for a circular contour with a sufficiently large radius. For a general contour the kernel of the resolvent equations is shown to be A-proper. That allows us to find approximate solutions by projecting onto finite-dimensional subspaces.

The approach is applicable to short-range interactions, to the long-range Coulomb interaction and Coulomb-like interactions. For the sake of pedagogical simplicity the subsequent discussion concentrates on the two-body system, but a generalization to the N-body system is sketched.

II. THE TWO-BODY SYSTEM

We use the notation of Ref. 9. We assume that the center of mass motion can be split off and we consider only the relative motion between the two particles. Consider the Hilbert space $\mathscr{H} = L_2(\mathbb{R}^3)$. Let H^0 denote the free Hamiltonian which reads in momentum space, with p the relative momentum between the two particles and m its reduced mass,

$$h^0(p) = p^2/2m. \tag{2.1}$$

Let V^s denote a strong short-range interaction defined in coordinate representation by a real-valued function $v^s(\mathbf{r})$, $\mathbf{r} \in \mathbb{R}^3$, where $\mathbf{r}$ denotes the relative distance between the particles. One demands

$$v^s \in L_2(\mathbb{R}^3),$$
$$|v^s(\mathbf{r})| < c r^{-\beta}, \quad c > 0, \quad \beta > 1. \tag{2.2}$$

V^s is a bounded and self-adjoint operator on $\mathscr{H}$.[1] Let

$$H = H^0 + V^s. \tag{2.3}$$

H is self-adjoint but unbounded on the domain $D(H) = D(H^0)$.[1] The long-range Coulomb potential V^c is defined in coordinate space by

$$v^c(\mathbf{r}) = e_1 e_2 / r\,. \tag{2.4}$$

V^c is an unbounded operator due to the $1/r$ singularity. The full Hamiltonian (Coulomb-like)

$$H = H^0 + V^s + V^c \tag{2.5}$$

is self-adjoint but unbounded on $D(H) = D(H^0)$.[1] Now we consider the wave operators which contain all the scattering information on the system. According to Dollard,[11] an auxiliary time-dependent Hamiltonian is introduced

$$A^{0c}(t) = \mathrm{sgn}(t)(m/2H^0)^{1/2} e_1 e_2 \log(4H^0 |t|),$$
$$H^{0c}(t) = H^0 t + A^{0c}(t). \tag{2.6}$$

Then an appropriate wave operator for a Coulomb-like Hamiltonian is given by[11]

$$\Omega^{(\pm)} = \operatorname*{s-lim}_{t \to \mp\infty} \exp(iHt)\exp(-iH^{0c}(t)). \tag{2.7}$$

In order to carry out the steps indicated above we want to approximate $\Omega^{(\pm)}$ in the strong sense by exponentials of bounded operators. A sufficient condition for that is that $H^{0c}(t)$ and H are approximated in the sense of strong resolvent convergence. There are many ways to do so, the following definition gives an example:

$$H_u^0 = (E/u)\arctan(uH^0/E),$$
$$H_u^{0c}(t) = H_u^0 t + A^{0c}(t)(2/\pi)\arctan(H^0/uE). \tag{2.8}$$

V_u^c is defined in coordinate representation via

$$v_u^c(\mathbf{r}) = v^c(\mathbf{r})(2/\pi)\arctan(r/uR),$$
$$V_u^{sc} = V^s + V_u^c, \tag{2.9a}$$

$$H_u = H_u^0 + V_u^{sc}. \tag{2.9b}$$

E,R are fixed positive scaling parameters of dimension energy and length, respectively, u is the approximation parameter. $H_u^{0c}(t)$, H_u are bounded, self-adjoint, and approximate $H^{oc}(t)$ and H, respectively, in the sense of strong resolvent convergence if $u \to 0$.[9] Moreover,

$$\Omega_{u,T} = \exp(iH_u T)\exp(-iH_u^{0c}(T)) \tag{2.10}$$

approximates $\Omega^{(\pm)}$ in the sense of strong convergence for a suitable choice of parameter pairs (u,T) where $u \to 0$, $T \to \mp\infty$.[9]

Because H_u, H_u^0 are bounded self-adjoint operators its spectra lie in a finite interval on the real axis. Thus the Dunford calculus can be applied

$$\exp(iH_u T) = \frac{1}{2\pi i}\int_{C_u} d\xi \,\frac{\exp(i\xi T)}{\xi - H_u}, \tag{2.11}$$

where C_u is a closed contour around the spectrum of H_u chosen such that there is always a finite distance $d(C_u, \sigma(H_u)) \geqslant d_0 > 0$.

In practical calculations, where one has no *a priori* knowledge on $\sigma(H_u)$, one is on the safe side if one defines

$$I(H_u) = \{r \mid r \in \mathbb{R},\ |r| \leqslant \|H_u^0\| + \|V_u^{sc}\|\} \tag{2.12}$$

and chooses C_u such that $d(C_u, I(H_u)) \geqslant d_0 > 0$, which also guarantees $d(C_u, \sigma(H_u^0)) \geqslant d_0 > 0$. The same calculus could be applied to $\exp(-iH_u^0 T)$ but H_u^0, $H_u^{0c}(T)$ are diagonal in momentum space and hence $\exp(-iH_u^0 T)$, $\exp(-iH_u^{0c}(T))$ can be calculated easily.

Defining the resolvent or Green's function of an operator A via $G(\xi, A) = (\xi - A)^{-1}$, the Green's function $G(\xi, H_u)$, which appears in the contour integral [Eq. (2.11)] fulfills the resolvent identity

$$G(\xi, H_u) = G(\xi, H_u^0) + G(\xi, H_u^0) V_u^{sc} G(\xi, H_u), \quad \xi \in C_u. \tag{2.13}$$

In the following we want to study spectral properties of its kernel $K = G(\xi, H_u^0) V_u^{sc}$ and ways to solve the equation.

Proposition 1: The real interval $I = [\min(1, \|K\|), \max(1, \|K\|)]$ lies in the resolvent set $\rho(K)$. Moreover, there is an open connected region D_I being connected with $E_K = \{z \mid z \in \mathbb{C},\ |z| > \|K\|\}$ such that $I \subset D_I$ and $D_I \cup E_K \subset \rho(K)$ [tongue-shaped extension of $\rho(K)$ from the outer region, with $|z| > \|K\|$, to the point $z = 1$, for the case $\|K\| > 1$, see Fig. 1].

Proof: We distinguish the cases $\|K\| < 1$, and $\|K\| \geqslant 1$.

(i) In the case $\|K\| < 1$, $I = [\|K\|, 1] \subset E_K$. $E_K \subset \rho(K)$ because $r(K) < \|K\|$. E_K is open, hence one can find an open connected region D_I with $I \subset D_I \subset E_K$.

(ii) Now let us consider the case $\|K\| \geqslant 1$, $I = [1, \|K\|]$. Firstly we want to show that $[1, \infty) \subset \rho(K)$. Let us assume on the contrary $1 + \eta \in \sigma(K)$ for $\eta \in [0, \infty)$.

(a) If $1 + \eta \in \sigma_{\mathrm{point}}(K)$, there is $\psi \in \mathscr{H}$, $\psi \neq 0$ with

$$G(\xi, H_u^0) V_u^{sc} \psi = (1 + \eta)\psi. \tag{2.14}$$

Multiplication with $\xi - H_u^0$ yields

$$\xi\psi = (H_u^0 + V_u^{sc}/(1 + \eta))\psi, \tag{2.15}$$

i.e., $\xi \in \sigma(H_{u,\eta})$ where $H_{u,\eta} = H_u^0 + V_u^{sc}/(1 + \eta)$. $H_{u,\eta}$ is self-adjoint for all η. Thus if ξ has a nonvanishing imaginary part, it leads to a contradiction. If $\xi \in \mathbb{R}$, one can estimate

$$\|H_{u,\eta}\psi\| \leqslant (\|H_u^0\| + \|V_u^{sc}\|)\|\psi\| \tag{2.16}$$

and due to the choice of C_u

$$\|\xi\psi\| \geqslant (\|H_u^0\| + \|V_u^{sc}\| + d_0)\|\psi\| \tag{2.17}$$

which gives a contradiction.

(b) If $1 + \eta \in \sigma_{\mathrm{continuous}}(K)$ one arrives similarly at a contradiction.

(c) If $1 + \eta \in \sigma_{\mathrm{resolvent}}(K)$, one has that

$$\overline{\mathrm{range}(1 + \eta - K)}^{\perp} = \mathrm{null}(1 + \eta - K^+) \tag{2.18}$$

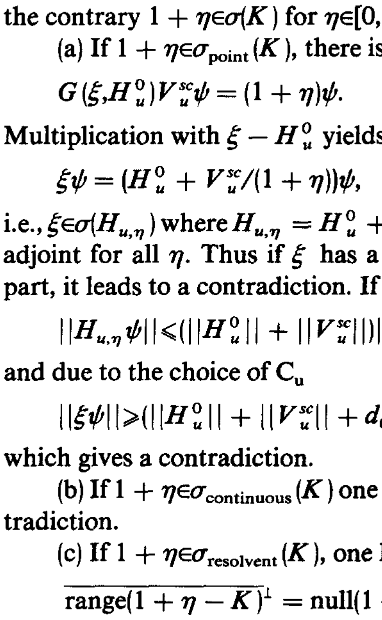


FIG. 1. Schematic plot of the spectrum $\sigma(K)$ with a tongue-shaped extension of $\rho(K)$ from the region outside of the disk with radius $\|K\|$ to the point 1.

is not empty, i.e., there is $\psi \in \mathscr{H}$, $\psi \neq 0$,

$$(1+\eta-K^{+})\psi=0,$$
$$K^{+}=V_{u}^{sc}G(\xi^{*},H_{u}^{0}). \tag{2.19}$$

With $\xi \in C_u$, one has $\xi^* \in \rho(H_u^0)$ and thus there is $\phi \in \mathscr{H}$, $\phi \neq 0$,

$$\psi=(\xi^{*}-H_{u}^{0})\phi, \tag{2.20}$$

thus

$$\xi^{*}\phi=(H_{u}^{0}+V^{sc}/(1+\eta))\phi, \tag{2.21}$$

i.e., $\xi^* \in \sigma_{\text{point}}(H_{u,\eta})$, but that leads to a contradiction similarly as in paragraph (a) $\xi \in \sigma_{\text{point}}(H_{u,\eta})$ has led to a contradiction. The paragraphs (a), (b), (c) imply $1+\eta \in \rho(K)$ for all $\eta \in [0,\infty)$.

Now we want to show that there is an open connected region D_I with $I \subset D_I$ and $D_I \subset \rho(K)$. Because $\rho(K)$ is open and $I=[1,\|K\|] \subset \rho(K)$, there is an open disk $D_{x,r}=\{z \mid z \in \mathbb{C}, |z-x|<r\} \subset \rho(K)$ for all $x \in I$. To each $D_{x,r}$ there corresponds a real interval $I_{x,r}=\{t \mid t \in \mathbb{R}, |t-x|<r\}$, open in $\mathbb{R}$. $\cup_{x \in I} I_{x,r}$ is an open cover of I in $\mathbb{R}$. But I is a compact interval in $\mathbb{R}$, hence there is a finite open cover of I in $\mathbb{R}$: $\cup_{\nu=1,\dots,N} I_{x_\nu,r_\nu}$. Hence, $D_I=\cup_{\nu=1,\dots,N} D_{x_\nu,r_\nu}$ is an open connected region in $\mathbb{C}$ having I as interior points. Because $z=\|K\|$ is an interior point of D_I, D_I is connected with E_K. Clearly $D_I \subset \rho(K)$, which finishes the proof.

In particular, Proposition 1 implies $1 \in \rho(K)$, which means that the resolvent equation has a unique solution $G(\xi,H_u)$, which holds for all $\xi \in C_u$.

Proposition 2: The solution $G(\xi,H_u)$ of the resolvent equation can be obtained by summing up the Neumann series of the resolvent equation, if the contour C_u is a circle centered at the origin with a sufficiently large radius R. The convergence is uniform in $\xi \in C_u$.

Proof: Let α: $0<\alpha<1$, R: $R \geqslant \|V_u^{sc}\|/\alpha+\|H_u^0\|$ and $C_u=\{Re^{i\phi} \mid \phi: 0 \to 2\pi\}$. Then one can estimate

$$\|G(\xi,H_{u}^{0})\|=\|(\xi-H_{u}^{0})^{-1}\| \leqslant 1/d(C_{u},\sigma(H_{u}^{0})) \leqslant 1/(R-\|H_{u}^{0}\|) \leqslant \alpha/\|V_{u}^{sc}\| \tag{2.22}$$

and

$$\|K\|=\|G(\xi,H_{u}^{0})V_{u}^{sc}\| \leqslant \|G(\xi,H_{u}^{0})\| \, \|V_{u}^{sc}\| \leqslant \alpha<1, \tag{2.23}$$

which holds for all $\xi \in C_u$. That is a sufficient condition for the convergence of the Neumann series

$$G(\xi,H_{u})=\sum_{\nu=0,1,2\dots} K^{\nu}G(\xi,H_{u}^{0}). \tag{2.24}$$

The convergence is uniform in $\xi \in C_u$, because the estimate on $\|K\|$ holds for all $\xi \in C_u$, which proves Proposition 2.

Although the Dunford integral is independent of the contour, as long as it is chosen properly to encircle the spectrum, it may be numerically inconvenient to use a circle with a large radius as required in the assumption of Proposition 2. Thus one is interested in solving the resolvent equation for a contour C_u without that restriction. To solve functional equations with compact kernels one can use powerful projection methods. In our case, however, the kernel $G(\xi,H_u^0)V_u^{sc}$ is not compact, because in momentum space the operator $G(\xi,H_u^0)$ does not vanish if the momentum approaches infinity, but rather it tends towards $1/(\xi-\pi E/2u)$ within the particular choice of H_u^0. There is a generalization of compact operators, the so-called A-proper operators, introduced by Petryshin,[12] which also allows us to apply projector methods to solve functional equations. This will be applicable in our case. Let us consider the resolvent equations when applied on a Hilbert state χ and let us denote

$$g=G(\xi,H_{u}^{0})\chi, \quad f=G(\xi,H_{u})\chi. \tag{2.25}$$

Then the resolvent equation reads

$$f=g+Kf \tag{2.26}$$

or with $B=1-K$,

$$Bf=g. \tag{2.27}$$

We define

$$T_{B,g}^{(n)}=\text{linear span of } \{B^{\nu}g \mid \nu=0,1,\dots,n\},$$
$$T_{B,g}=\overline{\bigcup_{n=0,1,\dots} T_{B,g}^{(n)}} \tag{2.28}$$

and in analogy $T_{B,Bg}^{(n)}$, $T_{B,Bg}$. Let $P^{(n)}$ denote the orthogonal projection onto $T_{B,g}^{(n)}$ and $Q^{(n)}$ the orthogonal projection onto $T_{B,Bg}^{(n)}$. Let $f^{(n)}$ be defined as the solution of

$$Q^{(n)}BP^{(n)}f^{(n)}=Q^{(n)}g. \tag{2.29}$$

Then we claim the following.

Theorem: $P^{(n)}f^{(n)}$ converges strongly to f, the approximation scheme is projectionally complete and B is an A-proper mapping.

Proof: In Proposition 1 it has been shown that $1 \in \rho(K)$ and there is an open connected region extending from E_K to the point 1. Then the Theorem, Corollary 1 and Corollary 2 of Ref. 13 can be applied which proves the claim.

III. THE N-BODY SYSTEM

In this section we want to sketch the generalization to the N-body system.

Assume there are N distinguishable particles, denoted by i,j considered as "elementary" as opposite to "composite" particles which are formed as bound states of the elementary particles. To each elementary particle i we ascribe a mass m_i and charge e_i. The channel index α denotes which of the elementary particles cluster and form a composite particle. The channel without composite particles is denoted by 0. We assume that the total interaction is a sum of pair interactions

$$V=\sum_{1 \leqslant i<j \leqslant N} V^{ij}, \tag{3.1}$$

where V^{ij} can have a short-range plus a Coulomb part. $V^{ij\alpha}$ is defined to be V^{ij} if i and j are contained both in any one of the composite particles of channel α, otherwise $V^{ij\alpha}=0$. A channel interaction is defined by

$$V^{\alpha}=\sum_{1 \leqslant i<j \leqslant N} V^{ij\alpha}. \tag{3.2}$$

Let H^0 denote the total kinetic energy Hamiltonian, let

$$H^{\alpha}=H^{0}+V^{\alpha}, \quad H=H^{0}+V \tag{3.3}$$

denote the Hamiltonian corresponding to channel α and the full Hamiltonian, respectively.

Dollard's time-dependent auxiliary Hamiltonian is generalized as follows.[11] Let m_{ij} denote the reduced mass of particles i,j and let H^{0ij} denote the kinetic energy Hamiltonian of the relative motion of particles i,j. For channel $\alpha = 0$ let

$$A^{0c}(t) = \mathrm{sgn}(t) \sum_{1<i<j<N} e_i e_j \left(\frac{m_{ij}}{2H^{0ij}}\right)^{1/2} \times \log(4H^{0ij}|t|),$$
$$H^{0c}(t) = H^0 t + A^{0c}(t). \qquad (3.4)$$

For channel $\alpha \neq 0$ let I, J denote either an elementary or a composite particle. $I \subset \alpha$ means that if I is an elementary particle, it is contained as an elementary particle in channel α, if I is a composite particle, it is contained as a composite particle in channel α. Let e_I denote the charge of particle I, m_I its mass, m_{IJ} the reduced mass of particles I, J and H^{0IJ} the kinetic energy Hamiltonian of the relative motion of particles I,J. Let

$$A^{\alpha c}(t) = \mathrm{sgn}(t) \sum_{I,J\subset\alpha}^{1<I<J<N} e_I e_J \left(\frac{m_{IJ}}{2H^{0IJ}}\right)^{1/2} \log(4H^{0IJ}|t|),$$
$$H^{\alpha c}(t) = H^\alpha t + A^{\alpha c}(t). \qquad (3.5)$$

The wave operator

$$\Omega^{\alpha(\pm)} = s - \lim_{t\to\mp\infty} \exp(iHt)\exp(-iH^{\alpha c}(t)) \qquad (3.6)$$

is defined on the domain D^α which is a projection of the Hilbert space onto a subspace where the bound-state wave function describes the internal motion of any composite particle in channel α.

We introduce the following bounded approximations:

$$V_u = \sum_{1<i<j<N} V_u^{ij},$$
$$V_u^\alpha = \sum_{1<i<j<N} V_u^{ij\alpha}, \qquad (3.7)$$

where each pair interaction V_u^{ij} is defined as in the two-body case [Eq. (2.9a)].

Let H_u^0 be defined as in the two-body case [Eq. (2.8)] and let

$$H_u^\alpha = H_u^0 + V_u^\alpha, \quad H_u = H_u^0 + V_u. \qquad (3.8)$$

For channel $\alpha = 0$ let

$$A_u^{0c}(t) = \mathrm{sgn}(t) \sum_{1<i<j<N} e_i e_j \left(\frac{m_{ij}}{2H^{0ij}}\right)^{1/2} \times \log(4H^{0ij}|t|)(2/\pi)\arctan(H_{0ij}/uE), \qquad (3.9a)$$

and for channel $\alpha \neq 0$ let

$$A_u^{\alpha c}(t) = \mathrm{sgn}(t) \sum_{I,J\subset\alpha}^{1<I<J<N} e_I e_J \left(\frac{m_{IJ}}{2H^{0IJ}}\right)^{1/2} \times \log(4H^{0IJ}|t|)(2/\pi)\arctan(H^{0IJ}/uE), \qquad (3.9b)$$
$$H_u^{\alpha c}(t) = H_u^\alpha t + A_u^{\alpha c}(t). \qquad (3.10)$$

The approximate wave operator takes the form

$$\Omega_{u,T}^\alpha = \exp(iH_u T)\exp(-iH_u^{\alpha c}(T)). \qquad (3.11)$$

In Ref. 9 it has been conjectured that $\Omega_{u,T}^\alpha$ tends strongly to $\Omega^{\alpha(\pm)}$ for suitably chosen $u \to 0$, $T \to \mp\infty$. However, all the results obtained in Sec. II of this paper for the two-body system can be taken over directly to the N-body case and formally read the same if the asymptotic Hamiltonian $H^{0c}(t)$ is substituted by $H^{\alpha c}(t)$ and the resolvent equation reads

$$G(\xi,H_u) = G(\xi,H_u^\alpha) + G(\xi,H_u^\alpha)(V_u - V_u^\alpha)G(\xi,H_u). \qquad (3.12)$$

IV. CONCLUSION

In nonrelativistic few-body quantum scattering theory a new approach has been suggested to calculate wave operators for short-range and long-range Coulomb interactions. Strongly approximated wave operators are related via a contour integral of finite length to Green's functions of Hamiltonians approximated in the sense of strong resolvent convergence. The resolvent equation, which relates Green's functions of the approximate full and asymptotic Hamiltonians, is studied and is found to have a unique solution, which can be obtained by a Neumann series. Its kernel is found to be A-proper which also allows projector methods to solve the resolvent equations.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

[1] W. O. Amrein, J. M. Jauch, and K. B. Sinha, "Scattering Theory in Quantum Mechanics," *Lecture Notes and Supplements in Physics* (Benjamin, Reading, MA, 1977), Vol. 16.
[2] H. Kröger and C. C. Fenske, Rep. Math. Phys. **17**, 459–470 (1980).
[3] H. Kröger, J. Math. Phys. **22**, 185–198 (1981).
[4] J. C. Y. Chen, in *Advances in Atomic and Molecular Physics*, edited by D. R. Bates and I. Esterman (Academic, New York, 1972), Vol. 8, pp. 71–129.
[5] Z. Bajzer, Z. Phys. A **278**, 97–103 (1976).
[6] Z. Bajzer, "The Relation Between S-Matrices and Scattering Amplitudes for Three Charged Particles," in *Proceedings of Workshop on Few-body Nuclear Physics, Trieste*, 1978, edited by G. Pisent, V. Vanzani, and L. Fonda (IAEA, Vienna, 1978), pp. 365–385.
[7] E. O. Alt, W. Sandhas, H. Zankel, and H. Ziegelmann, Phys. Rev. Lett. **37**, 1537–1540 (1976).
[8] C. Chandler, Nucl. Phys. A **353**, 129–142 (1981).
[9] H. Kröger, J. Math. Phys. **24**, 1509–1515 (1983).
[10] H. Kröger, J. Math. Phys. **25**, 1875–1880 (1984).
[11] J. D. Dollard, J. Math. Phys. **5**, 729–738 (1964).
[12] W. V. Petryshyn, Bull. Amer. Math. Soc. **81**, 223–312 (1975).
[13] H. Kröger, J. Math. Phys. **23**, 715–719 (1982).

A Fock–Krein realization of the Landau gauge

Dario Pierotti
Sezione Teorica, Dipartimento di Fisica, Universita' di Milano, 20133 Milano, Italy

(Received 15 March 1984; accepted for publication 18 May 1984)

An explicit Fock realization of the Landau gauge of the free electromagnetic field with a Hilbert space structure which is simply the same as in the Gupta–Bleuler gauges is given.

I. INTRODUCTION

As clarified in Ref. 1, when a given set of Wightman functions $\{\mathscr{W}\}$ do not satisfy the positivity condition one may associate with them different Hilbert space structures, leading in general to different spaces of states. A Hilbert space structure, $(\eta, \mathscr{H})$ associated to the Wightman functions $\{\mathscr{W}\}$, with η a nondegenerate metric operator, is called maximal if there is no other Hilbert space structure $(\tilde{\eta}, \tilde{\mathscr{H}})$ associated with $\{\mathscr{W}\}$, with a nondegenerate metric operator $\tilde{\eta}$, such that $\mathscr{H}\subset\tilde{\mathscr{H}}$ properly. The following theorem holds.[1]

A Hilbert space structure $(\eta, \mathscr{H})$ associated to a set of Wightman functions $\{\mathscr{W}\}$ is maximal if and only if η^{-1} is bounded. Moreover, if η^{-1} is bounded, it is possible to redefine the metric without changing η in such a way that the new metric satisfies $\eta^2 = 1$ (Krein spaces). This property is very convenient from a technical point of view.

All known realizations of the Landau gauge of the free electromagnetic field[2–4] do not seem to provide maximal Hilbert space structures, because the metric operator very likely has no bounded inverse. In this note we construct an explicit Fock realization of the Landau gauge with $\eta^2 = 1$ (Fock–Krein realization).

II. THE ONE-PARTICLE SPACE AND THE METRIC OPERATOR

Recall that, in the Landau gauge, the two-point function of the free vector potential $A_\mu(x)$ is[2]

$$\langle \Psi_0, A_\mu(x) A_\nu(y)\Psi_0\rangle = -(g_{\mu\nu} - \partial_\mu\partial_\nu\Box^{-1})(1/i)\mathscr{D}^{(+)}(x-y), \tag{1}$$

where

$$\Box^{-1}(1/i)D^{(+)}(x-y) = -[1/(4\pi)^2]\log(-x^2 + i\epsilon x_0)$$

is the two-point function of the dipole field model[1,5] whose Fourier transform is $2\pi\theta(p_0)\delta'(p^2)$.

Thus the two-point function (1) defines in momentum space the indefinite sesquilinear form

$$\langle f,g\rangle_L = \pi\int d^4p\, 2\theta(p_0)\delta'(p^2)[p^2\hat{f}_\mu(p)\hat{g}^\mu(-p) - p^\mu\hat{f}_\mu(p)\cdot p^\mu\hat{g}_\mu(-p)].$$

From the definition of the distribution $\delta'(p^2)$ we get further[4]

$$\langle f,g\rangle_L = \pi\int\frac{d^3p}{p_0}\Big\{-\hat{f}_\mu(p)\hat{g}^\mu(-p) + \frac{1}{2p_0}\frac{\partial}{\partial p_0}[p^\mu\hat{f}_\mu(p)\cdot p^\mu\hat{g}_\mu(-p)] - \frac{1}{2p_0^2}[p^\mu\hat{f}_\mu(p)\cdot p^\mu\hat{g}_\mu(-p)]\Big\}, \tag{2}$$

with $p_0 = |\mathbf{p}|$.

In our construction the one-particle space is simply defined by

$$\mathscr{H}^{(1)} = \Big\{\Psi_\mu(p):\ \Psi_\mu\in L^2\Big(\frac{d^3p}{p_0}, C_+\Big),\quad \forall_\mu = 0,1,2,3\Big\}$$

(C_+ is the future cone) provided with the standard scalar product

$$(\Phi,\Psi) = \sum_\mu(\Phi_\mu, \Psi_\mu) = \sum_\mu\int\frac{d^3p}{p_0}\bar{\Phi}_\mu(p)\Psi_\mu(p).$$

The metric operator η is given by the tensor $-g^{\mu\nu}$; obviously we have $\eta^2 = 1$. Then we get in $\mathscr{H}^{(1)}$ the indefinite inner product

$$\langle\Phi,\Psi\rangle = -(\Phi_\mu,\Psi^\mu) = -\int\frac{d^3p}{p_0}\bar{\Phi}_\mu(p)\Psi^\mu(p).$$

III. DEFINITION OF THE ELECTROMAGNETIC POTENTIAL

Now we perform a quantization over $\mathscr{H}^{(1)}$ with respect to $\langle\cdot,\cdot\rangle$ (see Refs. 6 and 7) and define [for each real valued $f\in\mathscr{S}(\mathbb{R}^4)$] the field operator as

$$A(f) = (1/\sqrt{2})[a(\pi f) + a^+(\pi f)],$$

where a, a^+ are the annihilation and creation operators and the vectors $(\pi f)\in\mathscr{H}^{(1)}$ are given by

$$(\pi f)_\mu(p) = \sqrt{2\pi}\Big\{\hat{f}_\mu(-p) - \frac{p_\mu}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p^\nu\hat{f}_\nu(-p)) - \frac{(p^\nu\hat{f}_\nu(-p))}{2|\mathbf{p}|}\Big]\Big\}\Big|_{C_+}. \tag{3}$$

It is not difficult to verify that $\overline{\{\pi f, f\in\mathscr{S}(\mathbb{R}^4)\}} = \mathscr{H}^{(1)}$ (see the Appendix). Moreover, if $\hat{f}_\mu = p_\mu f$, we obtain from (3)

$$(\pi f) = 0,$$

which assures that the vector potential A is transverse.

Finally we have the two-point function

$$\begin{aligned}
&\langle\Psi_0, A(f)A(g)\Psi_0\rangle\\
&= -\pi\int\frac{d^3p}{p_0}\Big\{\hat{f}_\mu(p) - \frac{p_\mu}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p^\nu\hat{f}_\nu(p)) - \frac{p^\nu\hat{f}_\nu(p)}{2|\mathbf{p}|}\Big]\Big\}\Big\{\hat{g}^\mu(-p) - \frac{p^\mu}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p^\nu\hat{g}_\nu(-p)) - \frac{p^\nu\hat{g}_\nu(-p)}{2|\mathbf{p}|}\Big]\Big\}_{(p_0=|\mathbf{p}|)}\\
&= \pi\int\frac{d^3p}{p_0}\Big\{-\hat{f}_\mu(p)\hat{g}^\mu(-p) + \frac{p^\mu\hat{f}_\mu(p)}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p^\nu\hat{g}_\nu(-p)) - \frac{p^\nu\hat{g}_\nu(-p)}{2|\mathbf{p}|}\Big] + \frac{p^\mu\hat{g}_\mu(-p)}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p^\nu\hat{f}_\nu(p)) - \frac{p^\nu\hat{f}_\nu(p)}{2|\mathbf{p}|}\Big]\Big\}\\
&= \pi\int\frac{d^3p}{p_0}\Big\{-\hat{f}_\mu(p)\hat{g}^\mu(-p) + \frac{1}{2|\mathbf{p}|}\frac{\partial}{\partial p_0}[p^\nu\hat{f}_\nu(p)p^\nu\hat{g}_\nu(-p)] - \frac{1}{2|\mathbf{p}|^2}[p^\nu\hat{f}_\nu(p)p^\nu\hat{g}_\nu(-p)]\Big\},
\end{aligned}$$

 0022-2488/85/010143-02$02.50

which is the required form (2).

In $\mathscr{H}^{(1)}$ the η-unitary representation $U(a,\Lambda)$ of the Poincaré group is given on the dense set of vector $\{\pi f\}$ by

$$U(a,\Lambda)(\pi f) = (\pi f_{a,\Lambda}),$$

where

$$(f_{a,\Lambda})_\mu(x) = \Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}(x-a)).$$

In particular, for the translations one has explicitly

$$(U(a,1)\Psi))_\mu(p) = e^{ip\cdot a}\Big(\Psi_\mu(p) - ia_0\frac{(p^\nu\Psi_\nu)}{2|\mathbf{p}|}p_\mu\Big)$$

for each $\Psi\in\mathscr{H}^{(1)}$ in the domain of $U(a,1)$. The form $\langle\cdot,\cdot\rangle$ is non-negative on the subspace $\mathscr{H}^{(1)\prime}\subset\mathscr{H}^{(1)}$ of vectors that satisfy $p^\mu\Psi_\mu = 0$.

The elements in $\mathscr{H}^{(1)\prime}$ with zero norm have the form $\Psi_\mu(p) = p_\mu\Psi(p)$ and constitute the subspace $\mathscr{H}^{(1)\prime\prime}$.

Note that

$$\Box A(f)\Psi_0 = -P_\mu(P^\nu\hat{f}_\nu)\in\mathscr{H}',$$

Ψ_0 being the vacuum state. In the Fock space $\mathscr{H}$ one has the corresponding subspaces $\mathscr{H}'$ and $\mathscr{H}''$.

It can be easily seen that the electromagnetic field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ leaves invariant the subspace $\mathscr{H}'$ and that the Maxwell equations are valid in mean value on $\mathscr{H}'$. Thus, in our realization, the Hilbert space structure is the same as in Gupta–Bleuler gauges. The nontrivial point is the definition of the vector potential giving the two-point function (1) and satisfying the condition $\partial^\mu A_\mu = 0$.

ACKNOWLEDGMENT

Professor F. Strocchi is warmly thanked for very useful suggestions and comments concerning this work.

APPENDIX: A DENSITY THEOREM IN $\mathscr{H}^{(1)}$

We show that vectors of the type $(\pi f)_\mu$ $(f_\mu\in\mathscr{S}(\mathbb{R}^4)$, $\mu = 0,1,2,3)$ are dense in $\mathscr{H}^{(1)}$ by proving that $\Psi\in\mathscr{H}^{(1)}$ and $\Psi\perp(\pi f)$ for each f implies $\Psi = 0$.

Choose f_μ such that $f_2 = f_3 = 0$ and $p_0 f_0 = p_1 f_1$. From definition (3) we have

$$(\pi f) = \begin{pmatrix} f_0\\ f_1\\ 0\\ 0\end{pmatrix}\quad (p^\mu f_\mu = 0).$$

Then the orthogonality conditon reads

$$\int\frac{d^3p}{p_0}(\bar{f}_0(p)\Psi_0(p) + \bar{f}_0(p)\Psi_1(p)) = \int\frac{d^3p}{p_0}\bar{f}_0(p)\Big(\Psi_0(p) + \frac{p_0}{p_1}\Psi_1(p)\Big) = 0,\quad \forall f_0,$$

with

$$f_0\in\{p_1 f, f\in\mathscr{S}(\mathbb{R}^4)\}.$$

Since this set is dense in $L^2(d^3p/p_0, C_+)$ we get $\Psi_0(p) + (p_0/p_1)\Psi_1(p) = 0$ a.e., that is $\Psi_1(p) = (-p_1/p_0)\Psi_0(p)$ a.e. In the same manner, with a suitable choice of test functions, we obtain

$$\Psi_2 = (-p_2/p_0)\Psi_0,\quad \Psi_3 = -(p_3/p_0)\Psi_0.$$

This means that the vector Ψ_μ has the form

$$\Psi_\mu(p) = \frac{\bar{p}_\mu}{p_0}\Psi_0(p),$$

where

$$\bar{p}_0 = p_0,\quad \bar{p}_i = -p_i = p^i,\quad i = 1,2,3.$$

now the condition $(\bar{p}_\mu/p_0)\Psi_0\perp(\pi f)$, $\forall f$, yields

$$\sum_\mu\int\frac{d^3p}{p_0}\Big\{\bar{f}_\mu - \frac{p_\mu}{2|\mathbf{p}|}\Big[\frac{\partial}{\partial p_0}(p\cdot\bar{f}) - \frac{(p\bar{f})}{2|\mathbf{p}|}\Big]\Big\}\frac{\bar{p}_\mu}{p_0}\Psi_0 = \int\frac{d^3p}{p_0}\frac{(p\cdot\bar{f})}{p_0}\Psi_0 = 0.$$

Choosing further

$$f_0 = f\in\mathscr{S}(\mathbb{R}^4),\quad f_i = 0,\quad i = 1,2,3,$$

we get

$$\int\frac{d^3p}{p_0}\bar{f}(p)\Psi_0(p) = 0,\quad \forall f\in\mathscr{S}(\mathbb{R}^4),$$

which implies $\Psi_0 = 0$.

[1]G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré XXXIII 3, 251 (1980).
[2]F. Strocchi and A. S. Wightman, J. Math. Phys. 15, 2198 (1974); J. Math. Phys. 17, 1930 (1976).
[3]G. Rideau, Lett. Math. Phys. 1, 17 (1975).
[4]M. Mintchev, Rep. Math. Phys. 17, 257 (1980).
[5]D. Zwanziger, Phys. Rev. D 17, 457 (1978).
[6]F. Strocchi and A. S. Wightman (private communication).
[7]M. Mintchev, J. Phys. A 13, 1841 (1980).

The short-range expansion for multiple well scattering theory

H. Holden,[a),b)] R. Høegh-Krohn,[a),b)] and M. Mebkhout[c)]
Centre de Physique Théorique, Centre National de la Recherche Scientifique, Luminy, Case 907, F-13288 Marseille Cedex 9, France

(Received 7 February 1984; accepted for publication 8 June 1984)

We expand the scattering amplitude and the scattering matrix around the zero-range limit, i.e., with point interactions, in the case when the potential has a finite number of wells. The lower-order terms are largely independent of the shape of the potentials.

I. INTRODUCTION

The explicitly solvable models serve as testing ground for what one can expect in general. Therefore it is of interest to study these models in great detail.

In nonrelativistic quantum mechanics there are essentially only three explicitly solvable models: The hydrogen atom (i.e., the Coulomb interaction), the harmonic oscillator, and, finally, the point interaction. The impact of the first model in the early days of quantum mechanics can hardly be overestimated, while the harmonic oscillator has been very important in, e.g., quantum field theory.

The last model to be mentioned, the point interaction or zero-range interaction, was studied already in the thirties by Fermi, Breit, Peierls, and Thomas in neutron–proton scattering in three dimensions while Kronig and Penney introduced the model that was later called the Kronig–Penney model of an infinite crystal in *one* dimension with a zero-range interaction. The rigorous study of these models was started much later by Berezin and Faddeev in the sixties and later made into a systematic theory by Høegh-Krohn and various co-workers. See Ref. 1 for an extensive and detailed exposition of this.

The n-center case, i.e., where one has n δ-potentials in $\mathbb{R}^3$, which we are going to discuss here, was first studied by Albeverio, Fenstad, and Høegh-Krohn[2] by means of nonstandard analysis and later extended by Grossmann, Høegh-Krohn, and Mebkhout.[3,4] The use of nonstandard analysis was initiated by Nelson[5] and Friedman.[6]

When one has a solvable model, it is reasonable to ask in what sense one can use the solvable model to deduce properties of models which are close to the solvable one in some sense, i.e., how well the solvable model is approximated by other more general models.

To approximate the point interactions with short-range potentials of the form $\epsilon^{-2}V(x/\epsilon)$ is an idea going back actually to Thomas in 1935, but it was put into a systematic and detailed study only in 1980 by Albeverio and Høegh-Krohn[7] where both the case of a finite and infinite number of centers are discussed. Their results were later improved in the one-center case by Albeverio, Gesztesy, and Høegh-Krohn,[8] and in the finite and infinite center case by Holden, Høegh-Krohn, and Johannesen.[9,10] In Ref. 8 the low-energy behavior of the scattering matrix is studied in great detail in the one-center case, and we rely very heavily on these results in this paper. The low-energy behavior of the scattering matrix in the one-center case has also been studied by Jensen and Kato[11] where weaker results are obtained (asymptotic expansions instead of analytic expansions) under weaker assumptions using other methods then the ones we advocate here.

Scattering theory for point interactions has also been studied by Karloukovski,[12] Thomas,[13] and Pavlov and Faddeev[14] while impurity scattering for point interactions in three dimensions has been studied in Ref. 15.

II. POINT INTERACTIONS

We will start by introducing the Hamiltonian with point interactions which is formally

$$H_{\nu,X} = -\Delta - \sum_{j=1}^{N} \nu_j \delta(\cdot - x_j), \tag{2.1}$$

where $\nu = (\nu_1,\ldots,\nu_N) \in \mathbb{R}^N$, $X = (x_1,\ldots,x_N)$ with $x_i \in \mathbb{R}^3$. As this is not a self-adjoint operator on $L^2(\mathbb{R}^3)$ we have to be careful how to define a rigorous analog of (2.1). There are by now several ways of doing this, e.g., by use of nonstandard analysis (see Ref. 2), self-adjoint extensions of symmetric operators (see Ref. 1), by a renormalization procedure on the coupling constant ν_j (Ref. 4), and finally as a limit of operators with decreasing support (see Refs. 1, 7, and 9). All these methods give the same operator, namely that $H_{\nu,X}$ can be rigorously defined as the unique self-adjoint operator $-\Delta_{(\alpha,X)}$ with resolvent

$$\begin{aligned}(-\Delta_{(\alpha,X)} - k^2)^{-1} &= G_k + \sum_{j,l=1}^{N}\left[\left(\alpha_j - \frac{ik}{4\pi}\right)\delta_{jl} - \tilde{G}_k(x_j - x_l)\right]^{-1}_{jl} \\ &\quad \times |G_k(\cdot - x_j)\rangle\langle \overline{G_k(\cdot - x_l)}|,\end{aligned} \tag{2.2}$$

with $\operatorname{Im} k > 0$ and

$$G_k = (-\Delta - k^2)^{-1} \tag{2.3}$$

has kernel

$$G_k(x-y) = e^{ik|x-y|}/4\pi|x-y| \tag{2.4}$$

and

$$\tilde{G}_k(x) = \begin{cases} G_k(x), & x \neq 0, \\ 0, & x = 0, \end{cases} \tag{2.5}$$

[a)] Permanent address: Matematisk institutt, Universitetet i Oslo, Blindern, Oslo 3, Norway.
[b)] Permanent address: Université de Provence, Centre St. Charles, France.
[c)] Permanent address: Université d'Aix-Marseille II, Faculté des Sciences de Luminy, Marseille, France.

where $[\ \ ]_{jl}^{-1}$ denotes the jlth element of the inverse of the matrix $[\ \ ]$ and finally the operator $S = |f\rangle\langle g|$ is defined to be $Sh = f(g,h)$.

For an extensive discussion of various aspects of this definition and with many properties of the operator, see Refs. 1 and 3.

With the operator (2.2) at hand we can start to approximate it by Hamiltonians with more regular, short-range interaction.

Let

$$H_\epsilon = -\Delta + \varepsilon^{-2}\sum_{j=1}^{N}\lambda_j(\epsilon)V_j\Big(\frac{1}{\epsilon}(\cdot - x_j)\Big), \tag{2.6}$$

where λ_j is analytic in a neighborhood of zero with $\lambda_j(0) = 1$ and V_j is a Rollnik function (i.e., $\int\int |x-y|^{-2}|V_j(x)||V_j(y)|dx\,dy < \infty$) with compact support. The operator H_ϵ is well-defined as a quadratic forms sum. Observe that while $(1/\epsilon^3)V_j((1/\epsilon)(\cdot - x_j)) \to \delta(\cdot - x_j)$ as $\epsilon \to 0$ we have $\epsilon((1/\epsilon^3)V_j((1/\epsilon)(\cdot - x_j))) \approx \epsilon\delta(\cdot - x_j)$ which indicates that the coupling constant v_j in the formal expression (2.1) has to be infinitesimally small in some sense.

One can prove that H_ϵ converges in the norm resolvent sense to $-\Delta_{(\alpha,X)}$ as ϵ tends to zero (see Refs. 1, 7, and 9).

The limit is, however, very delicate in the sense that it depends very crucially on detailed properties on the spectral point zero for the one-center operator:

$$H_j = -\Delta + V_j. \tag{2.7}$$

Let

$$v_j = |V_j|^{1/2}, \quad u_j = v_j \operatorname{sgn} V_j. \tag{2.8}$$

Then we have

$$(H_j - k^2)^{-1} = G_k - G_k v_j(1 + u_j G_k v_j)^{-1} u_j G_k, \tag{2.9}$$

where $\operatorname{Im} k > 0$ and we recall

$$G_k = (-\Delta - k^2)^{-1}. \tag{2.10}$$

This implies that k^2 is an eigenvalue for H_j iff -1 is an eigenvalue for $u_j G_k v_j$, $\operatorname{Im} k > 0$. If

$$(1 + u_j G_k v_j)\phi_j = 0, \tag{2.11}$$

then

$$\psi_j = G_k v_j \phi_j \tag{2.12}$$

will satisfy $\psi_j \in L^2(\mathbb{R}^3)$ and

$$H_j\psi_j = k^2\psi_j. \tag{2.13}$$

If, however, $k = 0$, i.e.,

$$(1 + u_j G_0 v_j)\phi_j = 0, \tag{2.14}$$

then

$$\psi_j = G_0 v_j \phi_j \tag{2.15}$$

will not in general be in $L^2(\mathbb{R}^3)$, but $\psi_j \in L^2_{\text{loc}}(\mathbb{R}^3)$ and satisfies

$$H\psi_j = 0 \tag{2.16}$$

in the sense of distributions (see Refs. 1 and 8).

If $\psi_j \notin L^2(\mathbb{R}^3)$ we will say that 0 is a resonance for H_j. It is necessary to make a further distinction (see Refs. 1 and 8), but first we state the following nice lemma to decide when ψ is in $L^2(\mathbb{R}^3)$.

Lemma 2.1: Let $V_j \in R$ and assume that supp V_j is compact. Then

$$\psi_j = G_0 v_j \phi_j \in L^2(\mathbb{R}^3) \Leftrightarrow (v_j,\phi_j) = 0. \tag{2.17}$$

Proof: See Ref. 8 where it is proved under weaker conditions on V_j.

Case I: -1 is not an eigenvalue for $u_j G_0 v_j$.

Case II: -1 is a simple eigenvalue for $u_j G_0 v_j$ and $(v_j,\phi_j) \neq 0$.

Case III: -1 is an eigenvalue for $u_j G_0 v_j$ with multiplicity $N \geqslant 1$ and $(v_j,\phi_{j1}) = \cdots = (v_j,\phi_{jN}) = 0$.

Case IV: -1 is a degenerate eigenvalue for $u_j G_0 v_j$ with multiplicity $N \geqslant 2$ and at least one ϕ_{jl} is such that $(v_j,\phi_{jl}) \neq 0$. In case IV we can always arrange such that $(v_j,\phi_{j1}) \neq 0$ and $(v_j,\phi_{j2}) = \cdots = (v_j,\phi_{jN}) = 0$. In addition we define

$$\tilde{\phi}_j = \phi_j \operatorname{sgn} V_j. \tag{2.18}$$

With this at hand we recall the basic result.

Theorem 2.2: Let H_ϵ be given by (2.6) where $V_j:\mathbb{R}^3 \to \mathbb{R}$ is a Rollnik function with a compact support.

If H_j is in case III or IV, assume in addition that $\lambda_j'(0) \neq 0$.

Then H_ϵ converges in the norm resolvent sense to $-\Delta_{(\alpha,X)}$ as $\epsilon \to 0$, where $\alpha = (\alpha_1,\ldots,\alpha_N)$ is given by

$$\alpha_j = \lambda_j'(0)|(v_j,\phi_j)|^{-2}(\tilde{\phi}_j,\phi_j). \tag{2.19}$$

Remarks: (1) In case I where $\phi_j = 0$, (2.19) is interpreted as $\alpha_j = \infty$, in case III where $(v_j,\phi_j) = 0$, we also interpret this as $\alpha_j = \infty$, and, finally, in case IV, $\phi_j = \phi_{j1}$, where $(v_j,\phi_{j1}) \neq 0$.

(2) If $\alpha_j = \infty$ for some j, this point has to be removed, i.e., we have $-\Delta_{(\tilde{\alpha},\tilde{X})}$, where $\tilde{X}$ consists of the points x_j where $\alpha_j \neq \infty$ and similar for $\tilde{\alpha}$.

(3) When $\lambda_j'(0) = 0$ in case III or IV we will not have norm convergence in general. See Refs. 8 and 9.

Proof: See Refs. 1 or 8.

We now introduce the scattering quantities.

Formally we can define for a Schrödinger operator

$$H = -\Delta + V \tag{2.20}$$

the corresponding t-matrix by [where $R = (H - k^2)^{-1}$]

$$\begin{aligned} t_k &= V - VRV \\ &= V - G_k^{-1}G_k VRV \\ &= V - G_k^{-1}(G_k - R)V = G_k^{-1}RV \\ &= G_k^{-1}RVG_kG_k^{-1} \\ &= G_k^{-1}(G_k - R)G_k^{-1} = v(1 + uG_kv)^{-1}u, \end{aligned} \tag{2.21}$$

using first the resolvent identity twice and then (2.9).

From the t-matrix we obtain the scattering amplitude f by

$$\begin{aligned} f(p,q,k) &= -\frac{1}{4\pi}(e^{ip\cdot},t_k e^{iq\cdot}) \\ &= -\frac{1}{4\pi}\int\int dx\,dy\, e^{-ipx}t_k(x,y)e^{iqy}. \end{aligned} \tag{2.22}$$

We can finally now define the on-shell scattering matrix S_k

$$S_k: L^2(S^{(2)}) \to L^2(S^{(2)}) \tag{2.23}$$

(where $S^{(2)} = \{x \in \mathbb{R}^3 | |x| = 1\}$) as a unitary operator given by

$$(S_k h)(\omega) = h(\omega) - (2\pi i)^{-1} k \int_{S^{(2)}} d\omega' f_{\rm on}(\omega,\omega',k) h(\omega'), \quad (2.24)$$

when $h \in L^2(S^{(2)})$, where

$$f_{on}(\omega,\omega',k) = f(p,q,k), \quad (2.25)$$

where $|p| = |q| = k$, $\omega = p/|p|$, and $\omega' = q/|q|$. See, e.g., Berthier[16] for an introduction to scattering theory.

III. THE ϵ-EXPANSION

The scheme for the ϵ-expansion of the scattering amplitude and the S-matrix is now clear.

First we derive the t-matrix for H_ϵ and then we expand it around the point interaction limit.

In order to do that, we relate the operator H_ϵ to an operator

$$H(\epsilon) = -\Delta + \sum_{j=1}^{N} \lambda_j(\epsilon) V_j\Big(\frac{\cdot - x_j}{\epsilon}\Big). \quad (3.1)$$

Using the unitary scaling operator U_ϵ given by

$$(U_\epsilon g)(x) = \epsilon^{-3/2} g(x/\epsilon), \quad (3.2)$$

when $g \in L^2(\mathbb{R}^3)$ we immediately see that

$$H_\epsilon = \epsilon^{-2} U_\epsilon H(\epsilon) U_\epsilon^{-1}, \quad (3.3)$$

and we then obtain the t-matrix for $H(\epsilon)$ first.

Lemma 3.1: The resolvent of $H(\epsilon)$ reads

$$(H(\epsilon) - k^2)^{-1} = G_k - \sum_{j,l=1}^{N} G_k \tilde v_j (1 + \tilde B_{\epsilon,k})_{jl}^{-1} \lambda_l(\epsilon) \tilde u_l G_k, \quad (3.4)$$

where

$$\tilde v_j(x) = v_j(x - x_j/\epsilon), \quad \tilde u_l(x) = u_l(x - x_j/\epsilon), \quad (3.5)$$

and $\tilde B_{\epsilon,k} = [\tilde B_{jl}]$ is given by

$$\tilde B_{jl} = \lambda_j(\epsilon) \tilde u_j G_k \tilde v_l. \quad (3.6)$$

The t-matrix $t_k(\epsilon)$ for $H(\epsilon)$ is

$$t_k(\epsilon) = \sum_{j,l=1}^{N} \tilde v_j (1 + \tilde B)_{jl}^{-1} \lambda_l(\epsilon) \tilde u_l. \quad (3.7)$$

Proof: Using Appendix A in Ref. 17 we obtain the stated formula for the resolvent of $H(\epsilon)$. From the formula $t_k = G_k^{-1}(G_k - R)G_k^{-1}$ we obtain the t-matrix as stated.

Using this and the stated unitary equivalence (3.3) between H_ϵ and $H(\epsilon)$ we obtain the t-matrix for H_ϵ.

Lemma 3.2: H_ϵ has a t-matrix given by

$$t_\epsilon(k) = \epsilon^{-2} U_\epsilon \Big[\sum_{j,l=1}^{N} \tilde v_j (U_j^\epsilon)^{-1} (1 + B_{\epsilon,k})_{jl}^{-1} \times \lambda_l(\epsilon) U_l^\epsilon \tilde u_l \Big] U_\epsilon^{-1}, \quad (3.8)$$

where $B_{\epsilon,k} = [B_{jl}]$ has the integral kernel

$$B_{jl}(x,y) = \epsilon \lambda_j(\epsilon) u_j(x) G_k(\epsilon(x-y) + x_j - x_l) v_l(y), \quad (3.9)$$

and U_j^ϵ and U_l^ϵ are given by (3.14).

Proof: We have that

$$\begin{aligned}(H_\epsilon - k^2)^{-1} &= \epsilon^2 U_\epsilon (H(\epsilon) - (\epsilon k)^2)^{-1} U_\epsilon^{-1} \\ &= G_k - \epsilon^{-2} \sum_{j,l=1}^{N} G_k U_\epsilon \tilde v_j \\ &\quad \times (1 + \tilde B_{\epsilon,\epsilon k})_{jl}^{-1} \lambda_l \tilde u_l U_\epsilon^{-1} G_k,\end{aligned} \quad (3.10)$$

which implies

$$t_\epsilon(k) = \sum_{j,l=1}^{N} \epsilon^{-2} U_\epsilon \tilde v_j (1 + \tilde B_{jl})_{jl}^{-1} \lambda_l(\epsilon) \tilde u_l U_\epsilon^{-1}. \quad (3.11)$$

By introducing matrix notation we can write this as

$$t_\epsilon(k) = \epsilon^{-1} U_\epsilon \begin{bmatrix} \tilde v_1 & & 0 \\ & \ddots & \\ 0 & & \tilde v_N \end{bmatrix} [1 + \tilde B_{\epsilon,\epsilon k}]^{-1} \times \begin{bmatrix} \lambda_1(\epsilon)\tilde u_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N(\epsilon)\tilde u_N \end{bmatrix} U_\epsilon^{-1}. \quad (3.12)$$

Defining the unitary operator

$$\tilde U^\epsilon = \begin{bmatrix} U_1^\epsilon & & 0 \\ & \ddots & \\ 0 & & U_N^\epsilon \end{bmatrix}, \quad (3.13)$$

where

$$(U_j^\epsilon g)(x) = g(x + x_j/\epsilon) \quad (3.14)$$

for $g \in L^2(\mathbb{R}^3)$, we can introduce $\tilde U^\epsilon \tilde U^{\epsilon\,-1}$ between the two matrix products to obtain that

$$t_\epsilon(k) = \epsilon^{-2} U_\epsilon \sum_{j,l=1}^{N} \tilde v_j (U_j^\epsilon)^{-1} \times [1 + B_{\epsilon,k}]_{jl}^{-1} \lambda_l(\epsilon) U_l^\epsilon \tilde u_l U_\epsilon^{-1}, \quad (3.15)$$

where $B_{\epsilon,k} = [B_{jl}]$ has the integral kernel

$$B_{jl} = \epsilon \lambda_j(\epsilon) u_j(x) G_k(\epsilon(x-y) + x_j - x_l) v_l(y). \quad (3.16)$$

From this lemma we see that $t_\epsilon(k)$ is defined whenever $-1 \notin \sigma(B_{\epsilon,k})$. For $\epsilon = 0$, $B_{\epsilon,k}$ reduces to

$$B_{0,k} = [\delta_{jl} u_j G_0 v_j], \quad (3.17)$$

which implies from the analysis in Sec. II that the detailed properties of $t_\epsilon(k)$ around $\epsilon = 0$ depends on the zero-energy properties of the operators $H_j = -\Delta + V_j, j = 1,\dots,N$.

We recall here the following.

Proposition 3.3: We have the expansion

$$\epsilon(1 + \lambda_j(\epsilon) u_j G_{\epsilon k} v_j)^{-1} = Q_j + \epsilon R_j + o(\epsilon), \quad (3.18)$$

where

$$Q_j = \begin{cases} 0, & \text{in case I,} \\ (ik/4\pi - \alpha_j)^{-1}|\phi_j\rangle\langle\tilde{\phi}_j|, & \text{in case II,} \\ -\dfrac{1}{\lambda_j'(0)} \displaystyle\sum_{l=1}^{N_j} \dfrac{|\phi_{jl}\rangle\langle\tilde{\phi}_{jl}|}{(\tilde{\phi}_{jl},\phi_{jl})}, & \text{in case III,} \\ \displaystyle\sum_{l,m=1}^{N_j} \left[\frac{ik}{4\pi}(\phi_{jl},v_j)(v_j,\phi_{jm}) - \lambda_j'(0)(\tilde{\phi}_{jl},\phi_{jm})\right]^{-1}_{lm} |\phi_{jl}\rangle\langle\tilde{\phi}_{jm}|, & \text{in case IV,} \end{cases} \tag{3.19}$$

and

$$R_j = \begin{cases} (1+u_jG_0v_j)^{-1}, \quad \text{in case I,} \\ T_j - \dfrac{ik}{4\pi}\left(\dfrac{ik}{4\pi}-\alpha_j\right)^{-1}(|\phi_j\rangle\langle T_j^*v_j| + |T_ju_j\rangle\langle\tilde{\phi}_j|) + \left(\dfrac{ik}{4\pi}-\alpha_j\right)^{-2}\Bigg[\left(\dfrac{ik}{4\pi}\right)^2(v_j,T_ju_j)|\phi_j\rangle\langle\tilde{\phi}_j| \\ \quad - \left(\dfrac{ik}{4\pi}\lambda_j'(0) - \dfrac{1}{2}\lambda_j''(0)(\tilde{\phi}_j,\phi_j) - \dfrac{1}{8\pi}\displaystyle\int\int dx\,dy\,\phi_j(x)v_j(x)|x-y|v_j(y)\phi_j(y)\right)|\phi_j\rangle\langle\tilde{\phi}_j|\Bigg], \quad \text{in case II,} \\ T_j - \lambda_j'(0)^{-2}\displaystyle\sum_{l,m=1}^{N_j}\left[-\frac{\lambda_j''(0)}{2}(\tilde{\phi}_{jl},\phi_{jm}) + \frac{ik}{4\pi}\lambda_j'(0) - \frac{k^2}{8\pi}\int\int dx\,dy\,\phi_{jl}(x)v_j(x)|x-y|v_j(y)\phi_{jm}(y)\right] \\ \quad \times \dfrac{|\phi_{jl}\rangle\langle\tilde{\phi}_{jm}|}{(\tilde{\phi}_{jl},\phi_{jl})(\tilde{\phi}_{jm},\phi_{jm})}, \quad \text{in case III,} \\ T_j - \dfrac{ik}{4\pi}(|T_ju_j\rangle\langle Q_j^*v_j| + |Q_ju_j\rangle\langle T_j^*v_j|) + \left(\dfrac{ik}{4\pi}\right)^2(v_j,T_ju_j)|Q_ju_j\rangle\langle Q_j^*u_j| \\ \quad - \displaystyle\sum_{\substack{l,m=1 \\ l',m'=1}}^{N_j}\left[\frac{ik}{4\pi}(\phi_{jl},v_j)(v_j,\phi_{jm}) - \lambda_j'(0)(\tilde{\phi}_{jl},\phi_{jm})\right]^{-1}_{lm}\left[\frac{ik}{4\pi}(\phi_{jl'},v_j)(v_j,\phi_{jm'}) - \lambda_j'(0)(\tilde{\phi}_{jl'},\phi_{jm'})\right]^{-1}_{l'm'} \\ \quad \times\Bigg(-\dfrac{1}{2}\lambda_j''(0)(\tilde{\phi}_{jm'},\phi_{jl'}) + \dfrac{ik}{4\pi}\lambda_j'(0)(\phi_{jm},v_j)(v_j,\phi_{jl'}) \\ \quad - \dfrac{k^2}{8\pi}\displaystyle\int\int dx\,dy\,\phi_{jm}(x)v_j(x)|x-y|v_j(y)\phi_{jl'}(y)\Bigg)|\phi_{jl'}\rangle\langle\tilde{\phi}_{jm'}|, \quad \text{in case IV,} \end{cases} \tag{3.20}$$

where T_j is the reduced resolvent, i.e.,

$$T_j = n - \lim_{\epsilon\to 0}(1+\epsilon+u_jG_0v_j)^{-1}(1-P_j), \tag{3.21}$$

where P_j is the projection onto the eigenspace corresponding to the eigenvalue -1, viz.,

$$P_j = \sum_{l=1}^{N_j}\frac{|\phi_{jl}\rangle\langle\tilde{\phi}_{jl}|}{(\tilde{\phi}_{jl},\phi_{jl})}. \tag{3.22}$$

Proof: See Theorems 3.1–3.4 in Ref. 8 and Ref. 18.

From this we infer the expansion for the full $\epsilon(1+B_{\epsilon,k})^{-1}$ term.

Theorem 3.4: Let the points $x_1,\ldots,x_N$ be numbered such that $x_1,\ldots,x_n$ (resp. $x_{n+1},\ldots,x_N$) are in case II or IV (resp. I or III). Then we have the following expansion:

$$\begin{aligned}\epsilon(1+B_{\epsilon,k})^{-1} &= X + \epsilon Y + o(\epsilon) \\ &\equiv (1+QT_0)^{-1}Q + \epsilon((1+QT_0)^{-1}R \\ &\quad - (1+QT_0)^{-1}(RT_0+QU)(1+QT_0)^{-1}Q) \\ &\quad + o(\epsilon),\end{aligned} \tag{3.23}$$

where $Q=[\delta_{jl}Q_j]$, $R=[\delta_{jl}R_j]$ are given in Proposition 3.3 and

$$T_0 = [\tilde{G}_k(x_j-x_l)|u_j\rangle\langle v_l|] \tag{3.24}$$

and $U=[(1-\delta_{jl})U_{jl}]$ has the integral kernel

$$\begin{aligned}U_{jl}(x,y) &= \lambda_j'(0)\tilde{G}_k(x_j-x_l)u_j(x)v_l(y) \\ &\quad + u_j(x)\nabla G_k(x_j-x_l)\cdot(x-y)v_l(y).\end{aligned} \tag{3.25}$$

Proof: Separating off the diagonal elements of $B_{\epsilon,k}$ we obtain

$$1+B_{\epsilon,k} = 1+S_\epsilon+\epsilon T_\epsilon, \tag{3.26}$$

where $S_\epsilon=[\delta_{jl}\lambda_j(\epsilon)u_jG_{\epsilon k}v_j]$ and $T_\epsilon=[(1-\delta_{jl})T_{jl}]$ has the integral kernel

$$T_{jl}(x,y) = \lambda_j(\epsilon)u_j(x)G_k(\epsilon(x-y)+x_j-x_l)v_l(y). \tag{3.27}$$

Expanding T_ϵ we find

$$T_\epsilon = T_0+\epsilon U+o(1), \tag{3.28}$$

where U is as stated in the theorem. From Proposition 3.3 we know the expansion of $\epsilon(1+S_\epsilon)^{-1}=Q+\epsilon R+o(\epsilon)$. Hence

$$
\begin{aligned}
\epsilon(1+B_{\epsilon,k})^{-1} &= \epsilon(1+S_\epsilon+\epsilon T_\epsilon)^{-1}\\
&=(1+\epsilon(1+S_\epsilon)^{-1}T_\epsilon)^{-1}\epsilon(1+S_\epsilon)^{-1}\\
&=(1+QT_0+\epsilon(RT_0+QU)+o(\epsilon))^{-1}\\
&\quad\times(Q+\epsilon R+o(\epsilon))\\
&=(1+\epsilon((1+QT_0)^{-1}(RT_0+QU)+o(1)))^{-1}\\
&\quad\times(1+QT_0)^{-1}(Q+\epsilon R+o(\epsilon))\\
&=(1-\epsilon(1+QT_0)^{-1}(RT_0+QU)+o(\epsilon))\\
&\quad\times(1+QT_0)^{-1}(Q+\epsilon R+o(\epsilon))\\
&=(1+QT_0)^{-1}Q+\epsilon((1+QT_0)^{-1}\\
&\quad\times R-(1+QT_0)^{-1}\\
&\quad\times(RT_0+QU)(1+QT_0)^{-1}Q)+o(\epsilon).
\end{aligned}
\tag{3.29}
$$

We can now turn to the scattering amplitude and state the following theorem.

Theorem 3.5: The scattering amplitude $f_\epsilon(p,q,k)$ for H_ϵ is analytic and has the expansion

$$
\begin{aligned}
f_\epsilon(p,q,k) &= -(1/4\pi)(e^{ip\cdot},t_\epsilon(k)e^{iq\cdot})\\
&=-\frac{1}{4\pi}\sum_{j,l=1}^{n} e^{-i(px_j-qx_l)}\\
&\quad\times\Big[\Big(\alpha_j-\frac{ik}{4\pi}\Big)\delta_{jl}-\tilde G_k(x_j-x_l)\Big]^{-1}_{jl}\\
&\quad-\frac{\epsilon}{4\pi}\Big(\sum_{j,l=1}^{n} e^{-i(px_j-qx_l)}\lambda_l'(0)\\
&\quad\times\Big[\Big(\alpha_j-\frac{ik}{4\pi}\Big)\delta_{jl}-\tilde G_k(x_j-x_l)\Big]^{-1}_{jl}\\
&\quad+\sum_{j,l=1}^{n} ie^{-i(px_j-qx_l)}\\
&\quad\times[-(p\cdot v_j,X_{jl}u_l)+(v_j,X_{jl}u_lq\cdot)]\\
&\quad+\sum_{j,l=1}^{N} e^{-i(px_j-qx_l)}(v_j,Y_{jl}u_l)\Big)+o(\epsilon).
\end{aligned}
\tag{3.30}
$$

Proof: Using the expansion for $\epsilon(1+B_{\epsilon,k})^{-1}$ from the preceding proposition and expanding the other terms we obtain

$$
\begin{aligned}
f_\epsilon(p,q,k) &= -\frac{1}{4\pi}(e^{ip\cdot},t_\epsilon(k)e^{iq\cdot})\\
&=-\frac{1}{4\pi}\sum_{j,l=1}^{N}(e^{ip(\epsilon\cdot+x_j)},v_j\epsilon(1+B_{\epsilon,k})^{-1}_{jl}\\
&\quad\times\lambda_l(\epsilon)u_le^{iq(\epsilon\cdot+x_l)})\\
&=-\frac{1}{4\pi}\sum_{j,l=1}^{N} e^{-i(px_j-qx_l)}[(v_j,X_{jl}u_l)\\
&\quad+\epsilon(\lambda_\lambda'(0)(v_j,X_{jl}u_l)+(p\cdot v_j,X_{jl}u_l)\\
&\quad+(v_j,X_{jl}u_lq\cdot)+(v_j,Y_{jl}u_l))].
\end{aligned}
\tag{3.31}
$$

By inserting the explicit expressions for X_{jl} and Y_{jl} we see that in all terms except the last one, we only have to sum over $j,l\leqslant n$.

Recall now from Sec. I that the on-shell scattering matrix S_k is the unitary operator

$$S_k: L^2(S^{(2)})\to L^2(S^{(2)}), \tag{3.32}$$

given by

$$(S_kh)(\omega)=h(\omega)-(2\pi i)^{-1}k\int_{S^{(2)}}d\omega'\,f_{\rm on}(\omega,\omega',k)h(\omega'), \tag{3.33}$$

where

$$f_{\rm on}(\omega,\omega',k)=f(\omega k,\omega'k,k). \tag{3.34}$$

Theorem 3.6: The scattering matrix $S_{\epsilon,k}$ for H_ϵ is analytic and has the expansion

$$
\begin{aligned}
S_{\epsilon,k} &= 1+\frac{k}{8\pi^2 i}\sum_{j,l=1}^{n}|e^{-ikx_j}\rangle\\
&\quad\times\Big[\Big(\alpha_j-\frac{ik}{4\pi}\Big)\delta_{jl}-\tilde G_k(x_j-x_l)\Big]^{-1}_{jl}\langle e^{-ikx_l}|\\
&\quad+\frac{\epsilon k}{8\pi^2 i}\Big[\sum_{j,l=1}^{n}|e^{-ikx_j}\rangle\lambda_l'(0)\\
&\quad\times\Big[\Big(\alpha_j-\frac{ik}{4\pi}\Big)\delta_{jl}-\tilde G_k(x_j-x_l)\Big]^{-1}_{jl}\langle e^{-ikx_l}|\\
&\quad+\sum_{j,l=1}^{N}|e^{-ikx_j}\rangle(v_j,Y_{jl}u_l)\\
&\quad\times\langle e^{-ikx_l}|+\sum_{j,l=1}^{n}Z_{jl}\Big]+O(\epsilon),
\end{aligned}
\tag{3.35}
$$

where Z_{jl} has integral kernel

$$
\begin{aligned}
Z_{jl}(\omega,\omega') &= ke^{-ik(x_j\omega-x_l\omega')}((\omega\cdot v_j,X_{jl}u_l)\\
&\quad+(v_j,X_{jl}u_l\omega'\cdot)).
\end{aligned}
\tag{3.36}
$$

Proof: The expansion follows directly from (3.33) and Theorem 3.5.

Remark: It is, of course, possible to expand all the relevant quantities to an arbitrary order. However, as should be apparent already from the first-order terms, the formulas become more complicated.

Point interactions are also related to another interesting limiting case, namely, as we already have seen, to the operator $H(\epsilon)$ given by (3.1).

This operator corresponds to the situations where the potentials move apart as $\epsilon\to0$. For simplicity we only state the results when $\lambda_j(\epsilon)\equiv1$, for $j=1,\ldots,N$, i.e., $\alpha_j=0$ in cases II and IV.

Theorem 3.7: Let

$$H(\epsilon)=-\Delta+\sum_{j=1}^{N}V_j\Big(\frac{\cdot-x_j}{\epsilon}\Big), \tag{3.37}$$

and let $f(\epsilon,p,q,k)$ denote its scattering amplitude. Then we have

$$
\begin{aligned}
f(\epsilon,\epsilon p,\epsilon q,\epsilon k) &= \epsilon^{-1} f_\epsilon(p,q,k) \\
&= -\frac{\epsilon^{-1}}{4\pi} \sum_{j,l=1}^{n} e^{-i(px_j - qx_l)} \\
&\quad \times \Big[-\frac{ik}{4\pi}\delta_{jl} - \tilde{G}_k(x_j - x_l) \Big]^{-1}_{jl} \\
&\quad -\frac{1}{4\pi} \sum_{j,l=1}^{n} e^{-i(px_j - qx_l)} \\
&\quad \times [(p\cdot v_j, X_{jl} u_l) + (v_j, X_{jl} u_l q\cdot)] \\
&\quad + \sum_{j,l=1}^{N} e^{-i(px_j - qx_l)} (v_j, Y_{jl} u_l) + o(1), \qquad (3.38)
\end{aligned}
$$

where x and Y are given by (3.23) with $\alpha_j = 0$.

Proof: Follows immediately from Lemma 3.1, Lemma 3.2, and Theorem 3.5.

Theorem 3.8: Let $S_k(\epsilon)$ denote the on-shell scattering matrix for $H(\epsilon)$. We then have

$$
S_{\epsilon k}(\epsilon) = S_{\epsilon,k}, \qquad (3.39)
$$

which implies that the expansion for $S_{\epsilon k}(\epsilon)$ is given in Theorem 3.6.

Proof: We have

$$
\begin{aligned}
&(S_{\epsilon k}(\epsilon)h)(\omega) \\
&\quad = h(\omega) - (2\pi i)^{-1}\epsilon k \textstyle\int d\omega' f(\epsilon,\omega\epsilon k,\omega'\epsilon k,\epsilon k) h(\omega') \\
&\quad = h(\omega) - (2\pi i)^{-1} k \textstyle\int d\omega' f_\epsilon(\omega k,\omega' k,k) h(\omega') \\
&\quad = (S_{\epsilon,k} h)(\omega).
\end{aligned}
$$

IV. THE GENERIC CASE

In Sec. III we were mostly interested in cases II and IV which give rise to the point interaction limit. However neither case II nor case IV is generic, so in this section we will study the generic case, case I, in more detail.

Recall that the Rollnik class R is a Banach space consisting of functions V which satisfy

$$
\|V\|_R^2 \equiv (4\pi)^{-2} \textstyle\int\int |x-y|^{-2} |V(x)V(y)| dx\, dy < \infty. \qquad (4.1)
$$

Consider $(V_1,\ldots,V_N) \in R^N \equiv R \oplus \ldots \oplus R$ with norm

$$
\|(\mathrm{V}_1,\ldots,V_N)\|^2 = \sum_{i=1}^{N} \|V_i\|_R^2. \qquad (4.2)
$$

Then we have the following theorem.

Theorem 4.1: The set A of $(V_1,\ldots,V_N) \in R^N$ such that V_j is in case I for $j = 1,\ldots,N$ is a dense open set in R^N.

Proof: Recall that V_j is in case I iff -1 is not an eigenvalue of $u_j G_0 v_j$. Now

$$
\|u_j G_0 v_j\|_2^2 = \|V_j\|_R^2, \qquad (4.3)
$$

where $\|\cdot\|_2$ is the Hilbert–Schmidt norm on $L^2(\mathbb{R}^3)$. Hence

$$
(V_1,\ldots,V_N) \mapsto (u_1 G_0 v_1,\ldots,u_N G_0 v_n) \qquad (4.4)
$$

is an isometric map from R^N into $\mathbb{B}_2(L^2(\mathbb{R}^3))^N \equiv \mathbb{B}_2(L^2(\mathbb{R}^3)) \oplus \ldots \oplus \mathbb{B}_2(L^2(\mathbb{R}^3))$, where $\mathbb{B}_2$ denotes the set of Hilbert–Schmidt operators and $\mathbb{B}_2(L^2(\mathbb{R}^3))^N$ has norm

$$
\|(B_1,\ldots,B_N)\|_2^2 = \sum_{j=1}^{N} \|B_j\|_2^2. \qquad (4.5)
$$

Since the nonzero eigenvalues of $u_j G_0 v_j$ depend continuously on the Hilbert–Schmidt norm of $u_j G_0 v_j$, and hence on the Rollnik norm of V_j, we see that the set R of all $(V_1,\ldots,V_N) \in R^N$, where V_j is in case I for $j = 1,\ldots,N$ is an open subset of R^N.

That this set is dense, follows from the following argument. Consider Rollnik potentials $V_1,\ldots,V_N$ such that $(V_1,\ldots,V_N) \notin A$, i.e., -1 is an eigenvalue of $u_j G_0 v_j$ for some $j = 1,\ldots,N$. Since $\sigma(u_j G_0 v_j)$ has only zero as a possible accumulation point, we can find a sequence $\{\lambda_n\}$ such that $\lambda_n \to 1$ as $n \to \infty$ and

$$
-\lambda_n^{-1} \notin \bigcup_{j=1}^{N} \sigma(u_j G_0 v_j).
$$

Hence $(\lambda_n V_1,\ldots,\lambda_n V_N) \in A$ for all n. But $(\lambda_n V_1,\ldots,\lambda_n V_N) \mapsto (V_1,\ldots,V_N)$ in R^N which proves that A is dense.

From now on we assume that $V_1,\ldots,V_N$ are in case I. We can then expand the scattering amplitude to third order in ϵ as the next theorem shows.

Theorem 4.2: Assume that $V_1,\ldots,V_N \in R$, V_j has compact support for $j = 1,\ldots,N$, and that $H_j = -\Delta + V_j$ is in case I for $j = 1,\ldots,N$. Then the scattering amplitude for

$$
H_\epsilon = -\Delta + \epsilon^{-2} \sum_{j=1}^{N} \lambda_j(\epsilon) V_j\Big(\frac{1}{\epsilon}(\cdot - x_j)\Big) \qquad (4.6)
$$

is analytic and has the following expansion:

$$
\begin{aligned}
f_\epsilon(p,q,k) = &-\frac{\epsilon}{4\pi}\sum_{j=1}^{N} e^{-i(p-q)x_j}(v_j,R_j u_j) + \frac{\epsilon^2}{4\pi}\sum_{j=1}^{N} e^{-i(p-q)x_j}(i(p\cdot v_j,R_j u_j) \\
&- i(v_j,R_j q\cdot u_j) + \lambda_j'(0)(v_j,R_j u_j)) + \sum_{j,l=1}^{N} e^{-i(px_j - qx_l)}(v_j,(R(M+T_0)R)_{jl} u_l) + \frac{\epsilon^3}{4\pi}\Big[\frac{1}{2}\sum_{j=1}^{N} e^{-i(p-q)x_j}(\lambda_j''(0)(v_j,R_j u_j) \\
&+ (v_j,\tilde{R}_j(p,q)u_j) + 2i\lambda_j'(0)((p\cdot v_j,R_j u_j) - (v_j,R_j q\cdot v_j))) + \sum_{j,l=1}^{N} e^{-i(px_j - qx_l)}(i(p\cdot v_j,(R(M+T_0)R)_{jl} u_l) \\
&- (v_j,(R(M+T_0)R)_{jl} q\cdot u_l) + (v_j,[R(M+T_0)R(M+T_0)R - R(W+U)R]_{jl} u_l) \\
&+ \lambda_l'(0)(v_j,(R(M+T_0)R)_{jl} u_l))\Big] + o(\epsilon^2), \qquad (4.7)
\end{aligned}
$$

where $\tilde{R}_j(p,q)$ has kernel

$$\tilde{R}_j(p,q)(x,y) = u_j(x)G_0(x-y)v_j(y)(px-qy)^2 \qquad (4.8)$$

and the operators R, M, W, T_0, U are given by (3.20), (4.11), (4.11), (3.24), and (3.25), respectively.

Proof: From the definition of the scattering amplitude we see that we have to expand $(1+B_{\epsilon,k})^{-1}$ to second order in ϵ, thus extending Theorem 3.4 to next order when all the potentials are in case I.

Using the notation from Proposition 3.3 and Theorem 3.4 we have

$$1+B_{\epsilon,k} = 1+S_\epsilon+\epsilon T_\epsilon, \qquad (4.9)$$

where now

$$1+S_\epsilon = 1+S_0+\epsilon M+\epsilon^2 W+O(\epsilon^2), \qquad (4.10)$$

and $M=[\delta_{jl}M_j]$, $W=[\delta_{jl}W_j]$, and

$$M_j = \lambda_j'(0)u_jG_0u_j + \frac{ik}{4\pi}|u_j\rangle\langle v_j|,$$

$$W_j = \frac{1}{2}\lambda_j''(0)u_jG_0v_j + \frac{ik}{4\pi}\lambda_j'(0)|u_j\rangle\langle v_j| - \frac{k^2}{16\pi}D_j, \qquad (4.11)$$

where

$$D_j(x,y) = u_j(x)|x-y|v_j(y). \qquad (4.12)$$

Hence

$$\begin{aligned}(1+B_{\epsilon,k})^{-1} &= (1+S_0+\epsilon(M+T_0)+\epsilon^2(W+U)+o(\epsilon^2))^{-1}\\ &= (1+\epsilon(1+S_0)^{-1}[(M+T_0)+\epsilon(W+U)+o(\epsilon)])^{-1}\\ &\quad\times(1+S_0)^{-1}\\ &= R-\epsilon R(M+T_0)R+\epsilon^2\\ &\quad\times[-R(W+U)R+R(M+T_0)\\ &\quad\times R(M+T_0)R]+o(\epsilon^3).\end{aligned} \qquad (4.13)$$

Using now the definition of the scattering amplitude $f_\epsilon(p,q,k)$, viz.

$$f_\epsilon(p,q,k) = -\frac{\epsilon}{4\pi}\sum_{j,l=1}^{N}(e^{ip(\epsilon\cdot+x_j)}, v_j(1+B_{\epsilon,k})_{jl}^{-1} \times\lambda_l(\epsilon)u_le^{iq(\epsilon\cdot+x_l)}), \qquad (4.14)$$

together with (4.13) we obtain the stated expansion.

We end this section with a corollary.

Corollary 4.3: Assume in addition to the assumptions in Theorem 4.2 that

$$V_j(-x) = V_j(x), \qquad (4.15)$$

for $j=1,\ldots,N$ and

$$\lambda_j(\epsilon)\equiv 1, \qquad (4.16)$$

for $j=1,\ldots,N$. Then the scattering amplitude has the expansion

$$\begin{aligned}f_\epsilon(p,q,k) = &-\frac{\epsilon}{4\pi}\sum_{j=1}^{N}e^{-i(p-q)x_j}a_j\\ &-\frac{\epsilon^2}{4\pi}\sum_{j,l=1}^{N}e^{-i(px_j-qx_l)}a_j\\ &\times\Big[-\frac{ik}{4\pi}\delta_{jl}-\tilde{G}_k(x_j-x_l)\Big]a_l\\ &+\frac{\epsilon^2}{4\pi}\sum_{j=1}^{N}e^{-i(p-q)x_j}(b_j(p,q)+c_j)\\ &+\sum_{j,l,m=1}^{N}e^{-i(px_j-qx_l)}\Big(\frac{ik}{4\pi}\delta_{jm}+\tilde{G}_k(x_j-x_m)\Big)\\ &\times\Big(\frac{ik}{4\pi}\delta_{ml}+\tilde{G}_k(x_m-x_l)\Big)a_ja_ma_l+o(\epsilon^2),\end{aligned} \qquad (4.17)$$

where in addition to the notation used in Theorem 4.2 we have used

$$\begin{aligned}a_j &= (v_j, r_ju_j),\\ b_j(p,q) &= (v_j,\tilde{R}_j(p,q)u_j),\\ c_j &= (k^2/16)(v_j,R_jD_jR_ju_j).\end{aligned} \qquad (4.18)$$

ACKNOWLEDGMENTS

We are indebted to Professor S. Albeverio and Professor F. Gesztesy for valuable discussions and to Professor L. Streit for his kind invitation to Zentrum für interdisziplinäre Forschung, Universität Bielefeld, where part of this work was done.

[1] S. Albeverio, F. Gesztesy, H. Holden, and R. Høegh-Krohn, *Solvable Models in Quantum Mechanics* (book in preparation).
[2] S. Albeverio, J. E. Fenstad, and R. Høegh-Krohn, Trans. Am. Math. Soc. **252**, 275 (1979).
[3] A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, J. Math. Phys. **21**, 2376 (1980), p. 106.
[4] A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, Commun. Math. Phys. **77**, 87 (1980).
[5] E. Nelson, Bull. Am. Math. Soc. **83**, 1165 (1977).
[6] C. N. Friedman, J. Funct. Anal. **10**, 346 (1972).
[7] S. Albeverio and R. Høegh-Krohn, J. Operator Theory **6**, 313 (1981).
[8] S. Albeverio, F. Gesztesy, and R. Høegh-Krohn, Ann. Inst. H. Poincaré A **37**, 1 (1982).
[9] H. Holden, R. Høegh-Krohn, and S. Johannesen, Adv. Appl. Math. **4**, 402 (1983).
[10] H. Holden, R. Høegh-Krohn, and S. Johannesen, "The Short-Range Expansion in Solid State," to appear in Ann. Inst. H. Poincaré A.
[11] A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[12] V. I. Karloukovski, Rep. Math. Phys. **10**, 87 (1976).
[13] L. E. Thomas, "Scattering from point interaction," in *Lecture Notes in Physics*, Vol. 130, edited by J. A. De Santo, A. W. Saenz, and W. W. Zachary (Springer-Verlag, Berlin, 1980), pp. 159–162.
[14] B. S. Pavlov and M. D. Faddeev, Theor. Math. Phys. **55**, 485 (1983).
[15] S. Albeverio, R. Høegh-Krohn, and M. Mebkhout, J. Math. Phys. **25**, 1327 (1984).
[16] A. M. Berthier, *Spectral Theory and Wave Operators for the Schrödinger Equation* (Pitman, Boston, 1982).
[17] S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and W. Kirsch, "On point interactions in one dimension," to appear in J. Operator Theory.
[18] H. Holden, "Konvergens mot punkt-interaksjoner," Cand. real. thesis, University of Oslo, 1981 (in Norwegian).

The $\mathscr{H}\mathscr{H}$ equation in arbitrary canonical coordinates

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Pue., Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 1 February 1984; accepted for publication 24 May 1984)

Extending previous results, the equation which determines the algebraically special solutions to the Einstein vacuum field equations is given in an arbitrary system of coordinates adapted to the congruence of totally null two-dimensional surfaces that these space-times possess. The action of the transformations which relate any pair of these coordinate systems is also given.

I. INTRODUCTION

By considering the analytic continuation of Einstein's field equations, Plebański and Robinson[1] showed that all the vacuum solutions whose conformal curvature is algebraically special from one side (i.e., the self-dual or the anti-self-dual part of the Weyl tensor is algebraically special) are determined by the solutions of a single second-order partial differential equation for a scalar function—the $\mathscr{H}\mathscr{H}$ equation. By using the spinorial notation introduced in Ref. 2 for these space-times, in Ref. 3 it was shown that this reduction can be deduced from the Bianchi identities and the existence of a congruence of totally null two-dimensional surfaces. The extension of this reduction process to include some sources under certain restrictions was also given there.

Integration of the $\mathscr{H}\mathscr{H}$ equation would allow us to find all the algebraically special (real) solutions of the Einstein vacuum field equations. However, by now, the systematic efforts made towards this goal have not been successful. Nevertheless, since the metric of these space-times is given in terms of, essentially, a single unknown function, it has been possible to investigate, using this approach, some properties such as the Killing vectors,[4] the $D(k,0)$ Killing spinors,[5] and the massless spinor fields[3,6] on these backgrounds.

In the references cited above the computations were made in systems of coordinates—canonical coordinates—adapted to the congruence of null surfaces already mentioned; however, the expression for the $\mathscr{H}\mathscr{H}$ equation given there is valid in a restricted class of these coordinate systems, in which a lot of gauge-dependent terms vanish, simplifying, in a sense, the form of the equation. With that restriction there is still a considerable freedom in the choice of the canonical coordinates,[7] but the $\mathscr{H}\mathscr{H}$ equation imposes a strong condition on the unknown function. On the other hand, maintaining in the $\mathscr{H}\mathscr{H}$ equation all the gauge-dependent terms, some parts of the unknown function can be accommodated in them, simplifying, in this way, the form of the solution (see Sec. IV). In this paper the $\mathscr{H}\mathscr{H}$ equation, applicable to the case of vacuum with cosmological constant, is given in an arbitrary system of canonical coordinates. The effect of an arbitrary change of canonical coordinates is also given.

It is also possible to avoid the use of a particular set of coordinates by working with covariant expressions. A procedure to obtain such expressions for this kind of space-times is given in Ref. 8. However, in order to find explicit solutions it is necessary, at one point or another, to introduce coordinates. The canonical coordinates, being distinguished by the geometry of the space-time, lead to relatively simple expressions.

In Sec. II we give the expression for the metric and the $\mathscr{H}\mathscr{H}$ equation in an arbitrary set of canonical coordinates. In Sec. III the group of coordinate transformations which preserve the form of the metric, and its effect on the objects that appear in the metric, is given. We use throughout the spinorial notation with the conventions $\psi_A = \epsilon_{AB}\ \psi^B$, $\psi^B = \psi_A \epsilon^{AB}$ for *all* the spinorial indices and similarly for the dotted ones.

II. THE $\mathscr{H}\mathscr{H}$ EQUATION

For any algebraically special solution of the Einstein vacuum field equations one can find, locally, coordinate functions $q^{\dot A}$, $p^{\dot A}$ such that the metric has the form[1,3]

$$ds^2 = -\tfrac{1}{2} g_{A\dot B} \otimes g^{A\dot B}, \tag{2.1a}$$

where

$$g^{2\dot A} = -\sqrt{2}\phi^{-2}\, dq^{\dot A}, \qquad g^{1\dot A} = -\sqrt{2}(dp^{\dot A} - Q^{\dot A\dot B}\, dq_{\dot B}). \tag{2.1b}$$

Then, from Einstein's equations, it follows that ϕ and $Q_{\dot A\dot B}$ are given in terms of some arbitrary functions of $q^{\dot A}$ only and a single function which has to satisfy a second-order partial differential equation with quadratic nonlinearities, the so-called hyperheavenly, or $\mathscr{H}\mathscr{H}$, equation. In particular, the function ϕ is given by

$$\phi = J_{\dot A} p^{\dot A} + \kappa, \tag{2.2}$$

where $J_{\dot A}$ and κ are functions of $q^{\dot A}$ only. In Refs. 1–3 two possible cases were distinguished, called case I and case II according to whether $J_{\dot A}$ is zero or not. There, using the available freedom in choosing the coordinates $q^{\dot A}$, $p^{\dot A}$, in case I, κ was made equal to 1 and, in case II, $J_{\dot A}$ and κ were made constant. This leads to a simplification in the computations but restricts the coordinate transformations which maintain the metric and the $\mathscr{H}\mathscr{H}$ equation form-invariant.

Here we will eliminate the above-mentioned restriction treating the case of vacuum with cosmological constant. We will make use of the general expressions given in Ref. 3, presenting the cases I and II separately.

 0022-2488/85/010152-03$02.50

A. Case I

Taking $G_{\dot{A}}$ equal to zero and $R = -4\lambda$ in Eqs. (3.9b) and (3.26) of Ref. 3 we get

$$Q_{\dot{A}\dot{B}} = -\partial_{\dot{A}}\partial_{\dot{B}}\Theta - \tfrac{2}{3}\kappa^2 L_{(\dot{A}}p_{\dot{B})} + (\lambda/3)\kappa^{-2}p_{\dot{A}}p_{\dot{B}}, \quad (2.3)$$

where $\partial_{\dot{A}} = \partial/\partial p^{\dot{A}}$, Θ is some function, $L_{\dot{A}} = L_{\dot{A}}(q^{\dot{R}})$, and λ is the cosmological constant. Then, by a procedure analogous to that followed in Ref. 3, we find that the Einstein equations reduce to [9]

$$\begin{aligned}&\tfrac{1}{2}(\partial^{\dot{A}}\partial^{\dot{B}}\Theta)\partial_{\dot{A}}\partial_{\dot{B}}\Theta - \partial^{\dot{A}}\Theta_{,\dot{A}} - \kappa^2 L^{\dot{A}}\partial_{\dot{A}}\Theta + \tfrac{1}{18}(\kappa^2 L_{\dot{A}}p^{\dot{A}})^2\\ &+ \tfrac{2}{3}\kappa^2 L^{\dot{A}}p^{\dot{B}}\partial_{\dot{A}}\partial_{\dot{B}}\Theta - \tfrac{1}{6}\kappa^{-2}(\kappa^4 L_{\dot{A}})_{,\dot{B}}p^{\dot{A}}p^{\dot{B}}\\ &- (\lambda/3)\kappa^{-2}p^{\dot{A}}p^{\dot{B}}\partial_{\dot{A}}\partial_{\dot{B}}\Theta + \lambda\kappa^{-2}(p^{\dot{A}}\partial_{\dot{A}}\Theta - \Theta)\\ &+ (\ln\kappa)_{,\dot{A}}\partial^{\dot{A}}\Theta + \tfrac{1}{3}\kappa^{-1}\kappa_{,\dot{A}\dot{B}}p^{\dot{A}}p^{\dot{B}} = \kappa^2(N_{\dot{A}}p^{\dot{A}} + \gamma),\end{aligned} \quad (2.4)$$

where ${}_{,\dot{A}} \equiv \partial/\partial q^{\dot{A}}$ and $N_{\dot{A}}$ and γ are functions of $q^{\dot{R}}$ only.

B. Case II

With $G_{\dot{A}} = 0$ and $R = -4\lambda$, from Eqs. (3.9a) and (3.23) of Ref. 3 we obtain

$$\begin{aligned}Q^{\dot{A}\dot{B}} = &-\phi^3\partial^{(\dot{A}}\phi^{-2}\partial^{\dot{B})}W + 4K^{\dot{C}}\phi_{,\dot{C}}J^{(\dot{A}}K^{\dot{B})}\\ &- 2\phi K^{\dot{C}}J^{(\dot{A}}_{,\dot{C}}K^{\dot{B})} + [\mu\phi^3 + \lambda/6 + \phi K^{\dot{C}}J^{\dot{D}}_{,\dot{C}}J_{\dot{D}}\\ &- J^{\dot{C}}\phi_{,\dot{C}}]K^{\dot{A}}K^{\dot{B}},\end{aligned} \quad (2.5)$$

where $K_{\dot{A}}$ is a spinor such that $K^{\dot{A}}J_{\dot{A}} = 1$, W is some function and $\mu = \mu(q^{\dot{A}})$. Then, by using repeatedly the identity

$$K^{\dot{A}}J_{\dot{B}} - K_{\dot{B}}J^{\dot{A}} = \delta^{\dot{A}}_{\dot{B}}, \quad (2.6)$$

after a lengthy computation along the lines indicated in Refs. 2 and 3 we find that the Einstein vacuum field equations require[9]

$$\begin{aligned}&\tfrac{1}{2}\phi^4(\partial^{\dot{A}}\phi^{-2}\partial^{\dot{B}}W)\partial_{\dot{A}}\phi^{-2}\partial_{\dot{B}}W - \phi^{-1}\partial^{\dot{A}}W_{,\dot{A}}\\ &- (\lambda/6)\phi^{-1}\partial_\phi\partial_\phi W - \mu\phi^4\partial_\phi\phi^{-1}\partial_\phi\phi^{-1}W\\ &+ \tfrac{1}{2}\eta[\eta J^{\dot{A}} - (\phi+\kappa)K^{\dot{A}}]\mu_{,\dot{A}} + (\phi^{-1}J^{\dot{A}}\phi_{,\dot{A}}\\ &+ K^{\dot{A}}J^{\dot{B}}_{,\dot{A}}J_{\dot{B}})\partial_\phi\partial_\phi W\\ &- 2(2\phi^{-1}K^{\dot{A}}\phi_{,\dot{A}} + K^{\dot{A}}J^{\dot{B}}_{,\dot{A}}K_{\dot{B}})J^{\dot{C}}\partial_{\dot{C}}\partial_\phi W\\ &- 2\phi^{-1}K^{\dot{A}}J^{\dot{B}}_{,\dot{A}}J_{\dot{B}}\partial_\phi W\\ &- \phi^{-1}K^{\dot{A}}J^{\dot{B}}_{,\dot{A}}K_{\dot{B}}J^{\dot{C}}\partial_{\dot{C}}W - \tfrac{3}{2}\phi^{-1}K^{\dot{A}}K^{\dot{B}}\phi_{,\dot{A}\dot{B}}\\ &- \phi^{-1}K^{\dot{A}}K^{\dot{B}}_{,\dot{A}}\phi_{\dot{B}} - 2\phi^{-1}K^{\dot{A}}_{,\dot{A}}K^{\dot{B}}\phi_{,\dot{B}}\\ &+ (\lambda/6)\phi^{-1}K^{\dot{A}}K^{\dot{B}}_{,\dot{A}}K_{\dot{B}} - \mu(2\eta^2K^{\dot{A}}J^{\dot{B}}_{,\dot{A}}J_{\dot{B}}\\ &- \tfrac{1}{2}\eta^2J^{\dot{A}}_{,\dot{A}} + \phi K^{\dot{A}}\eta_{,\dot{A}} + \phi\eta K^{\dot{A}}_{,\dot{A}} + \tfrac{1}{2}K_{\dot{A},\dot{B}}p^{\dot{A}}p^{\dot{B}})\\ &= N_{\dot{A}}p^{\dot{A}} + \gamma,\end{aligned} \quad (2.7)$$

where $\partial_\phi \equiv K^{\dot{A}}\partial_{\dot{A}}$, $\eta \equiv K^{\dot{A}}p_{\dot{A}}$ and $N_{\dot{A}}$ and γ are functions of $q^{\dot{A}}$ only.

III. COORDINATE TRANSFORMATIONS

In place of the coordinates $q^{\dot{A}}$, any other pair of independent functions $q'^{\dot{A}} = q'^{\dot{A}}(q^{\dot{B}})$ can be used. Therefore the matrix $(T^{\dot{A}}{}_{\dot{B}}) \equiv (\partial q'^{\dot{A}}/\partial q^{\dot{B}})$ must be nonsingular. Simultaneously, the function ϕ can be replaced by $\phi' = \rho^{-1/2}\phi$, where ρ is a nonvanishing function of $q^{\dot{A}}$ only. Then, in order to maintain the metric form-invariant, the new "longitudinal" coordinates $p'^{\dot{A}}$ must be given by

$$p'^{\dot{A}} = -\rho^{-1}T_{\dot{B}}^{-1\dot{A}}p^{\dot{B}} + \sigma^{\dot{A}}, \quad (3.1)$$

where $(T^{-1\dot{A}}{}_{\dot{B}})$ denotes the inverse matrix of $(T^{\dot{A}}{}_{\dot{B}})$ and $\sigma^{\dot{A}}$ are arbitrary functions of $q^{\dot{A}}$ only and

$$Q'^{\dot{A}\dot{B}} = \rho^{-1}T_{\dot{C}}^{-1\dot{A}}T_{\dot{D}}^{-1\dot{B}}Q^{\dot{C}\dot{D}} - T^{-1\dot{C}(\dot{A}}\partial p'^{\dot{B})}/\partial q^{\dot{C}}. \quad (3.2)$$

Writing $\phi' = J'_{\dot{A}}p'^{\dot{A}} + \kappa'$ [see Eq. (2.2)], in view of (3.1) we have

$$\begin{aligned}J'_{\dot{A}} &= -\rho^{1/2}T_{\dot{A}}{}^{\dot{B}}J_{\dot{B}},\\ \kappa' &= \rho^{-1/2}\kappa + \rho^{1/2}T_{\dot{A}}{}^{\dot{B}}J_{\dot{B}}\sigma^{\dot{A}}.\end{aligned} \quad (3.3)$$

We shall now list the transformation properties of the other objects introduced in the integration of the field equations.

A. Case I

Expressing $Q'_{\dot{A}\dot{B}}$ in terms of primed quantities as in Eq. (2.3) and using Eqs. (3.1)–(3.3) (with $J_{\dot{A}} = 0$) we obtain

$$\kappa'^2L'_{\dot{A}} = T^{-1\dot{B}}{}_{\dot{A}}[\kappa^2L_{\dot{B}} - (\ln\rho^{3/2}T)_{,\dot{B}}] + \lambda\rho\kappa^{-2}\sigma_{\dot{A}}, \quad (3.4)$$

$$\begin{aligned}\rho^2T^2\Theta' = &\rho^{-1}\Theta + \tfrac{1}{6}\rho^{-1}T^{-1}_{\dot{A}\dot{D}}T^{\dot{D}}{}_{\dot{B},\dot{C}}p^{\dot{A}}p^{\dot{B}}p^{\dot{C}}\\ &- \tfrac{1}{6}\lambda\rho\kappa^{-2}(T^{\dot{A}}{}_{\dot{B}}\sigma_{\dot{A}}p^{\dot{B}})^2 + \tfrac{1}{2}T^{\dot{C}}{}_{\dot{A}}\sigma_{\dot{C},\dot{B}}p^{\dot{A}}p^{\dot{B}}\\ &- \tfrac{1}{3}T^{\dot{C}}{}_{\dot{A}}\sigma_{\dot{C}}[\kappa^2L_{\dot{B}} - (\ln\rho^{3/2}T)_{,\dot{B}}]p^{\dot{A}}p^{\dot{B}}\\ &+ a_{\dot{A}}p^{\dot{A}} + b,\end{aligned} \quad (3.5)$$

where $T \equiv \det(T^{\dot{A}}{}_{\dot{B}})$ and $a_{\dot{A}}$ and b are functions of $q^{\dot{A}}$ only.

B. Case II

Taking $K'_{\dot{A}} = \rho^{-1/2}T^{-1\dot{B}}{}_{\dot{A}}K_{\dot{B}}$ [cf. Eq. (3.3)] and expressing $Q'_{\dot{A}\dot{B}}$ in a form analogous to (2.5), from Eqs. (3.1) and (3.2) we get

$$\mu' = \rho^{3/2}\mu, \quad (3.6)$$

$$\begin{aligned}T^2W' = &\rho^{-5/2}W + a\phi^3 + \tfrac{1}{2}\rho^{-5/2}(K^{\dot{A}}T^{\dot{B}}{}_{\dot{A}}J^{\dot{C}}T^{-1\dot{D}}{}_{\dot{B},\dot{C}}J_{\dot{D}}\\ &+ \tfrac{1}{2}\rho^{-1}J^{\dot{A}}\rho_{,\dot{A}})\eta^2 + \tfrac{1}{2}\rho^{-7/2}K^{\dot{A}}\rho_{,\dot{A}}\eta\phi\\ &- \tfrac{1}{2}\rho^{-5/2}(\phi+\kappa)K^{\dot{A}}T^{\dot{B}}{}_{\dot{A}}K^{\dot{C}}T^{-1\dot{D}}{}_{\dot{B},\dot{C}}p_{\dot{D}}\\ &+ \rho^{-3/2}(2K^{\dot{A}}\sigma_{\dot{B},\dot{A}}T^{\dot{B}}{}_{\dot{C}}J^{\dot{C}} - \tfrac{1}{2}J^{\dot{A}}\sigma_{\dot{B},\dot{A}}T^{\dot{B}}{}_{\dot{C}}K^{\dot{C}}\\ &- \rho^{-2}\kappa K^{\dot{A}}\rho_{,\dot{A}} - \rho^{-1}\kappa J^{\dot{A}}T^{\dot{B}}{}_{\dot{A}}K^{\dot{C}}T^{-1\dot{D}}{}_{\dot{B},\dot{C}}K_{\dot{D}})\eta\\ &- \tfrac{1}{2}\rho^{-3/2}T^{\dot{A}}{}_{\dot{B}}p^{\dot{B}}K^{\dot{C}}\sigma_{\dot{A},\dot{C}} + b,\end{aligned} \quad (3.7)$$

where a and b are arbitrary functions of $q^{\dot{A}}$ only.

In case II, due to the ambiguity in the definition of $K_{\dot{A}}$, there exists an additional transformation available. If $K'_{\dot{A}}$ also satisfies the condition $K'^{\dot{A}}J_{\dot{A}} = 1$, then $K'_{\dot{A}} = K_{\dot{A}} + \zeta J_{\dot{A}}$, where ζ is a function of $q^{\dot{A}}$ only. Since this transformation does not involve a change of coordinates, $Q_{\dot{A}\dot{B}}$ remains invariant. Writing $Q_{\dot{A}\dot{B}}$ in Eq. (2.5) in terms of $K'_{\dot{A}}$ and a function W', with the other objects unchanged, we find

$$\begin{aligned}W' = &W + \tfrac{1}{4}\zeta\mu\phi^3(\zeta\phi - 2\eta) - \tfrac{1}{2}[\zeta^3J^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}J_{\dot{B}}\\ &+ \zeta^2(K^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}J_{\dot{B}} + 2J^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}K_{\dot{B}}) + 2\zeta K^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}K_{\dot{B}}]\phi^2\\ &- 2\zeta\phi K^{\dot{A}}\phi_{,\dot{A}} - \tfrac{3}{2}\zeta^2\phi J^{\dot{A}}\phi_{,\dot{A}}\\ &+ \zeta(J^{\dot{A}}\phi_{,\dot{A}} + \phi J^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}K_{\dot{B}})\eta\\ &+ \tfrac{1}{2}\zeta J^{\dot{A}}J^{\dot{B}}{}_{,\dot{A}}J_{\dot{B}}\eta^2 - (\lambda/12)\zeta(\zeta\phi - 2\eta) + a\phi^3 + b,\end{aligned} \quad (3.8)$$

where a and b are functions of q^A only.

IV. CONCLUSIONS

The form of the $\mathscr{H}\mathscr{H}$ equation given here is invariant under the full group of coordinate transformations which maintain the metric (2.1) form-invariant. By specializing to particular cases, the action of this group can be used to simplify the equations as convenient.

An example of the simplification which can be obtained by allowing $J_{\dot{A}}$ and κ to depend on q^A is provided by the type N vacuum metric[10]

$$\begin{aligned} ds^2 = {} & 2r^2(d\zeta - f\,dt)(d\bar{\zeta} - \bar{f}\,dt) \\ & + 2\,dt\,[dr + (r/2)(f_\zeta + \bar{f}_{\bar{\zeta}})dt\,], \end{aligned} \tag{4.1}$$

where $f = f(\zeta, t)$ is an arbitrary complex function. By using the procedure outlined in Ref. 3, the metric (4.1) can be brought to the form (2.1). A natural choice of the canonical coordinates is given by $q^{\dot{1}} = -\zeta$, $q^{\dot{2}} = t$, $p^{\dot{1}} = 1/r + f\bar{\zeta}$, $p^{\dot{2}} = -\bar{\zeta}$. Taking $\phi = 1/r$, from Eq. (2.2) we have, $J_{\dot{1}} = 1$, $J_{\dot{2}} = f$, $\kappa = 0$. Then, choosing $K_{\dot{1}} = 0$, $K_{\dot{2}} = -1$, we find $\mu = 0$ and

$$W = -\frac{1}{2}\int^{-p^{\dot{2}}} \bar{f}(s, q^{\dot{2}})ds + \frac{3}{4} f_{,\dot{1}}(p^{\dot{2}})^2. \tag{4.2}$$

On the other hand, in a system of canonical coordinates in which $J_{\dot{A}} = \text{const}$, the function W takes a very involved form.

The self-dual part of the Weyl tensor of the metric (2.1) is algebraically special, while the spinorial components of the anti-self-dual part with respect to the tetrad (2.1b) are given by

$$C_{\dot{A}\dot{B}\dot{C}\dot{D}} = \phi^3\,\partial_{\dot{A}}\,\partial_{\dot{B}}\,\partial_{\dot{C}}\,\partial_{\dot{D}}\,\widetilde{W}, \tag{4.3}$$

where $\widetilde{W} = \kappa^{-1}\Theta$ in case I and $\widetilde{W} = W - (\mu/4)\phi^2\eta^2$ in case II. If we require the spinor (4.3) to be also algebraically special, which is necessary for a real solution, then this restriction gives additional information concerning the form of $\widetilde{W}$ (see Ref. 11).

[1] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976). The details can be found in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[2] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).

[3] G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).

[4] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **19**, 760 (1978); S. A. Sonnleitner and J. D. Finley, III, J. Math. Phys. **23**, 116 (1982); G. F. Torres del Castillo, J. Math. Phys. **25**, 1980 (1984).

[5] G. F. Torres del Castillo, Commun. Math. Phys. **92**, 485 (1984).

[6] A. García D., J. F. Plebański, and I. Robinson, Gen. Relativ. Gravit. **8**, 841 (1977); J. D. Finley, III and J. F. Plebański, J. Math. Phys. **18**, 1662 (1977).

[7] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **22**, 667 (1981).

[8] G. F. Torres del Castillo, J. Math. Phys. **25**, 342 (1984).

[9] In Ref. 2 there are some slight errors in the λ-dependent terms that appear in the hyperheavenly equation.

[10] A. García D. and J. F. Plebański, J. Math. Phys. **22**, 2655 (1981).

[11] J. F. Plebański and G. F. Torres del Castillo, J. Math. Phys. **23**, 1349 (1982).

Magnetic generalization of the Kerr–Newman metric

Alberto García Díaz[a)]

Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México D. F., Mexico

(Received 6 March 1984; accepted for publication 15 June 1984)

The exact expression of the magnetized Kerr–Newman metric is presented. The obtained solution possesses five continuous arbitrary parameters; mass, angular momentum, electric charge, and electric and magnetic field parameters.

I. INTRODUCTION

The main purpose of the present work is to give the exact form of a magnetized Kerr–Newman (KN) metric, extending in this way the results concerning Kerr–Newman black holes in a magnetic universe.[1,2] In the paper by Ernst and Wild (Ref. 1), which will be referred to as I, there were presented the exact forms of the electromagnetic field and metrical corrections up to the first order in the magnetic parameter for a magnetic KN field. We succeeded in integrating the whole problem, obtaining the general solution in terms of rational functions. Moreover, by canceling out the "proper" charge in our solution one arrives at the right expression of the magnetized Kerr metric.

We start our study from the KN metric given in the form

$$g = \Delta /P\, dp^2 + \Delta /Q\, dq^2 + f^{-1}PQ\, dT^2 - f[d\phi + W\, dT]^2, \tag{1}$$

where the structural functions are

$$\begin{aligned}
&\Delta = q^2 + p^2, \quad P = a^2 - p^2, \quad Q = a^2 + e^2 - 2mq + q^2,\\
&f = -\Delta^{-1}AP := -\Delta^{-1}[(a^2+q^2)^2 - PQ]P,\\
&W = -A^{-1}[a^2 + q^2 - Q]a^2.
\end{aligned} \tag{2}$$

The complex Ernst potentials Φ and ϵ for the given metric are

$$\begin{aligned}
a\Phi &= e(a^2 - ipq)/(q + ip),\\
a^2\mathscr{E} &= a^2 f - (e^2/\Delta)(a^4 + p^2q^2)\\
&\quad + 2i\{mp[P^2\Delta^{-1} + (3a^2 - p^2)] - e^2pqP\Delta^{-1}\}.
\end{aligned} \tag{3}$$

Executing in the above expressions the coordinate transformations

$$p = a\cos\theta, \quad q = r, \quad \phi = \phi, \quad T \to Ta^{-1}, \tag{4}$$

one arrives just at the starting KN metric of I, with coordinates $\{T, r, \theta, \phi\}$ running the values: $-\infty < T < \infty$, $0 < r < \infty$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. Furthermore, one obtains, applying (4) to (3), the potentials Φ and ϵ given in I [see also Ref. 2, formulas (5.1) and (5.2)]. The orthonormal components of the electromagnetic field for a locally nonrotating observer are derivable from Eqs. (1.14) of I, namely,

$$\begin{aligned}
H_r + iE_r &= (A^{1/2}\sin\theta)^{-1}\frac{\partial\Phi}{\partial\theta},\\
H_\theta + iE_\theta &= -(A^{1/2}\sin\theta)^{-1}Q\frac{\partial\Phi}{\partial r}.
\end{aligned} \tag{5}$$

II. THE MAGNETIZED KN METRIC

To determine the magnetic generalization of a given solution, one accomplishes the Harrison transformations[3]

$$\Phi' = \Psi^{-1}[\Phi + (E + iB)\mathscr{E}], \quad \mathscr{E}' = \Psi^{-1}\mathscr{E}, \tag{6}$$

with

$$\Psi = 1 - 2(E - iB)\Phi - \delta\mathscr{E}, \quad \delta := E^2 + B^2, \tag{7}$$

where the constants E and B are the additional electric and magnetic field parameters. In the studied case, the resulting metric has the form

$$g = |\Psi|^2\{(\Delta/P)dp^2 + (\Delta/Q)dq^2 + f^{-1}PQ\, dT^2\} - |\Psi|^{-2}f[d\phi + W'\, dT]^2, \tag{8}$$

with the function W' constrained to the equation

$$dW' = |\Psi|^2 dW + (i\Delta/AP)[P\, dq(\Psi\bar{\Psi}_{,p} - \bar{\Psi}\Psi_{,p}) - Q\, dp(\Psi\bar{\Psi}_{,q} - \bar{\Psi}\Psi_{,q})]. \tag{9}$$

Note that the common denominator of this equation is A^2P^2. Thus, the sought solution ought to be of the form $\omega = X/AP$, with X being a polynominal in p and q.

The general solution of Eq. (9) can be given as

$$\begin{aligned}
W' = W + A^{-1}\{&4Be\, pQa - 4Ee\, q(a^2 + q^2)a\\
&+ 3\delta e^2[a^4 - q^4 + (a^2 + p^2)Q]\\
&+ 4\delta Be\, Sa^{-1} - 4\delta EeMa^{-1} + \delta^2 Na^{-2}\},
\end{aligned} \tag{10}$$

where the polynomials S, M, and N are

$$\begin{aligned}
S &= a^2(4a^2 + 3e^2)pq^2 + (3a^2 - p^2)q^4p\\
&\quad + a^4(a^2 + p^2)p + e^2p[a^4 + (2a^2 + e^2)p^2]\\
&\quad - 2mp[a^2(a^2 + p^2)q + e^2p^2q + (3a^2 - p^2)q^3],\\
M &= m\{a^4(2a^2 + 3p^2) - 3a^2(a^2 - 3p^2)q^2 - (a^2 - 2p^2)q^4\\
&\quad - a^2(2mq - e^2)(a^2 + 3p^2)\} + e^2q(2a^4 - 3a^2p^2\\
&\quad - p^2q^2) + a^4(a^2 \dot{-} 3p^2)q - (a^2 + p^2)q^5 - 4a^2p^2q^3,
\end{aligned}$$

[a)] Also at Sección de Graduados, Escuela Superior de Ingeniería Mecánica y Eléctrica del IPN, México D. F., Mexico.

$$
\begin{aligned}
N = {} & 2m^2a^2\{a^2[2a^4 + 5a^2p^2 + 2p^4] \\
& + e^2[a^4 + 5a^2p^2 + 2p^4]\} \\
& + (a^2 + e^2)p^2[a^6 + e^2a^2(2a^2 + p^2) + e^4p^2] \\
& + 2mq\{a^4[2a^4 - 7a^2p^2 - 3p^4] \\
& - 2m^2a^2[a^4 + 5a^2p^2 + 2p^4] \\
& + e^2[4a^2(a^4 - 2a^2p^2 - p^4) - e^2p^4]\} \\
& - 2m^2a^2q^2(7a^4 - 17a^2p^2 - 8p^4) + a^6q^2(a^2 + p^2) \\
& + 2e^2q^2[a^2(a^4 + 4a^2p^2 - p^4) + e^2p^2(3a^2 - p^2)] \\
& - 4mq^3[a^6 + 6a^4p^2 + a^2p^4 + 2e^2p^2(3a^2 - p^2)] \\
& - 2m^2q^4[a^4 - 12a^2p^2 + 2p^4] + a^6q^4 \\
& + e^2q^4[a^4 + 6a^2p^2 - 3p^4] \\
& - 2mq^5[3a^4 + 6a^2p^2 - p^4]. \qquad (11)
\end{aligned}
$$

Comparing our results for W' with formulas (2.4) an (5.2) of I, we observe that there is no reason to include in the denominator of W' of I the factor $(r^2 + a^2)$. The same comment applies for the W' of the magnetic Kerr metric shown in I. One arrives at the right expression for the magnetized Kerr metric simply by setting the charge e equal to zero in the metric structure presented here.

ACKNOWLEDGMENT

I highly appreciate the friendly attitude of my colleague Alla N. Volkova encouraging me to integrate the studied problem.

[1] F. J. Ernst and W. J. Wild, J. Math. Phys. **17**, 182 (1976).
[2] F. J. Ernst, J. Math. Phys. **17**, 54 (1976).
[3] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).

An effective algorithm for obtaining polynomials for dimer statistics. Application of operator technique on the topological index to two- and three-dimensional rectangular and torus lattices

Haruo Hosoya and Akiko Motoyama[a)]
Department of Chemistry, Ochanomizu University, Bunkyo-ku, Tokyo 112, Japan

(Received 26 March 1984; accepted for publication 27 July 1984)

Recursion formulas on the numbers of ways for placing an arbitrary number of indistinguishable dumbbells on the various rectangular lattice spaces ($3\times n$, $4\times n$) and torus spaces ($2\times n$, $3\times n$, $2\times 2\times n$) were obtained by the use of the operator technique to the counting polynomial on the topological characteristics (e.g., nonadjacent number, Kekulé number, and topological index). Perfect matching numbers or Kekulé numbers were also obtained for larger lattices such as the $4\times n$ torus and the $2\times 3\times n$ lattice. By deriving these new results the utility of the proposed method for dimer statistics is demonstrated and some of their mathematical features are discussed.

I. INTRODUCTION

Enumeration of the number of ways for placing indistinguishable dumbbells on various periodic lattice spaces has a key role in solving various statistical problems, e.g., adsorption of diatomic molecules on a crystalline surface, magnetic properties of antiferromagnetic metals, stability of ionic crystals,[1,4] etc. Enumeration of the number of the Kekulé structures for aromatic hydrocarbons is mathematically equivalent to this problem,[5,6] but in this paper we are concerned only with the dimer statistics of rectangular lattices. An analytical solution was obtained by Temperley and Fisher[7] and by Kasteleyn[8] for the complete coverage of dimers to a two-dimensional $n\times m$ square lattice, while the solution for imperfect matching (covering) was obtained only for very limited cases of $1\times n$ and $2\times n$ lattices by McQuistan and Lichtman.[9] The only known solution for the three-dimensional lattice was obtained by Hock and McQuistan for the $2\times 2\times n$ lattice.[10]

However, their method turns out to be very tedious and difficult for applying to larger lattices even of two-dimensional cases. From a quite different standpoint the present authors have developed a systematic method for counting various characteristic quantities of graphs, such as the nonadjacent number $p(G,k)$, Z-counting polynomial $Q_G(x)$, and topological index Z_G[11] for the study of electronic[12,13] and thermodynamic[14,15] properties of and for characterization[16] of hydrocarbon molecules. The recursion formulas for these quantities are found to be obtained easily by the use of the operator technique also proposed by the present authors.[17]

In the present paper this enumeration technique is applied to the dimer statistics to get further results on the number of ways for the complete and incomplete covering of several two-dimensional rectangular and torus lattices and preliminary results on three-dimensional lattices.[18] Interesting mathematical relations among the recursion formulas of this family of lattices are found and discussed. No thermodynamic treatment which should follow these results is expanded in this paper.

[a)] Present address: Japan Association for International Chemical Information, Bunkyo-ku, Tokyo 113, Japan.

II. DURAL GRAPH, POLYOMINO, AND POLYCUBE

Instead of treating the rectangular lattice spaces for the dimer statistics we will be concerned with their dual graphs,[4,19] which are obtained by transforming the adjacency relation of the cells into that of points through lines as shown in Fig. 1. A graph G is composed of points and lines, the latter of which must be defined by a pair of the former.[20] We are concerned with connected nondirected graphs without any loop and multiple lines, and particularly with polyomino graphs,[21] or square animals.[22] The problem of arranging indistinguishable dumbbells on an $n\times m$ rectangular lattice space is then transformed into the matching problem (*vide infra*) on the lines of the dual graph of $(n-1)\times(m-1)$ lattice. The dual of a three-dimensional rectangular lattice space of an $n\times m\times l$ polycube is also a three-dimensional $(n-1)\times(m-1)\times(l-1)$ polycube graph.

III. DEFINITIONS OF NONADJACENT NUMBER, Z-COUNTING POLYNOMIAL, AND TOPOLOGICAL INDEX[11]

Define a nonadjacent number $p(G,k)$ as the number of ways for choosing k disconnected lines from G, with $p(G,0)$ being taken as unity. The Z-counting polynomial $Q_G(x)$ is defined as the equation

$$Q_G(x)=\sum_{k=0}^{m} p(G,k)x^k, \tag{1}$$

where m is the maximum number of k. For G with an even number $N=2m$ of points let us denote $p(G,m)$ as $K(G)$ and

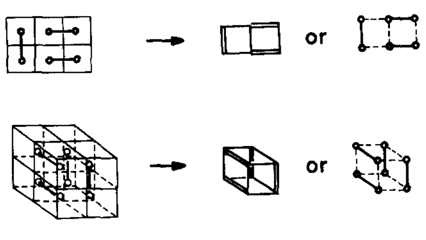
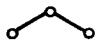
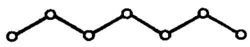


FIG. 1. Representation of the placement of indistinguishable dumbbells on two- and three-dimensional rectangular lattice spaces in terms of dual graphs. Perfect matching patterns are given as examples.

 0022-2488/85/010157-11$02.50

TABLE I. The Z-counting polynomial and the topological index of the path graphs.

N	G	$p(G,k)$[a] $k=0$	1	2	3	4	$Z_G(F_N)$[b]
0	∅	1					1
1		1					1
2		1	1				2
3		1	2				3
4		1	3	1			5
5		1	4	3			8
6		1	5	6	1		13
7		1	6	10	4		21
8		1	7	15	10	1	34

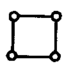
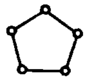
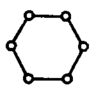
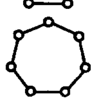
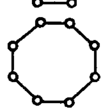
[a] $Q_G(x)$ can be obtained from Eq. (1).
[b] The Z_G values for this series of graphs form the well-known Fibonacci series $\{F_n\}$, such that $F_n = F_{n-1} + F_{n-2}$ with $F_0 = F_1 = 1$.

call it the perfect matching number or Kekulé number. The topological index Z_G is the sum of the $p(G,k)$ numbers, or

$$Z_G = \sum_{k=0}^{m} p(G,k) = Q_G(1). \tag{2}$$

In Tables I–III are given some examples of the $p(G,k)$'s and Z-values for typical smaller graphs relevant to the dimer statistics.

TABLE II. The Z-counting polynomial and the topological index of the path graphs.

N	G	$p(G,k)$ $k=0$	1	2	3	4	$Z_G(L_N)$[a]
3		1	3				4
4		1	4	2			7
5		1	5	5			11
6		1	6	9	2		18
7		1	7	14	7		29
8		1	8	20	16	2	47

[a] The Z_G values for this series of graphs form the well-known Lucas series $\{L_N\}$, with the same recursion formula as F_N but with different initial condition, $L_0 = 2$ and $L_1 = 1$.

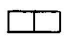
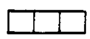
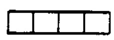
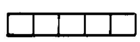
The $p(G,k)$ numbers for the linear polyomino graphs in Table III represent the number of ways for the incomplete and complete covering of k dumbbells on a $2 \times n$ rectangular lattice as derived by McQuistan and Lichtman,[9] and the $Q_G(x)$ polynomial is equivalent to the $f_N(x)$ defined as Eq. (3) in Ref. 10.

The graph-theoretical quantities can be applied to a number of chemical and physical problems.[11–17,19] Here, for dimer statistics several recursion formulas and the operator technique for obtaining these characteristic quantities of larger networks will be introduced.

IV. COMPOSITION PRINCIPLES FOR THE CHARACTERISTIC QUANTITIES

To get the characteristic quantities for larger graphs the use of several composition principles is inevitable. By using them, useful recursion formulas can be obtained for certain

TABLE III. The Z-counting polynomial and the topological index of the linear polyomino graphs ($2 \times n$ lattices).

G	$p(G,k)$ $k=0$	1	2	3	4	5	6	Z_G
	1	4	2					7
	1	7	11	3				22
	1	10	29	26	5			71
	1	13	56	94	56	8		228
	1	16	92	234	263	114	13	733

series of graphs that grow regularly. Choose any line l from G, and let the subgraphs $G–l$ and $G\ominus l$ be, respectively, defined as those obtained from G by deleting only l (retaining both the end points) and by deleting l together with all the lines adjacent to l (see Fig. 2). The $p(G,k)$ number is the sum of the contributions of l-exclusive choice and l-inclusive choice as expressed by the following composition principle (I):

$$p(G,k) = \underset{l\text{-exclusive}}{p(G - l,k)} + \underset{l\text{-inclusive}}{p(G\ominus l,k - 1)}. \tag{3}$$

The argument $k - 1$ in the second term comes from the fact that one line has already been chosen on l and the rest of the $k - 1$ disconnected lines are to be chosen from the graph $G\ominus l$. Accordingly $Q_G(x)$ and Z_G recur as

$$Q_G(x) = Q_{G-l} + xQ_{G\ominus l}(x), \tag{4}$$

$$Z_G = Z_{G-l} + Z_{G\ominus l}. \tag{5}$$

Note the factor x in the second term of Eq. (4).

Choose any point p from G and divide the lines $\{l_p\}$ incident to p into two groups $\{l_p^a\}$ and $\{l_p^b\}$ so that $\{l_p^a\} \cap \{l_p^b\} = 0$ and $|l_p^a| + |l_p^b| = |l_p|$, where a pair of vertical lines indicate the number of the elements in them (see Fig. 3). Let us define the subgraphs $G - \{l_p^a\}$, $G - \{l_p^b\}$, and $G\ominus p$ as in Fig. 3. If $p(G - \{l_p^a\},k)$ and $p(G - \{l_p^b\},k)$ are added to obtain the $p(G,k)$ value, we would overcount by $p(G\ominus p,k)$ ways, and thus we have the composition principle (II) as

$$p(G,k) = p(G - \{l_p^a\},k) + p(G - \{l_p^b\},k) - p(G\ominus p,k), \tag{6}$$

$$Q_G(x) = Q_{G-\{l_p^a\}}(x) + Q_{G-\{l_p^b\}}(x) - Q_{G\ominus p}(x), \tag{7}$$

$$Z_G = Z_{G-\{l_p^a\}} + Z_{G-\{l_p^b\}} - Z_{G\ominus p}. \tag{8}$$

In order to apply this $p(G,k)$ counting method to larger networks one needs to have jumbo composition principles which would be especially useful for G with high symmetry. Let us choose a set of lines $(l_1,l_2,...,l_n)$ from G (see Fig. 4). By repeating similar arguments as above one can derive the following jumbo composition principle (III):

$$\begin{aligned} p(G,k) = {} & p(G - (l_1 + l_2 + \cdots + l_n),k) + \sum_i^n p(G\ominus l_i,k - 1) \\ & - \sum_{i<j}^{n}{}' p(G\ominus(l_i + l_j),k - 2) \\ & + \sum_{i<j<k}^{n}{}' p(G\ominus(l_i + l_j + l_k),k - 3) + \cdots, \end{aligned} \tag{9}$$

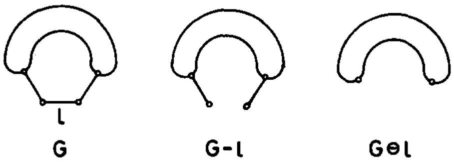


FIG. 2. Composition principle (I). See Eqs. (3)–(5).

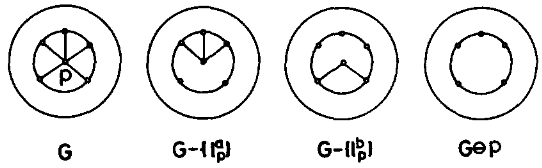


FIG. 3. Composition principle (II). See Eqs. (6)–(8).

$$\begin{aligned} Q_G(x) = {} & Q_{G-(l_1+l_2+\cdots l_n)}(x) + x\sum_i^n Q_{G\ominus l_i}(x) \\ & - x^2 \sum_{i<j}^{n}{}' Q_{G\ominus(l_i+l_j)}(x) \\ & + x^3 \sum_{i<j<k}^{n}{}' Q_{G\ominus(l_i+l_j+l_k)}(x) - \cdots, \end{aligned} \tag{10}$$

$$\begin{aligned} Z_G = {} & Z_{G-(l_1+l_2+\cdots+l_n)} + \sum_i^n Z_{G\ominus l_i} \\ & - \sum_{i<j}^{n}{}' Z_{G\ominus(l_i+l_j)} + \sum_{i<j<k}^{n}{}' Z_{G\ominus(l_i+l_j+l_k)}. \end{aligned} \tag{11}$$

The primes attached to the double summations mean that the summation is to be taken only for disjoint sets of l_i's. An example of using the composition principle (III) is given in Fig. 5 for the cube graph, where two alternative choices of the set of l_i's give the equivalent results.

A computer program making use of these composition principles has been coded into a HITAC M-280H computer of the University of Tokyo. The topological quantities of the lower members of the rectangular lattices and tori were calculated by this program while for those of the higher members the recursion formulas obtained in this study were used through the formula transformation language REDUCE implemented in the same computer.

V. OPERATOR TECHNIQUE FOR RECURSION FORMULA[17]

Let us demonstrate with an example the utility of the operator technique for obtaining the recursion formula of the counting polynomial. We are going to obtain the recursion formula of the $Q_G(x)$ for the linear polyomino graphs, or $2\times n$ rectangular lattices. By a successive application of (I) to the $2\times n$ lattice A_n and to the group of its subgraphs in Fig. 6 we can get a set of four simultaneous recursion formulas

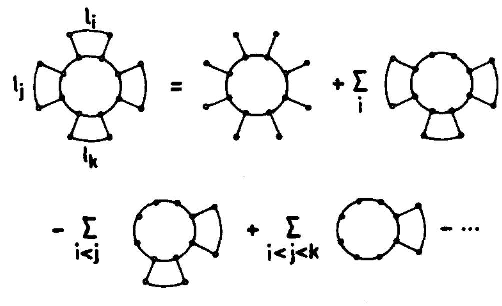


FIG. 4. Composition principle (III). See Eqs. (9)–(11).

$$\begin{cases} (A_n - xA_{n-1}) & -B_{n-1} & & & = 0, \\ & B_n & -C_n & -xD_{n-1} & = 0, \\ A_n & & -(C_n - xC_{n-1}) & & = 0, \\ xA_n & & +C_n & -D_n & = 0, \end{cases} \tag{12}$$

where A_n, etc., represent the corresponding Z-counting polynomials for graph A_n, etc.

Define a stepup operator $\hat{O}$ such that

$$\hat{O}F_{n-1} = F_n \quad (F = A,B,C,D), \tag{13}$$

and we get a set of four simultaneous linear equations

$$\begin{cases} (\hat{O} - x)A_n & -B_n & & & = 0, \\ & \hat{O}B_n & -\hat{O}C_n & -xD_n & = 0, \\ \hat{O}A_n & & -(\hat{O} - x)C_n & & = 0, \\ xA_n & & +C_n & -D_n & = 0. \end{cases} \tag{14}$$

In order to have nontrivial solutions of $\{F_n\}$ for Eq. (14) the coefficient determinant should be zero, namely,

$$\begin{vmatrix} \hat{O} - x & -1 & 0 & 0 \\ 0 & \hat{O} & -\hat{O} & -x \\ \hat{O} & 0 & -(\hat{O} - x) & 0 \\ x & 0 & 1 & -1 \end{vmatrix} = 0. \tag{15}$$

Expansion of this determinant gives the following operator polynomial:

$$\hat{O}^3 - (1 + 2x)\hat{O}^2 - x\hat{O} + x^3 = 0 \quad (2\times n), \tag{16}$$

which, on application to A_{n-3}, leads to the desired recursion formula of the Z-counting polynomial,

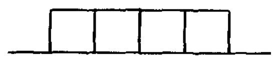
$$A_n(x) = (1 + 2x)A_{n-1}(x) + xA_{n-2}(x) - x^3A_{n-3}(x) \quad (2\times n), \tag{17}$$

which includes the recursion relation of the component $p(G,k)$ numbers,

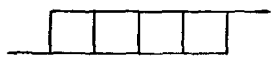
$$p(A_n,k) = p(A_{n-1},k) + 2p(A_{n-1},k-1) + p(A_{n-2},k-1) - p(A_{n-3},k-3) \quad (2\times n), \tag{18}$$

as obtained by McQuistan and Lichtman.[9,23] The Q polynomials of higher members of linear polyomino graphs can successively be obtained by using Eq. (17) or Eq. (18) with the lower three members of A_n given in Table III as the initial condition. Equation (17) is more practical than Eq. (18), since the former yields a whole set of $p(G,k)$ components by a single calculation.

Note that in some cases not all the series of the related graphs have the same recursion relation. For example, although the majority of the Q polynomials for the series of graphs derived from the $2\times n$ lattice, such as



recurs according to Eq. (16) or (17) as $A_n - D_n$ in Fig. 6, the graph



is found to recur according to

$$A_n(x) = (1 + x)A_{n-1}(x) + 2x(1 + x)A_{n-2}(x) + x^2(1 - x)A_{n-3}(x) - x^4A_{n-4}(x) \quad (\overline{2\times n}). \tag{19}$$

The torus $\overline{2\times n}$ obtained by joining the both ends of the $2\times n$ lattice is also found to obey Eq. (19). Note also the fact that Eqs. (17) and (19) are related to each other through the operator relation

$$\hat{O}^4 - (1 + x)\hat{O}^3 - 2x(1 + x)\hat{O}^2 - x^2(1 - x)\hat{O} + x^4 = (\hat{O} + x)\{\hat{O}^3 - (1 + 2x)\hat{O}^2 - x\hat{O} + x^3\},$$

which reveals an interesting algebraic structure of the set of the recursion formulas of the Z-counting polynomial for the rectangular lattice, its torus, and subgraphs.

The recursion formulas of the Kehulé number, $K(G) = p(G,N/2)$, or the perfect matching number, for the

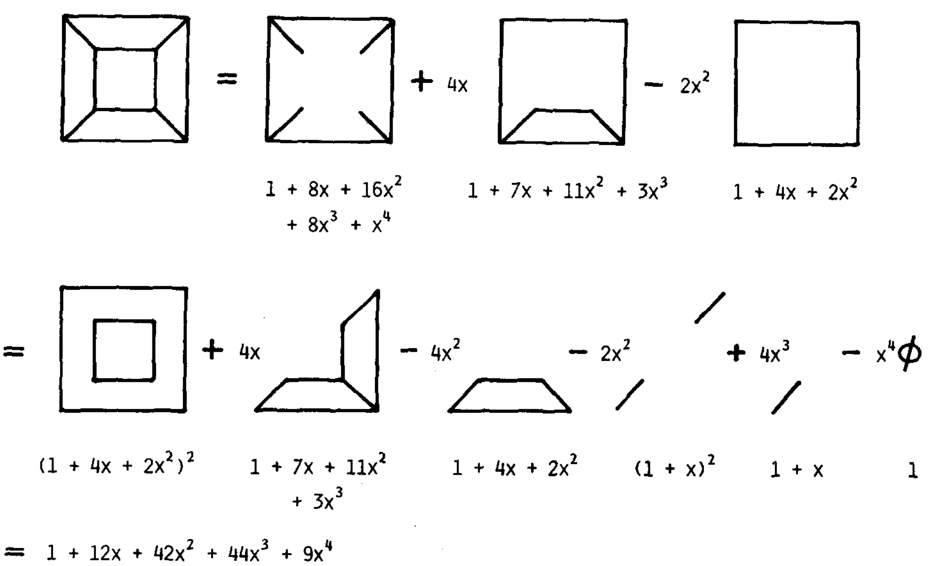


FIG. 5. Examples of composition principle (III) as applied to the cube graph. See Eq. (10). (a) Four lines of the inner square and (b) four lines connecting the inner and outer squares are, respectively, chosen as the sets of l_i's to be deleted.

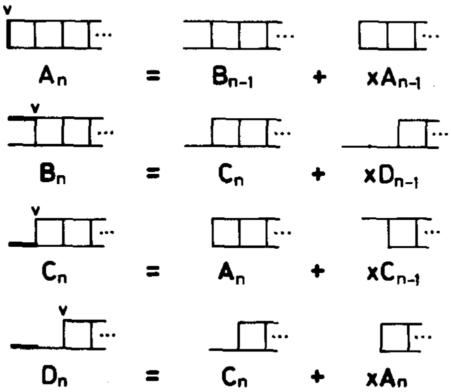


FIG. 6. Simultaneous recursion formulas obtained by decomposing the $2\times n$ lattice. The v mark represents the nth column of each subgraph. The bold line indicates the line to be deleted for applying the composition principles.

$2\times n$ lattice and its torus can be obtained similarly by the use of the operator technique. The results are as follows:

$$K(2\times n)=K_n=K_{n-1}+K_{n-2}, \tag{20}$$

$$K(\overline{2\times n})=K_{\bar n}=K_{\overline{n-1}}+2K_{\overline{n-2}}-K_{\overline{n-3}}-K_{\overline{n-4}}. \tag{21}$$

These two recursion formulas are found to be related to each other again through

$$(\hat O-1)(\hat O^2-\hat O-1)=\hat O^4-\hat O^3-2\hat O^2+\hat O+1.$$

In Table IV are listed the coefficients of the Z-counting polynomials and the Kekulé numbers for the $2\times n$ torus. The topological index Z_G for these graphs recurs as

$$Z_n=3Z_{n-1}+Z_{n-2}-Z_{n-3}\quad (2\times n), \tag{22}$$

$$Z_{\bar n}=2Z_{\overline{n-1}}+4Z_{\overline{n-2}}-Z_{\overline{n-4}}\quad (\overline{2\times n}), \tag{23}$$

which can be obtained by putting $x=1$ into Eqs. (17) and (19), respectively.

VI. APPLICATION TO OTHER LATTICES

A. $3\times n$ lattice

Let a $3\times n$ and its Z-counting polynomial be denoted again as A_n. If one applies the composition principle (I) to A_n successively, there are obtained a number of subgraphs which also form series of graphs with the same recursion relation as A_n. Among them the three series of graphs were found to be important for deriving the recursion relations of the whole system. They are shown in Fig. 7 and denoted as B_n, C_n, and D_n.

First choose A_n, delete one of the bold lines as depicted in Fig. 7, apply (I), and continue this process until we get the recursion relations involving only the A, B, C, and D series. By taking similar procedures for B_n, C_n, and D_n we get the following set of four simultaneous recursion formulas:

$$\begin{cases}(A_n-xA_{n-1})-(B_{n-1}+x^2B_{n-2})-xC_{n-2}=0,\\ (1+x)A_n+x^2A_{n-1}-(B_n-xB_{n-1}\\ \quad -x^3B_{n-2})+xD_{n-1}=0,\\ (2x+x^2)A_n+(1+x)B_n+x^2B_{n-1}-C_n+x^2D_{n-1}=0,\\ (A_n+x^2A_{n-1})+2xB_{n-1}-(D_n-x^3D_{n-2})=0.\end{cases} \tag{24}$$

According to the operator technique the following secular determinant is expanded into the polynomial with respect to the stepup operator $\hat O$:

$$\begin{vmatrix}\hat O^2-x\hat O & -(\hat O+x^2) & -x & 0\\ (1+x)\hat O^2+x^2\hat O & -(\hat O^2-x\hat O-x^3) & 0 & x\hat O\\ (2x+x^2)\hat O & (1+x)\hat O+x^2 & -\hat O & x^2\\ \hat O^2+x^2\hat O & 2x\hat O & 0 & -(\hat O^2-x^3)\end{vmatrix}$$
$$=-\hat O\{\hat O^6-(1+3x)\hat O^5-x(2+7x+5x^2)\hat O^4$$
$$-x^2(1+x-2x^2)\hat O^3+x^4(2+3x+5x^2)\hat O^2$$
$$-x^6(1-x)\hat O-x^9\}=0. \tag{25}$$

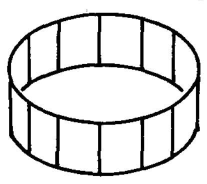
Then by applying the operator in the curly brackets of Eq. (25) to A_{n-6} we get the recursion formula of A_n,

$$\begin{aligned}A_n(x)&=(1+3x)A_{n-1}(x)+x(2+7x+5x^2)A_{n-2}(x)\\ &\quad+x^2(1+x-2x^2)A_{n-3}(x)\\ &\quad-x^4(2+3x+5x^2)A_{n-4}(x)+x^6(1-x)A_{n-5}(x)\\ &\quad+x^9A_{n-6}(x)\quad (3\times n),\end{aligned} \tag{26}$$

which on substitution of $x=1$ yields the recursion formula of the topological index Z_n $(=Z_{3\times n})$ of A_n as

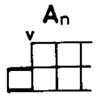
TABLE IV. The Z-counting polynomial and the topological index of the $2\times n$ torus.

	$p(G,k)$										
k	$n=2$	3	4	5	6	7	8	9	10	11	12
0	1	1	1	1	1	1	1	1	1	1	1
1	6	9	12	15	18	21	24	27	30	33	36
2	5	18	42	75	117	168	228	297	375	462	558
3		4	44	145	336	644	1 096	1 719	2 540	3 586	4 884
4			9	95	420	1225	2 834	5 652	10 165	16 940	26 625
5				11	192	1085	3 880	10 656	24 626	50 380	94 128
6					20	371	2 588	11 097	35 645	94 457	218 124
7						29	696	5 823	29 380	108 933	327 840
8							49	1 278	12 535	73 238	309 828
9								76	2 310	26 070	173 984
10									125	4 125	52 752
11										199	7 296
12											324
Z_G	12	32	108	342	1104	3544	11 396	36 626	117 732	378 424	1 216 380



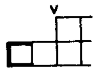
$= B_{n-1} + xA_{n-1} + x^2B_{n-2} + xC_{n-2}$

A_n



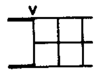
$= (1+x)A_n + xB_{n-1} + xD_{n-1} + x^2A_{n-1} + x^3B_{n-2}$

B_n



$= B_n + xA_n + x^2D_{n-1} + x(1+x)A_n + x^2B_{n-1} + xB_n$

C_n



$= A_n + xB_{n-1} + xB_{n-1} + x^2A_{n-1} + x^3D_{n-2}$

D_n

FIG. 7. Simultaneous recursion formulas obtained by decomposing the $3\times n$ lattice. See the caption of Fig. 6.

$$Z_n = 4Z_{n-1} + 14Z_{n-2} - 10Z_{n-4} + Z_{n-6} \quad (3\times n). \tag{27}$$

For graph A_n with even n the perfect matching number $K_{2n} = K(3\times 2n) = p(A_{2n}, 3n)$ can be shown to recur as

$$K_{2n} = 4K_{2n-2} - K_{2n-4} \quad (3\times 2n), \tag{28}$$

which has already been derived by Klarner and Pollack[24,25] and also by Read[26] together with the results of other higher members. The coefficients of $A_n(x)$, K_{2n}, and Z_n for the smaller members of this series of graphs are given in Table V.

B. $4\times n$ lattice

Quite similarly we can get a set of recursion formulas for the Z-counting polynomials of the $4\times n$ lattice (again denoted as A_n) and its subgraphs as shown in Fig. 8:

$$A_n = B_{n-1} + xC_{n-1}, \tag{29}$$

$$B_n = D_n + xE_n, \tag{30}$$

$$C_n = A_n + 2xF_{n-1} + x^2G_{n-1}, \tag{31}$$

$$D_n = E_n + xH_{n-1} + (x+x^2)I_{n-1} + x^2J_{n-2}, \tag{32}$$

$$E_n = (1+x)A_n + (x+x^2)E_{n-1} + xF_{n-1} + x^2H_{n-2}, \tag{33}$$

$$F_n = xA_n + E_n + x^2F_{n-1} + xI_{n-1} + x^2J_{n-2}, \tag{34}$$

$$G_n = (1+x)A_n + x^2C_{n-1} + 2xE_{n-1} + 2x^2H_{n-2}, \tag{35}$$

$$H_n = x^2A_n + x^3E_{n-1} + (1+x)F_n + xG_n + x^4H_{n-2} + xK_n, \tag{36}$$

$$I_n = (2x+x^2)A_n + (1+x)E_n + x^2E_{n-1} + x^2F_{n-1} + x^3H_{n-2}, \tag{37}$$

$$J_n = (2x+x^2)F_n + x^2G_n + (1+x)H_n + x^2K_n, \tag{38}$$

$$K_n = A_n + xE_{n-1} + xF_{n-1} + x^2H_{n-2} + x^2K_{n-1}. \tag{39}$$

By combining Eqs. (36) and (39) the K terms can be deleted as

$$(x+x^2)A_n - x^4A_{n-1} + (x^2-x^3)E_{n-1} - x^5E_{n-2} + (1+x)F_n - x^3F_{n-1} + xG_n - x^3G_{n-1} - H_n + x^2H_{n-1} + (x^3+x^4)H_{n-2} - x^6H_{n-3} = 0. \tag{40}$$

The C terms can be deleted from Eqs. (31) and (35) as

$$(1+x)A_n + x^2A_{n-1} + 2xE_{n-1} + 2x^3F_{n-2} - G_n + x^4G_{n-2} + 2x^2H_{n-2} = 0. \tag{41}$$

The I and J terms can be deleted from Eq. (34) by the aid of Eqs. (36)–(38) to give

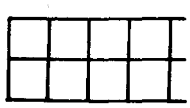
$$xA_n + (2x^2+x^3)A_{n-1} - x^5A_{n-2} + E_n + (x+x^2)E_{n-1} + x^3E_{n-2} - x^6E_{n-3} - F_n + x^2F_{n-1} + 2x^3F_{n-2} + (x^2+2x^3)H_{n-2} + x^4H_{n-3} - x^7H_{n-4} = 0. \tag{42}$$

The D, I, and J terms can be deleted from Eq. (32) by combining Eqs. (29)–(31) and (36)–(39) to give

$$A_n - xA_{n-1} - (2x^2+3x^3+x^4)A_{n-2} + x^5A_{n-3} - (1+x)E_{n-1} - (x+2x^2+x^3)E_{n-2} - (x^3+x^4)E_{n-3} + x^6E_{n-4} - 2x^2F_{n-2} - (2x^3+x^4)F_{n-3} - x^3G_{n-2} - xH_{n-2} - (x^2+2x^3)H_{n-3} - (x^4+x^5)H_{n-4} + x^7H_{n-5} = 0. \tag{43}$$

TABLE V. The Z-counting polynomial and the topological index of the $3\times n$ lattice.

	$p(G,k)$[a]								
k	$n=1$	2	3	4	5	6	7	8	9
0	1	1	1	1	1	1	1	1	1
1	2	7	12	17	22	27	32	37	42
2		11	44	102	185	293	426	584	767
3		$\underline{3}$	56	267	758	1 654	3 080	5 161	8 022
4			18	302	1597	5 256	13 254	28 191	53 292
5				123	1670	9 503	35 004	99 183	235 800
6				$\underline{11}$	757	9 401	56 456	227 262	708 881
7					106	4 603	53 588	355 396	1450 678
8						908	27 688	308 330	1993 990
9						$\underline{41}$	6 716	165 871	1786 876
10							540	46 801	991 849
11								5 580	313 290
12								$\underline{153}$	48 319
13									2 554
Z_G	3	22	131	823	5096	31 687	196 785	1 222 550	7 594 361



[a] The perfect matching numbers are underlined.

Now the operator technique can be applied to the simultaneous recursion equations (33) and (40)–(43) to give the following secular determinant whose rows and columns run in the order of A, E, F, G, and H:

$$\begin{vmatrix} (1+x)\hat{O}^2 & -\hat{O}^2+(x+x^2)\hat{O} & x\hat{O} & 0 & x^2 \\ (x+x^2)\hat{O}^3-x^4\hat{O}^2 & (x^2+x^3)\hat{O}^2-x^5\hat{O} & (1+x)\hat{O}^3-x^3\hat{O}^2 & x\hat{O}^3-x^3\hat{O}^2 & -\hat{O}^3+x^2\hat{O}^2+(x^3+x^4)\hat{O}-x^6 \\ (1+x)\hat{O}^2+x^2\hat{O} & 2x\hat{O} & 2x^3 & -\hat{O}^2+x^4 & 2x^2 \\ x\hat{O}^4+(2x^2+x^3)\hat{O}^3-x^5\hat{O}^2 & \hat{O}^4+(x+x^2)\hat{O}^3+x^3\hat{O}^2-x^6\hat{O} & -\hat{O}^4+x^2\hat{O}^3+2x^3\hat{O}^2 & 0 & (x^2+2x^3)\hat{O}^2+x^4\hat{O}-x^7 \\ \hat{O}^5-x\hat{O}^4-(2x^2+3x^3+x^4)\hat{O}^3+x^5\hat{O}^2 & -(1+x)\hat{O}^4-(x+2x^2+x^3)\hat{O}^3-(x^3+x^4)\hat{O}^2+x^6\hat{O} & -2x^2\hat{O}^3-(2x^3+x^4)\hat{O}^2 & -x^3\hat{O}^3 & -x\hat{O}^3-(x^2+2x^3)\hat{O}^2-(x^4+x^5)\hat{O}+x^7 \end{vmatrix} = 0. \quad (44)$$

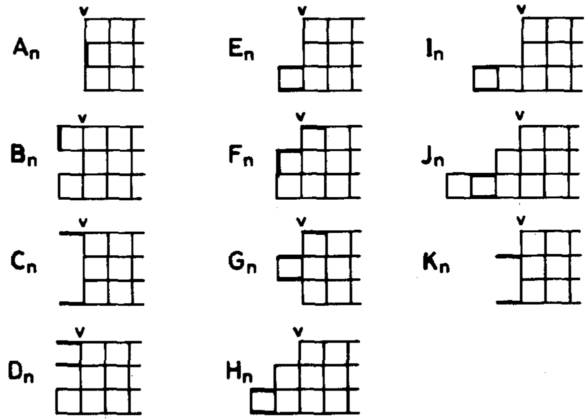


FIG. 8. Simultaneous recursion formulas obtained by decomposing the $4\times n$ lattice. See the caption of Fig. 6.

This determininant can be expanded into the polynomial as

$$\begin{aligned} &-\hat{O}^6(\hat{O}-x^2)\{\hat{O}^9-(1+5x+3x^2)\hat{O}^8 \\ &\quad-(2x+13x^2+21x^3+5x^4)\hat{O}^7 \\ &\quad-(x^2+4x^3-4x^4-27x^5-15x^6)\hat{O}^6 \\ &\quad+(3x^4+18x^5+41x^6+40x^7+9x^8)\hat{O}^5 \\ &\quad-(3x^6+14x^7+29x^8+24x^9+21x^{10})\hat{O}^4 \\ &\quad+(x^8-6x^{10}-19x^{11}-5x^{12})\hat{O}^3 \\ &\quad+(2x^{11}+3x^{12}+9x^{13}+9x^{14})\hat{O}^2 \\ &\quad+(x^{14}-x^{15}+x^{16})\hat{O}-x^{18}\}=0 \quad (4\times n), \end{aligned} \quad (45)$$

which gives the following recursion relation for A_n:

$$\begin{aligned} A_n &= (1+5x+3x^2)A_{n-1}+(2x+13x^2+21x^3+5x^4)A_{n-2} \\ &\quad+(x^2+4x^3-4x^4-27x^5-15x^6)A_{n-3} \\ &\quad-(3x^4+18x^5+41x^6+40x^7+9x^8)A_{n-4} \\ &\quad+(3x^6+14x^7+29x^8+24x^9+21x^{10})A_{n-5} \\ &\quad-(x^8-6x^{10}-19x^{11}-5x^{12})A_{n-6} \\ &\quad-(2x^{11}+3x^{12}+9x^{13}+9x^{14})A_{n-7} \\ &\quad-(x^{14}-x^{15}+x^{16})A_{n-8}+x^{18}A_{n-9} \quad (4\times n). \end{aligned} \quad (46)$$

The topological index Z_n and the number of perfect matching $K_n = K(4\times n)$ are found to recur, respectively, as

$$\begin{aligned} Z_n &= 9Z_{n-1}+41Z_{n-2}-41Z_{n-3}-111Z_{n-4}+91Z_{n-5} \\ &\quad+29Z_{n-6}-23Z_{n-7}-Z_{n-8}+Z_{n-9} \quad (4\times n) \end{aligned} \quad (47)$$

$$K_n = K_{n-1}+5K_{n-2}+K_{n-3}-K_{n-4} \quad (4\times n) \quad (48)$$

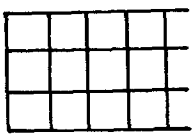
The smaller members of the Z-counting polynomials of $4\times n$ lattices are given in Table VI.

C. $3\times n$ and $4\times n$ torii

The $m\times n$ torus can be degraded into the $m\times n$ lattice and its subgraphs by applying the composition principles (I)–(III). Considering the mathematics of the operator technique one can infer that the operator polynomial for a torus can be

TABLE VI. The Z-counting polynomial and the topological index of the $4\times n$ lattice.

	$p(G,k)$									
k	$n=1$	2	3	4	5	6	7	8	9	10
0	1	1	1	1	1	1	1	1	1	1
1	3	10	17	24	31	38	45	52	59	66
2	1	29	102	224	395	615	884	1 202	1 569	1 985
3		26	267	1 044	2 696	5 566	9 997	16 332	24 914	36 086
4		5	302	2 593	10 769	31 106	72 277	145 356	263 818	443 539
5			123	3 388	25 835	111 882	350 894	893 980	1 970 796	3 906 352
6			11	2 150	36 771	261 965	1 169 511	3 906 894	10 716 696	25 493 632
7				552	29 580	395 184	2 689 527	12 288 090	43 153 390	125 765 134
8				36	12 181	372 109	4 230 941	27 870 240	129 672 547	474 237 944
9					2 111	206 206	4 452 310	45 253 920	290 890 497	1 373 442 102
10					95	60 730	3 014 229	51 741 942	484 175 443	3 053 582 362
11						7 852	1 232 561	40 527 972	590 385 544	5 184 745 628
12						281	274 258	20 847 772	516 766 986	6 655 476 596
13							27 403	6 602 264	315 084 296	6 359 057 492
14							781	1 160 009	128 090 556	4 422 387 941
15								93 674	32 522 360	2 169 365 254
16								2 245	4 648 352	717 928 045
17									310 496	150 132 840
18									6 336	17 898 019
19										1 013 474
20										18 061
Z_G	5	71	823	10 012	120 465	1 453 535	17 525 619	211 351 945	2 548 684 656	30 734 932 553



factored out by that for the parent rectangular lattice. The problem is then to find the quotient efficiently. First, for the Q polynomial A_n of several consecutive members of the $3\times n$ torus the same recursion relation F, or the relation (26) for the $3\times n$ lattice was formally applied. It was then found that the three consecutive difference polynomials between the estimated and correct Q polynomials form a recursion relation as

$$\begin{aligned} &A_n - F(A_{n-1}, A_{n-2}, \ldots, A_{n-6}) = G_n,\\ &A_{n+1} - F(A_n, A_{n-1}, \ldots, A_{n-5}) = H_n,\\ &A_{n+2} - F(A_{n+1}, A_n, \ldots, A_{n-4}) = x^3 G_n - xH_n. \end{aligned}$$

Then we get

$$\begin{aligned} &A_{n+2} + xA_{n+1} - x^3 A_n\\ &\quad = F(A_{n+1}, \ldots, A_{n-4}) + xF(A_n, \ldots, A_{n-5})\\ &\qquad - x^3 F(A_{n-1}, \ldots, A_{n-6}), \end{aligned} \tag{49}$$

meaning that the operator polynomial for A_n is obtained by expanding the following product:

$$\begin{aligned} &(\hat{O}^2 + x\hat{O} - x^3)\{\hat{O}^6 - (1+3x)\hat{O}^5 - x(2+7x+5x^2)\hat{O}^4\\ &\quad - x^2(1+x-2x^2)\hat{O}^3 + x^4(2+3x+5x^2)\hat{O}^2\\ &\quad - x^6(1-x)\hat{O} - x^9\} \end{aligned}$$

TABLE VII. The Z-counting polynomial and the topological index of the $3\times n$ torus.

	$p(G,k)$								
k	$n=2$	3	4	5	6	7	8	9	10
0	1	1	1	1	1	1	1	1	1
1									
1	10	15	20	25	30	35	40	45	50
2	24	69	142	240	363	511	684	882	1 105
3	12	107	440	1125	2 290	4 060	6 560	9 915	14 250
4		36	588	2710	8 139	19 222	38 934	70 875	119 270
5			288	3227	16 446	55 867	148 928	337 689	681 960
6			32	1645	18 141	99 085	371 008	1 093 524	2 731 225
7				240	9 870	103 231	594 880	2 410 182	7 731 110
8					2 148	58 310	593 260	3 565 728	15 419 490
9					108	15 267	345 216	3 434 867	21 360 480
10						1 274	104 688	2 044 575	20 005 489
11							13 280	689 058	12 126 350
12							392	112 221	4 439 005
13								6 138	873 870
14									73 980
15									1 452
Z_G	47	228	1511	9213	57 536	356 863	2 217 871	13 775 700	85 579 087

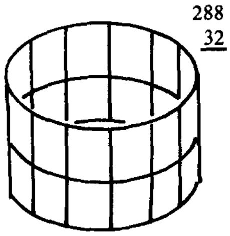
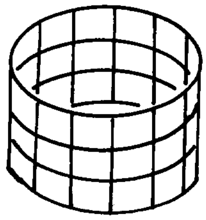


TABLE VIII. The Z-counting polynomial and the topological index of the $4\times n$ torus.

	$p(G,k)$							
k	$n=2$	3	4	5	6	7	8	9
0	1	1	1	1	1	1	1	1
1	14	21	28	35	42	49	56	63
2	59	156	306	505	753	1 050	1 396	1 791
3	82	501	1 672	3 910	7 562	12 971	20 480	30 432
4	29	672	4 863	17 725	46 938	102 543	196 986	345 114
5		285	7 416	48 193	187 530	544 642	1 310 136	2 762 766
6		19	5 470	77 405	487 241	1 985 823	6 193 408	16 105 872
7			1 620	69 510	813 486	4 991 717	21 068 296	69 537 870
8			121	31 060	843 342	8 566 530	51 645 071	224 005 914
9				5 360	509 542	9 797 963	90 500 088	538 378 440
10				176	160 653	7 160 181	111 379 032	958 971 672
11					21 438	3 118 822	93 487 528	1 248 917 526
12					725	718 753	51 147 646	1 163 887 980
13						69 321	17 004 352	751 691 466
14						1 471	3 063 292	320 964 705
15							241 992	84 350 355
16							5 041	12 136 608
17								766 431
18								11 989
Z_G	185	1 655	21 497	253 880	3 079 253	37 071 837	447 264 801	5 392 866 995
n	10	11	12	13	14	15	16	17
Z_G	37 584	97 021	290 521	783 511	2 289 869	6 323 504	18 241 441	51 026 011



$$\begin{aligned}&= \hat{O}^8 - (1+2x)\hat{O}^7 - x(3+10x+6x^2)\hat{O}^6 \\ &\quad - x^2(3+7x)\hat{O}^5 + x^3(-1+3x+12x^2+10x^3)\hat{O}^4 \\ &\quad + x^5(3+3x+4x^2)\hat{O}^3 - x^7(3+2x+6x^2)\hat{O}^2 \\ &\quad + x^9(1-2x)\hat{O} + x^{12} = 0 \quad (\overline{3\times n}). \qquad (50)\end{aligned}$$

This is nothing else but the recursion formula of the $3\times n$ torus,

$$\begin{aligned}A_n &= (1+2x)A_{n-1} + x(3+10x+6x^2)A_{n-2} \\ &\quad + x^2(3+7x)A_{n-3} - x^3(-1+3x \\ &\quad + 12x^2+10x^3)A_{n-4} - x^5(3+3x+4x^2)A_{n-5} \\ &\quad + x^7(3+2x+6x^2)A_{n-6} \\ &\quad - x^9(1-2x)A_{n-7} - x^{12}A_{n-8} \quad (\overline{3\times n}). \qquad (51)\end{aligned}$$

The recursion formula of the topological index of the $3\times n$ torus is then obtained by putting $x=1$ into Eq. (51) as

$$\begin{aligned}Z_n &= 3Z_{n-1} + 19Z_{n-2} \\ &\quad + 10Z_{n-3} - 24Z_{n-4} - 10Z_{n-5} \\ &\quad + 11Z_{n-6} + Z_{n-7} - Z_{n-8} \quad (\overline{3\times n}). \qquad (52)\end{aligned}$$

The Q's and Z's for the lower members of $3\times n$ torus graphs are given in Table VII, where the Kekulé number $K_n = K(\overline{3\times n})$ or the perfect matching number is underlined. Naturally it appears at the last term of every other polynomial A_n and the corresponding k value regularly increases by 3. By picking out the coefficients of the x^{3j} of every other $\hat{O}^{8-2j}$ term of Eq. (50) one gets the recursion polynomial for K_n as

$$\begin{aligned}&\hat{O}^8 - 6\hat{O}^6 + 10\hat{O}^4 - 6\hat{O}^2 + 1 \\ &\quad = (\hat{O}^2-1)^2(\hat{O}^4 - 4\hat{O}^2 + 1) = 0.\end{aligned}$$

A trial-and-error check shows us that the necessary and sufficient recursion formula for the Kekulé number of the $3\times n$ torus is

$$K_n = 5K_{n-2} - 5K_{n-4} + K_{n-6} \quad (\overline{3\times n}). \qquad (53)$$

For the $4\times n$ torus, however, the above-mentioned method was found to be inefficient, as will be inferred from the relatively large size of the following recursion relation of the Kekulé number of the $4\times n$ torus:

$$\begin{aligned}K_n &= K_{n-1} + 13K_{n-2} - 7K_{n-3} - 61K_{n-4} + 12K_{n-5} \\ &\quad + 128K_{n-6} - 128K_{n-8} - 12K_{n-9} + 61K_{n-10} \\ &\quad + 7K_{n-11} - 13K_{n-12} - K_{n-13} + K_{n-14} \quad (\overline{4\times n}). \qquad (54)\end{aligned}$$

This reveals that the corresponding operator polynomial is obtained by multiplying $\hat{O}^4 - \hat{O}^3 - 5\hat{O}^2 - \hat{O} + 1$ [see Eq. (48)] by such a high-ordered factor,

$$(\hat{O}^2-1)(\hat{O}^8 - 7\hat{O}^6 + 13\hat{O}^4 - 7\hat{O}^2 + 1).$$

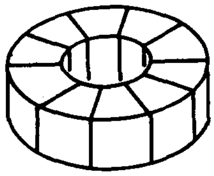
Equation (54) was actually derived by degrading the $4\times n$ torus into the $4\times n$ lattice and its subgraphs. All these component graphs are found to be classified into roughly three groups with different recursion polynomials whose least common multiple gives the relation (54). In Table VIII are given the Q polynomials and the Kekulé numbers for the lower members of the $4\times n$ torus.

D. $2\times 2\times n$ lattice and torus

By applying the jumbo composition principle (III) and the operator technique to the three-dimensional $2\times 2\times n$ lattice one can easily obtain the operator polynomial for the Z-counting polynomial as

TABLE IX. The Z-counting polynomial and the topological index of the $2\times2\times n$ torus.

	$p(G,k)$						
k	$n=2$	3	4	5	6	7	8
0	1	1	1	1	1	1	1
1	16	24	32	40	48	56	64
2	76	204	400	660	984	1 372	1 824
3	112	748	2 496	5 840	11 296	19 376	30 592
4	36	1149	8 256	30 195	80 058	174 993	336 248
5		588	14 208	93 324	364 464	1 060 416	2 553 280
6		50	11 648	169 660	1 075 876	4 402 580	13 761 856
7			3 712	171 820	2 033 328	12 569 260	53 277 952
8			272	86 725	2 376 105	24 425 086	148 298 136
9				17 300	1 610 560	31 528 252	294 311 296
10				722	566 124	25 921 168	409 048 704
11					83 472	12 677 728	386 622 464
12					3 108	3 288 761	237 566 272
13						364 476	88 543 488
14						10 082	17 889 792
15							1 599 232
16							39 952
Z_G	241	2764	41 025	576 287	8 205 424	116 443 607	1 653 881 153
n	9	10	11	12	13	14	
K_n	140 450	537 636	1 956 242	7 379 216	27 246 962	102 144 036	



$$\hat{O}^6-(1+7x+6x^2)\hat{O}^5-x(1+6x+6x^2-7x^3)\hat{O}^4 +2x^3(1+5x+13x^2+4x^3)\hat{O}^3 -x^5(1+2x+6x^2+9x^3)\hat{O}^2 -x^8(1-x+2x^2)\hat{O}+x^{12}=0 \quad (2\times2\times n), \tag{55}$$

giving the recursion formula

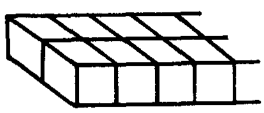
$$A_n(x)=(1+7x+6x^2)A_{n-1}(x) +x(1+6x+6x^2-7x^3)A_{n-2}(x) -2x^3(1+5x+13x^2+4x^3)A_{n-3}(x) +x^5(1+2x+6x^2+9x^3)A_{n-4}(x) +x^8(1-x+2x^2)A_{n-5}(x) -x^{12}A_{n-6}(x) \quad (2\times2\times n), \tag{56}$$

which is identical to what was obtained by Hock and McQuistan.[10] It is straightforward to derive the recursion formula for the $p(G,k)$ number from Eq. (56) as has been demonstrated in deriving Eq. (18) from Eq. (17). The result is identical to Eq. (2) of Ref. 10.

By picking up the coefficients of the terms $x^{2k}\hat{O}^{6-k}$

TABLE X. The Z-counting polynomial and the topological index of the $2\times3\times n$ lattice.

	$p(G,k)$					
k	$n=1$	2	3	4	5	6
0	1	1	1	1	1	1
1	7	20	33	46	59	72
2	11	142	436	899	1 531	2 332
3	3	440	2 984	9 798	23 073	45 006
4		588	11 434	65 722	224 640	577 933
5		288	24 766	282 586	1 487 214	5 222 662
6		32	29 180	787 131	6 864 336	34 261 363
7			16 984	1 404 402	22 287 124	165 933 320
8			3 993	1 553 006	50 747 188	597 679 423
9			229	1 001 258	79 898 508	1 601 033 392
10				338 099	84 714 756	3 168 578 566
11				48 790	57 970 292	4 571612 270
12				1 845	23 945 640	4 707 959 398
13					5 349 576	3 354 194 446
14					528 360	1 579 447 383
15					14 320	459 006 890
16						73 778 673
17						5 382 366
18						112 485
Z_G	22	1511	90 040	5 493 583	334 056 618	20 324 827 981



($k = 0, 1, \ldots, 6$) from Eq. (55) the operator polynomial for the Kekulé number can be constructed as

$$\hat{O}^6 - 6\hat{O}^5 + 7\hat{O}^4 + 8\hat{O}^3 - 9\hat{O}^2 - 2\hat{O} + 1 = (\hat{O}+1)(\hat{O}-1)(\hat{O}^2 - 2\hat{O} + 1)(\hat{O}^2 - 4\hat{O} + 1) = 0. \quad (57)$$

However, the smallest recursion formula of the Kekulé number K_n of the $2\times2\times n$ lattice is found to be[10]

$$K_n = 3K_{n-1} + 3K_{n-2} - K_{n-3} \quad (2\times2\times n), \quad (58)$$

which corresponds to the operator product $(\hat{O}+1)(\hat{O}^2 - 4\hat{O} + 1)$ chosen from the factors of Eq. (57) [see Eq. (28) of Ref. 10].

The operator polynomial for the Kekulé number of $2\times2\times n$ torus was obtained to be

$$\hat{O}^8 - 4\hat{O}^7 - 6\hat{O}^6 + 28\hat{O}^5 - 28\hat{O}^3 + 6\hat{O}^2 + 4\hat{O} - 1 = (\hat{O}+1)(\hat{O}-1)(\hat{O}^2 + 2\hat{O} - 1) \times (\hat{O}^2 - 2\hat{O} - 1)(\hat{O}^2 - 4\hat{O} + 1) = 0 \quad (\overline{2\times2\times n}). \quad (59)$$

In Table IX are given the coefficients of the Z-counting polynomial and Kekulé number for the lower members of the $2\times2\times n$ tori. The recursion relation for the former quantity is not yet obtained. The degree of the corresponding operator polynomial is estimated to be a little larger than 10 from the results of the operator polynomials for the relevant lattices and tori.

E. $2\times3\times n$ lattice

The operator polynomial for the Kekulé number of the $2\times3\times n$ lattice was similarly obtained to be

$$\hat{O}^{10} - 6\hat{O}^9 - 21\hat{O}^8 + 42\hat{O}^7 + 89\hat{O}^6 - 68\hat{O}^5 - 89\hat{O}^4 + 42\hat{O}^3 + 21\hat{O}^2 - 6\hat{O} - 1 = 0 \quad (2\times3\times n). \quad (60)$$

The degree of the operator polynomial for the Z-counting polynomial for the $2\times3\times n$ lattice is estimated as large as 20. In Table X are given the coefficients of the Z-counting polynomial and Kekulé number for the lower member of the $2\times3\times n$ lattices.

VII. CONCLUDING REMARKS

Thus far we have derived the recursion formulas for the Z-counting polynomial (including the perfect matching number) and topological index of several fundamental rectangular and torus lattices, which give the partition functions for the dimer statistics. In principle, a systematic application of this method can be performed to larger lattices. However, as the size of the lattice increases this strategy will soon come to a dead end. The graph-theoretic technique developed in this paper alone cannot readily be generalized to larger lattices. A further goal is to obtain the recursion relations among the operator polynomials for a set of $m\times n$ lattices with different m's and, of course, those for three-dimensional lattices.

There are several additional approaches for tackling this problem. The coefficients of the Z-counting polynomial for larger lattices could be approximated by a smooth or well-behaved function, whose mathematical properties are thoroughly understood or tractable. On the contrary, a number-theoretic approach might be possible with particular reference to the highly composite nature of the coefficients of the Z-counting polynomial, especially of the perfect matching number.[24] These possibilities are the reasons why the seemingly unnecessary lists of the coefficients of the Z-counting polynomials for different kinds of lattices have been presented in this paper. Further discussions derived from the manipulation of these numbers will be published elsewhere.

[1] P. W. Kasteleyn, in *Graph Theory and Theoretical Physics*, edited by F. Harary (Academic, London, 1967), pp. 43.
[2] J. K. Percus, *Combinatorial Methods* (Courant Institute of New York Univ., New York, 1969).
[3] H. N. V. Temperley, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, London, 1972), Vol. 1, pp. 227.
[4] H. N. V. Temperley, *Graph Theory and Applications* (Ellis Horwood, Chichester, England, 1981).
[5] M. Gordon and W. H. T. Davison, J. Chem. Phys. **20**, 428 (1952).
[6] N. Trinajstić, *Chemical Graph Theory* (CRC, Boca Raton, FL, 1983): extensive bibliographies on the related problems are given.
[7] H. N. V. Temperley and M. E. Fisher, Philos. Mag. **6**, 1061 (1961).
[8] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[9] R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).
[10] J. H. Hock and R. B. McQuistan, J. Math. Phys. **24**, 1859 (1983).
[11] H. Hosoya, Bull. Chem. Soc. Jpn. **44**, 2332 (1971); Theor. Chim. Acta **25**, 215 (1972).
[12] H. Hosoya, K. Hosoi, and I. Gutman, Theor. Chim. Acta **38**, 37 (1975).
[13] H. Hosoya and K. Hosoi, J. Chem. Phys. **64**, 1065 (1976).
[14] H. Hosoya, K. Kawasaki, and K. Mizutani, Bull. Chem. Soc. Jpn. **45**, 3415 (1972).
[15] H. Narumi and H. Hosoya, Bull. Chem. Soc. Jpn. **53**, 1228 (1980).
[16] H. Hosoya, J. Chem. Doc. **12**, 181 (1972).
[17] H. Hosoya and H. Okhami, J. Comput. Chem. **4**, 585 (1983).
[18] After submission of this manuscript the authors were informed by the referee that similar results had been obtained from a different technique by A. J. Phares, J. Math. Phys. **25**, 1756 (1984).
[19] A. Motoyama and H. Hosoya, J. Math. Phys. **18**, 1485 (1977).
[20] For standard graph-theoretical concepts and theorems see F. Harary, *Graph Theory* (Addison–Wesley, Reading, MA 1969); *Chemical Application of Graph Theory*, edited by A. Balaban (Academic, London, 1980), D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1980).
[21] S. W. Golomb, *Polyominoes* (Scribner's, New York, 1965).
[22] F. Harary in *Graph Theory and Theoretical Physics*, edited by F. Harary (Academic, London, 1967), pp. 1.
[23] M. A. Rashid, J. Math. Phys. **15**, 474 (1974).
[24] D. Klarner and J. Pollack, Discrete Math. **32**, 45 (1980).
[25] They have derived the recursion formulas of $K(m\times n)$ with $m = 2, 3, \ldots, 8$. However, the result for $m = 8$ is doubtful, since there are found several numerical errors in these entries. See also Ref. 26.
[26] R. C. Read, Fibonacci Q. **18**, 24 (1980).

Vacuum states and superselection rules in quantum field theory

R. Borek[a)]
Institut für Theoretische Physik, Universität Göttingen, D-3400 Göttingen, West Germany

(Received 24 April 1984; accepted for publication 9 August 1984)

Let $(\mathbf{A}(\mathbf{O}),5^4,\alpha)$ be a theory of local observables and let $\mathbf{A}$ denote the C^*-algebra of quasilocal observables. Assume that every representation of $\mathbf{A}$ which satisfies the spectrum condition is type I. Then if $(\pi,\mathbf{H})$ is an irreducible representation of $\mathbf{A}$ which satisfies the spectrum condition there exists a unique vacuum state asociated with $(\pi,\mathbf{H})$ by large spacelike translations.

I. INTRODUCTION

A superselection rule in quantum field theory is defined as any restriction of what is an observable in the theory. In Haag–Kastler theory[1] one regards the algebraic properties of local observables as the fundamental structure which embodies the relevant information about physically admissible states. If an algebra of local observables is given then one can derive, in principle, the properties of physically admissible states. Once it is known what is physically relevant in the theory then the superselection sectors of a Haag–Kastler theory are given by the unitary equivalent physically admissible representations of the quasilocal algebra. If there are superselection rules in a model there exist several such sectors, and one may regard the labels distinguishing them as charge quantum numbers. One can then construct a representation of the algebra of observables on a global physical Hilbert space by picking representations from each sector and taking their direct sum. In this sense a Haag–Kastler theory determines its own superselection rules.

The existence of a unique vacuum state in a quantum field theory is one of the axioms in Wightman theory. We shall show in an algebraic framework that the existence of a vacuum state does not need to be postulated—the vacuum state as a physically realizable state in a quantum field theory can be shown to exist if the superselection rules of the physical theory under consideration are known.

We shall state the Haag–Kastler–Araki axioms in the notation of Ref. 2.

(I) To every bounded open region $\mathbf{O}\subset\mathbb{R}^4$, one assigns a C^*-algebra $\mathbf{A}(\mathbf{O})$ such that (a) $\mathbf{O}_1\subset\mathbf{O}_2$ implies $\mathbf{A}(\mathbf{O}_1)\subset\mathbf{A}(\mathbf{O}_2)$; and (b) if the regions $\mathbf{O}_1$ and $\mathbf{O}_2$ are spacelike separated, then the elements of $\mathbf{A}(\mathbf{O}_1)$ commute with all elements of $\mathbf{A}(\mathbf{O}_2)$. The algebra of quasilocal observables $\mathbf{A}$ will denote the C^*-algebra generated by the union of $\{\mathbf{A}(\mathbf{O})\}$.

(II) There exists a representation of the vector group $\mathbb{R}^4$ as automorphisms of $\mathbf{A}$,

$$\alpha:\mathbf{R}^4\to\mathrm{Aut}\,\mathbf{A},$$

and, furthermore,

$$\alpha_{\mathbf{a}}\mathbf{A}(\mathbf{O})=\mathbf{A}(\mathbf{O}+\mathbf{a})\text{ for every }\mathbf{a}\text{ in }\mathbb{R}^4.$$

Let $\mathbf{H}$ be a complex separable Hilbert space.

(III) A representation π of $\mathbf{A}$ on $\mathbf{H}$ is called a representation satisfying the spectrum condition if the following holds: (a) there exists a strongly continuous unitary representation $U(\mathbf{a})$ of the vector group $\mathbb{R}^4$ on the Hilbert space $\mathbf{H}$; (b) the representation $U(\mathbf{a})$ implements the automorphisms $\alpha_{\mathbf{a}}$, that is,

$$U(\mathbf{a})\pi(x)U(\mathbf{a})^{-1}=\pi(\alpha_{\mathbf{a}}(x))\text{ for every }x\in\mathbf{A};$$

(c) the spectrum of the representation $U(\mathbf{a})$ is contained in the future light cone.

We shall consider C^*-dynamical systems which satisfy axioms I and II and have at least one faithful representation which satisfies the spectrum condition. Such a system will be called a theory of local observables and will be denoted by $(\mathbf{A}(\mathbf{O}),\mathbb{R}^4,\alpha)$.

A state ω on $\mathbf{A}$ is called a vacuum state if it is invariant under the automorphisms $\alpha_{\mathbf{a}}$ and the cyclic representation (π_ω, H_ω) induced by ω is a representation satisfying the spectrum condition.

It was shown recently by Buchholz and Fredenhagen[3] that if $(\pi,\mathbf{H})$ is a massive single-particle representation of $\mathbf{A}$, then there exists a unique vacuum state associated with the representation $(\pi,\mathbf{H})$ by the method of large spacelike translations.[4]

In order to make this notion more precise, we shall give the following definition.

Definition 1.1: Let $(\mathbf{A}(\mathbf{O}),\mathbb{R}^4,\alpha)$ be a theory of local observables. Let $(\pi,\mathbf{H})$ be a factor representation of $\mathbf{A}$ which satisfies the spectrum condition.

Assume that weak $\lim_{\lambda\to\infty}\pi(\alpha_{\lambda\mathbf{a}}(x))=\omega_0(x)\mathrm{I}$ exists for any spacelike vector $\mathbf{a}$. Then ω_0 is a vacuum state which will be called the vacuum state associated with the representation $(\pi,\mathbf{H})$.

Following Doplicher *et al.*,[5] we shall make the following assumptions.

(IV) To each bounded open region $\mathbf{O}$ in space-time, one associates a local field algebra $\mathscr{F}(\mathbf{O})$ which is a weakly closed $*$-subalgebra of $\mathbf{B}(\mathbf{H})$. $\mathbf{O}_1\subset\mathbf{O}_2$ implies that $\mathscr{F}(\mathbf{O}_1)\subset\mathscr{F}(\mathbf{O}_2)$. The field algebra $\mathscr{F}$ is then defined by $\mathscr{F}=\overline{\cup\mathscr{F}(\mathbf{O})}$ where the closure is in the norm topology.

One assumes that $\mathscr{F}''=\mathbf{B}(\mathbf{H})$.

(V) There is a compact gauge group G and a strongly continuous unitary representation $V(g)$ of it such that

$$V(g)FV(g)^{-1}=\alpha_g(F)\text{for all }F\text{ in }\mathscr{F},$$

and $\alpha_g(\mathscr{F}(\mathbf{O}))=\mathscr{F}(\mathbf{O})$ for all $\mathbf{O}$.

(VI) The local algebra $\mathbf{A}(\mathbf{O})$ is a C^*-subalgebra of $\mathscr{F}(\mathbf{O})$ which is invariant under gauge transformations, that is,

$$\mathbf{A}(\mathbf{O})=\mathscr{F}(\mathbf{O})\cap V(G)'\quad\text{for all }\mathbf{O}.$$

By Ref. 5 there exists a one-to-one correspondence

[a)] Present address: Department of Mathematics, Bedford College, Regents Park, London NW1 4NS, England.

 0022-2488/85/010168-04$02.50

between the elements of $\hat{G}$ and the physical spectrum of **A**. Since every representation with spectrum condition is type I it follows from Glimm's theorem[6] that the physical spectrum of the quasilocal algebra **A** is a smooth subspace of the algebraic dual space of **A**. We shall show that if there does not exist a unique vacuum state associated with every irreducible representation of **A** which satisfies the spectrum condition then the subspace of equivalence classes of these vacuum representations is not a standard Borel space. Therefore, it will follow that there exists a unique vacuum state associated with every irreducible representation which satisfies the spectrum condition.

It is well known that every vacuum representation of a free massless Fermi field is equivalent to the Fock representation. We would expect, however, that our results do not apply to an interacting theory which describes massless particles.

II. DEFINITIONS AND NOTATION

The notation we give will be that of Ref. 2. Let **A** be a C^*-algebra and $\mathbf{A}^{**}$ the double dual space of **A**. Then $\mathbf{A}^{**}$ becomes a von Neumann algebra, in a natural manner, if it is endowed with the weak topology induced by the topology of $\mathbf{A}^*$.

The set of spacelike directions in $\mathbb{R}^4$ will be denoted by **D**.

For any vector **a** in $\mathbb{R}^4$ and any x in $\mathbf{A}^{**}$, we shall define a set of operators $\mathbf{K}_\mathbf{a}(x)$.

Definition 2.1 (see Ref. 7): Let $(\mathbf{A}, \mathbb{R}^4, \alpha)$ be a C^*-dynamical system. For $x \in \mathbf{A}^{**}$ and $\mathbf{a} \in \mathbb{R}^4$, we define a set $\mathbf{K}_\mathbf{a}(x)$ by

$$\mathbf{K}_\mathbf{a}(x) = \bigcap_{M>0} [C_0\{\alpha_{\lambda\mathbf{a}}(x) : \lambda \geqslant M\}]^{-w}$$

where the closure is in the weak topology. Here, C_0 denotes the closed convex hull of the set $\{\alpha_{\lambda\mathbf{a}}(x) : \lambda \geqslant M\}$.

It follows from Ref. 7 that, for $x \in \mathbf{A}, \mathbf{K}_\mathbf{a}(x) \subset \mathbf{Z}(\mathbf{A}^{**})$ for any spacelike direction **a**, where $\mathbf{Z}(\mathbf{A}^{**})$ denotes the center of the vonNeumann algebra $\mathbf{A}^{**}$.

Let V denote the future light cone. If **A** is a C^*-algebra, we shall denote by $\mathbf{S}(\mathbf{A})$ the set of states of **A**.

$\mathbf{S}_0(\hat{V})$ will denote the set of states with the properties (a) $\omega(x\alpha_\mathbf{a}y)$ is continuous for every $x, y \in \mathbf{A}^{**}$; (b) $\omega(x\alpha_\mathbf{a}y) = f(\mathbf{a})$ is the boundary value of an analytic function $f(z)$ holomorphic in the tube $R^4 + iV^0 = T^+$, where V^0 denotes the interior of V; and (c) there exists a constant $m \geqslant 0$ depending on ω such that $|f(z)| \leqslant \|x\|\ \|y\| \exp\{m|\mathrm{Im}\, z|\}$.

$\mathbf{S}(\hat{V})$ will denote the norm closure of $\mathbf{S}_0(\hat{V})$.

The set $\mathbf{S}(\hat{V})$ has the following properties.

(1) If $(\pi, \mathbf{H})$ is a representation of **A**, there exists a strongly continuous unitary representation $U(\mathbf{a})$ of the translation group $\mathbb{R}^4$ which implements the automorphisms $\alpha_\mathbf{a}$, that is, $U(\mathbf{a})\pi(x)U(\mathbf{a})^{-1} = \pi(\alpha_\mathbf{a}(x))$ $(x \in \mathbf{A})$, and the spectrum of $U(\mathbf{a})$ is contained in $\hat{V}$ if and only if all normal states of π are in $\mathbf{S}(\hat{V})$.

(2) $\mathbf{S}(\hat{V})$ is a folium. This means that there exists a projection $E(\hat{V}) \in \mathbf{Z}(\mathbf{A}^{**})$ such that $\omega \in \mathbf{S}(\mathbf{A})$ is in $\mathbf{S}(\hat{V})$ if and only if $\omega(E(\hat{V})) = 1$.

(3) $\mathbf{S}(\hat{V})$ is invariant under $\alpha_\mathbf{a}$ for every $\mathbf{a} \in \mathbb{R}^4$. This implies that $E(\hat{V})$ is invariant under the automorphisms $\alpha_\mathbf{a}$.

Let $(\mathbf{A}(\mathbf{O}), \mathbb{R}^4, \alpha)$ be a theory of local observables. It follows from Ref. 8 that the automorphisms $\alpha_\mathbf{a}$ are spatial in $\mathbf{A}^{**}E(\hat{V})$. There exists a strongly continuous unitary representation $U(\mathbf{a})$ of $\mathbb{R}^4$ which implements the automorphisms $\alpha_\mathbf{a}$ and $U(\mathbf{a}) \in \mathbf{A}^{**}E(\hat{V})$. Furthermore, the representation $U(\mathbf{a})$ is minimal in the sense that, if $V(\mathbf{a})$ is a strongly continuous unitary representation of $\mathbb{R}^4$ which has the same properties, then $U(\mathbf{a})V(\mathbf{a})^{-1} \in \mathbf{Z}(\mathbf{A}^{**}E(\hat{V}))$.

III. THE MAIN THEOREM

Theorem 3.1: Let $(\mathbf{A}(\mathbf{O}), \mathbb{R}^4, \alpha)$ be a theory of local observables. Assume that every representation of **A** which satisfies the spectrum condition is a representation of type I. If $(\pi, \mathbf{H})$ is an irreducible representation of **A** which satisfies the spectrum condition, then there exists a unique vacuum state associated with the representation $(\pi, \mathbf{H})$.

Lemma 3.2 (see Ref. 7): Let $(\mathbf{A}, \mathbb{R}^4, \alpha)$ be a C^*-dynamical system. Let $x \in \mathbf{A}$. Then the set $\mathbf{K}_\mathbf{a}(x) \subset \mathbf{Z}(\mathbf{A}^{**})$ is a single point for any spacelike direction **a** if and only if $\alpha_{\lambda\mathbf{a}}(x)$ converges weakly to this point as λ tends to infinity.

If $(\pi, \mathbf{H})$ is a representation of **A** which satisfies the spectrum condition, let

$$\pi(\mathbf{K}_\mathbf{a}(x)) = \bigcap_{M>0} C_0\{\pi(\alpha_{\lambda\mathbf{a}}(x)) : \lambda \geqslant M\}^{-w},$$

where C_0 denotes the closed convex hull and the closure is in the weak topology.

Lemma 3.3 (see Ref. 7): Let $(\mathbf{A}(\mathbf{O}), \mathbb{R}^4, \alpha)$ be a theory of local observables. Let $(\pi, \mathbf{H})$ be a factor representation of **A** which satisfies the spectrum condition.

Then there exists a unique (up to a phase) translation invariant state associated with the representation $(\pi, \mathbf{H})$ if and only if the set $\pi(\mathbf{K}_\mathbf{a}(x))$ $(x \in \mathbf{A})$ is a single point for any spacelike direction **a**.

Proof: We define a function F on the subspace $\mathbf{A}^+$ (the subspace of the positive elements of **A**) by

$$F(x) = \operatorname*{lub}_{\mathbf{a} \in \mathbf{D}}\{y_\mathbf{a}(x) : y_\mathbf{a}(x) \in \pi(\mathbf{K}_\mathbf{a}(x))\}.$$

Then it follows from Lemma II.2 of Ref. 7 that the function $F: \mathbf{A}^+ \to \mathbf{Z}(\mathbf{A}^{**})$ is sublinear, that is, $F(\lambda x) = \lambda F(x)$ for any $\lambda \geqslant 0$ and $x \in \mathbf{A}^+$; $F(x+y) \leqslant F(x) + F(y)$ for any x, y in $\mathbf{A}^+$. Furthermore, $F(\alpha_\mathbf{a}(y)) = F(y)$ for any vector **a** in $\mathbb{R}^4$ and y in $\mathbf{A}^+$. Then, by the proof of the Hahn–Banach theorem, there exists a unique bounded linear map from **A** into $\mathbf{Z}(\mathbf{A}^{**})$ with $L(\mathbf{I}) = \mathbf{I}$, $L(x) \leqslant F(x)$ on $\mathbf{A}^+$, and $L(\alpha_\mathbf{a}(y)) = L(y)$ for y in **A** and any **a** in $\mathbb{R}^4$. If ω is any normal state of $(\pi, \mathbf{H})$ we let ω_L be the state $\omega_L(x) = \omega(L(x))$ for x in **A**. Then, by construction, ω_L is invariant under the automorphisms $\alpha_\mathbf{a}$.

Conversely, if $\pi(\mathbf{K}_\mathbf{a}(x))$ $(x \in \mathbf{A})$ is not a single point for some spacelike direction **a** in **D**, then for any net $\{y_{\lambda,\mathbf{a}}(x)\}$ in $\mathbf{Z}(\mathbf{A}^{**})$ with $y_{\lambda,\mathbf{a}}(x) \in \pi(\mathbf{K}_\mathbf{a}(x))$, we can associate by the above construction an invariant state with the representation $(\pi, \mathbf{H})$.

Lemma 3.4 (Ref. 9; Theorem 10): The invariant states defined by Lemma 3.3 are vacuum states on the quasilocal algebra **A**.

Remark: Theorem 3.1 is not valid in two-dimensional space-time. In so-called massive soliton theories in two space-time dimensions there are states where there exist different right and left vacua. Lemma 3.4 is proved under the assumption that there always exist states whose supports are arbitrarily close to the boundary of the future light cone. This assumption is not valid in two-dimensional space-time since in this case the Minkowski space is disconnected in arbitrary neighborhoods of the boundary of the future light cone.

Proof of Theorem 3.1: Let $\mathbf{A}^c$ denote the set of all representations of $\mathbf{A}$ on $\mathbf{H}$. Let $\mathbf{A}^c$ have the smallest Borel structure such that, for every ϕ, Ψ in $\mathbf{H}$, the complex-valued function $\rho(x) = (\pi(x)\phi, \Psi)$ defined on $\mathbf{A}^c$ is a Borel function. Let $\widehat{\mathbf{A}}^c$ be the subset of $\mathbf{A}^c$ which contains all irreducible representations of $\mathbf{A}$. Let $\widehat{\mathbf{A}}$ denote the set of unitary equivalence classes of representations of $\widehat{\mathbf{A}}^c$. $\widehat{\mathbf{A}}$ becomes a Borel space if it is endowed with the quotient topology induced by the Borel structure of $\widehat{\mathbf{A}}^c$.

Let $\mathbf{A}^c_{\rm sp}$ denote the set of all representations of $\mathbf{A}$ on $\mathbf{H}$ which satisfy the spectrum condition and $\widehat{\mathbf{A}}^c_{\rm sp}$ be the subset of $\mathbf{A}^c_{\rm sp}$ which contains all irreducible representations with the spectrum condition. In Lemma 3.5 we shall show that $\widehat{\mathbf{A}}^c_{\rm sp}$ is a Borel space with the Borel structure which it inherits as a subspace of $\widehat{\mathbf{A}}^c$. Let $\widehat{\mathbf{A}}_{\rm sp}$ denote the set of unitary equivalence classes of representations of $\widehat{\mathbf{A}}^c_{\rm sp}$. We shall show that $\widehat{\mathbf{A}}_{\rm sp}$ is a Borel subset of $\widehat{\mathbf{A}}$. Every representation of $\mathbf{A}$ which satisfies the spectrum condition is type I. Then it follows from Glimm's theorem[6] that $\widehat{\mathbf{A}}_{\rm sp}$ is smooth (Ref. 10, Corollary 1; p. 139). Therefore, $\widehat{\mathbf{A}}_{\rm sp}$ is a standard Borel space and there is a countable family of Borel subsets of $\widehat{\mathbf{A}}_{\rm sp}$ which separate points of $\widehat{\mathbf{A}}_{\rm sp}$. By Ref. 6, $\widehat{\mathbf{A}}_{\rm sp}$ is metrically countably separated; that is, for each σ-finite measure μ on $\widehat{\mathbf{A}}_{\rm sp}$ there is a Borel subset $\mathbf{N}$ of $\widehat{\mathbf{A}}_{\rm sp}$ such that $\mu(\mathbf{N}) = 0$ and $\widehat{\mathbf{A}}_{\rm sp} \sim \mathbf{N}$ is countably separated.

Lemma 3.5: $\widehat{\mathbf{A}}_{\rm sp}$ is a Borel subset of $\widehat{\mathbf{A}}$.

Proof: From the primitivity condition on the quasilocal algebra $\mathbf{A}$ it follows that $\widehat{\mathbf{A}}^c_{\rm sp}$ is a set of the second Baire category in $\widehat{\mathbf{A}}^c$ and therefore by the Banach–Steinhaus theorem it is a Borel subset of $\widehat{\mathbf{A}}^c$.

Let $\mathbf{B}_q$ be the quotient Borel structure on $\widehat{\mathbf{A}}_{\rm sp}$ derived from $\widehat{\mathbf{A}}^c_{\rm sp}$. If x is an element of $\widehat{\mathbf{A}}^c_{\rm sp}$, let $\tilde{x}$ be the corresponding unitary equivalence class in $\widehat{\mathbf{A}}_{\rm sp}$. If $\tilde{x} \notin \widehat{\mathbf{A}}$ we let $\rho(\tilde{x})$ be the unitary equivalence class in $\widehat{\mathbf{A}}$ containing $\tilde{x}$. Let $\mathbf{B}_s$ be the Borel structure on $\widehat{\mathbf{A}}_{\rm sp}$ which makes ρ a Borel isomorphism of $\widehat{\mathbf{A}}_{\rm sp}$ with $\rho(\widehat{\mathbf{A}}_{\rm sp})$, where $\rho(\widehat{\mathbf{A}}_{\rm sp})$ has a Borel structure as a subspace of $\widehat{\mathbf{A}}$. The lemma is proved if we show that the Borel structures $\mathbf{B}_q$ and $\mathbf{B}_s$ coincide. If $\widetilde{E} \subset \widehat{\mathbf{A}}$, let E be the set of elements of $\widetilde{E}$. If $\widetilde{E} \subset \widehat{\mathbf{A}}_{\rm sp}$, then $\widetilde{E} \in \mathbf{B}_s$ if and only if there is a Borel subset $\widetilde{F}$ of $\widehat{\mathbf{A}}$ such that $\rho(\widetilde{E}) = \rho(\widehat{\mathbf{A}}_{\rm sp}) \cap \widetilde{F}$ or, equivalently, $E = \widehat{\mathbf{A}}^c_{\rm sp} \cap F$ while $\widetilde{E} \in \mathbf{B}_q$ if and only if there is a Borel subset D *of* $\widehat{\mathbf{A}}^c$ such that $E = \widehat{\mathbf{A}}^c_{\rm sp} \cap D$. Therefore $\mathbf{B}_s \subset \mathbf{B}_q$. Since $\widehat{\mathbf{A}}^c_{\rm sp}$ is a Borel subset of $\widehat{\mathbf{A}}^c$ it follows from the same argument that $\mathbf{B}_q \subset \mathbf{B}_s$.

In view of Lemmas 3.3 and 3.4 the proof of Theorem 3.1 follows from the following Lemma.

Lemma 3.6: Let $(\mathbf{A}(\mathbf{O}), \mathbb{R}^4, \alpha)$ be a theory of local observables. Let $x \in \mathbf{A}^{**}E(V)$. If $\widehat{\mathbf{A}}_{\rm sp}$ is smooth then the set $\mathbf{K}_{\mathbf{a}}(x)$ is a single point for any spacelike direction $\mathbf{a}$.

Proof: Let $(\pi, \mathbf{H})$ be an irreducible representation with spectrum condition. Let $x \in \pi(\mathbf{A})''$ and assume that the set $\mathbf{K}_{\mathbf{a}}(x)$ is not a single point for some spacelike direction $\mathbf{a}$. It follows from the remarks above that the lemma is proved if we can show that $\widehat{\mathbf{A}}_{\rm sp}$ is not metrically countably separated. In fact, to show this, it suffices to find a subset $\mathbf{K}$ of $\widehat{\mathbf{A}}_{\rm sp}$ such that $\mathbf{K}$ as a subspace of $\widehat{\mathbf{A}}_{\rm sp}$ is not metrically countably separated.

To see this, we let μ be a σ-finite measure on $\mathbf{K}$ and if $\mathbf{N}$ is any Borel subset of $\mathbf{K}$ of μ-measure zero, we assume that $\mathbf{K} \sim \mathbf{N}$ is not metrically countably separated. Define $\tilde{\mu}(E) = \mu(E \cap \mathbf{K})$ for E a Borel subset of $\widehat{\mathbf{A}}_{\rm sp}$. Then $\tilde{\mu}$ is a σ-finite measure on $\widehat{\mathbf{A}}_{\rm sp}$. Then $\mathbf{N}$ is a Borel subset of $\widehat{\mathbf{A}}_{\rm sp}$ and $\tilde{\mu}(\mathbf{N}) = 0$. Let $E_1, E_2, \ldots$ be Borel subsets of $\widehat{\mathbf{A}}_{\rm sp}$. Then $\tilde{\mu}(\mathbf{N} \cap \mathbf{K}) = 0$ and the sets $E_1 \cap \mathbf{K}, E_2 \cap \mathbf{K}, \ldots$ do not separate $\mathbf{K} \sim \mathbf{N}$ and this implies that the sets $E_1, E_2, \ldots$ do not separate $\widehat{\mathbf{A}}_{\rm sp} \sim \mathbf{N}$ and hence $\widehat{\mathbf{A}}_{\rm sp}$ is not metrically countably separated.

For $x \in \mathbf{A}^{**}E(V)$, assume the set $\mathbf{K}_{\mathbf{a}}(x)$ is not a single point for some spacelike direction $\mathbf{a}$. Since the set $\mathbf{K}_{\mathbf{a}}(x)$ is convex, it follows that if y_1, y_2 are any two points of $\mathbf{K}_{\mathbf{a}}(x)$, then so is $\lambda_1 y_1 + \lambda_2 y_2$ for $\lambda_1 + \lambda_2 = 1$. Therefore, if the set $\mathbf{K}_{\mathbf{a}}(x)$ is not a single point for some spacelike direction $\mathbf{a}$, then it is infinite. For $x \in \mathbf{A}$, we choose nets $\{ y_{\lambda,\mathbf{a}}(x)\}$ with $y_{\lambda,\mathbf{a}}(x) \in \mathbf{K}_{\mathbf{a}}(x)$ and let F_λ be a bounded sublinear map from $\mathbf{A}^+$ into $\mathbf{Z}(\mathbf{A}^{**}E(\widehat{V}))$ defined by

$$F_\lambda(x^*x) = \operatorname*{lub}_{\mathbf{a} \in \mathbf{D}} \{ y_{\lambda,\mathbf{a}}(x^*x) : y_{\lambda,\mathbf{a}}(x) \in \mathbf{K}_{\mathbf{a}}(x)\}.$$

Let L_λ be a bounded linear map from $\mathbf{A}$ into $\mathbf{Z}(\mathbf{A}^{**}E(\widehat{V}))$ such that $\|L_\lambda\| = 1$, $L_\lambda(\alpha_{\mathbf{a}}(x)) = L_\lambda(x)$ for x in $\mathbf{A}$, any vector $\mathbf{a}$ in $\mathbb{R}^4$, and $L_\lambda(x) \leqslant F_\lambda(x)$ for x in $\mathbf{A}^+$. Let S be the collection of all nets in $\mathbf{Z}(\mathbf{A}^{**}E(\widehat{V}))$ such that for any distinct nets $\{ y_{\lambda_1,\mathbf{a}}(x)\}$, $\{ y_{\lambda_2,\mathbf{a}}(x)\}$ in S we have $L_{\lambda_1} \neq L_{\lambda_2}$. If $L_\lambda(x) = \omega_\lambda(x)\mathrm{I}$ and ω_λ is a vacuum state defined by the above relations for a net $\{ y_{\lambda,\mathbf{a}}(x)\}$ in S with $y_{\lambda,\mathbf{a}}(x) \in \pi(\mathbf{K}_{\mathbf{a}}(x))$, let π_λ be the induced representation of $\mathbf{A}$. Let $\mathbf{K}^c$ denote the subset of $\widehat{\mathbf{A}}^c_{\rm sp}$, which consists of all representations π_λ of $\mathbf{A}$. In particular, π is an element of $\mathbf{K}^c$.

Let X_0 be the measure space $\{0, 1\}$, let $\mathbf{B}_0$ be the set of subsets of $\{0, 1\}$, let ν_0 be the measure on X_0 defined by $\nu_0(\{0\}) = k_1$, $\nu_0(\{1\}) = k_2$ for $k_1 + k_2 = 1$. Let $(X_n, \mathbf{B}_n, \nu_n) = (X_0, \mathbf{B}_0, \nu_0)$ for $n = 1, 2, \ldots$, let

$$(X, \mathbf{B}', \nu') = \left(\prod_{i=1}^{\infty} X_i, \prod_{i=1}^{\infty} \mathbf{B}_i, \prod_{i=1}^{\infty} \nu_i \right),$$

and let $(X, \mathbf{B}, \nu)$ denote the measure space formed by the completion of ν'. If χ is in X, then χ is identified with the sequence (χ_n), where $\chi_n = 0$ or 1. If $\xi = (\xi_n)$ is in X, we define $\chi + \xi$ to be the sequence $(\chi_n + \xi_n)$ reduced mod 2. Then X is a group, and $\Delta = \{(\chi_n) : \chi_n \neq 0$ for at most a finite number of $n\}$ is a countable subgroup of X generated by the elements $((\gamma_k)_n)$, where $(\gamma_k)_n = \delta^n_k$. We define an action of Δ on X by $\chi\gamma = \chi + \gamma$. Then X/Δ becomes a Borel Δ-space with this right action of Δ on X.

The representation $(\pi, \mathbf{H})$ is a factor representation. Therefore, if $x \in \mathbf{A}$, then we assume that $\pi(\mathbf{K}_{\mathbf{a}}(x)) = \{z_{\lambda,\mathbf{a}}(x)\mathrm{I}\}$ for any spacelike direction $\mathbf{a}$, where $\{z_{\lambda,\mathbf{a}}(x)\}$ is a net of complex numbers. Let $\theta(\pi_\lambda(x^*x)) = \operatorname{lub}_{\mathbf{a} \in \mathbf{D}} z_{\lambda,\mathbf{a}}(x^*x)$ for x in $\mathbf{A}$. Then θ extends to a map $\bar{\theta}$ from $\mathbf{K}^c$ into X^2, where $\bar{\theta}(\pi_\lambda)$ is the net $\{\operatorname{lub}_{\mathbf{a} \in \mathbf{D}} z_{\lambda,\mathbf{a}}(x); x \in \mathbf{A}\}$ in X^2. Then by the construction of the representations π_λ it is easily seen that the map $\bar{\theta}$ is one-to-one. We shall also assume that $\bar{\theta}$ is onto. Otherwise there is a bijection between $\mathbf{K}^c$ and a subset Y of X^2 and the proof of the lemma would be valid with X^2 replaced by Y. Furthermore, the Borel structure of X^2 coincides with the quotient Borel structure induced by the Borel structure of $\mathbf{K}^c$ under $\bar{\theta}$. To show this, we let $V(b_1,\dots,b_n)$ be the cylindrical set

$$\{(\chi_1,\dots,\chi_n): \chi_k = b_k;\ k = 1,\dots,n\} \text{ in } X^2.$$

Then for $x \in \mathbf{A}$,

$$\begin{aligned}\bar{\theta}^{-1}(V(b_1,\dots,b_n)) &= \{\pi_{\lambda_j}: \pi_{\lambda_j} \in \mathbf{K}^c \text{ such that } \bar{\theta}(\pi_{\lambda_j}(x^*x)) \\ &= \operatorname*{lub}_{\mathbf{a}\in\mathbf{D}} z_{\lambda_j,\mathbf{a}}(x^*x) = b_j \text{ for } j = 1,\dots,n\}.\end{aligned}$$

Thus, $V(b_1,\dots,b_n)$ is a Borel set in the quotient Borel structure. The quotient Borel structure contains the original Borel structure of X^2 and thus the quotient Borel structure is countably separated. Since the identity map of X^2 onto itself from the quotient Borel structure to the original Borel structure is a Borel map, it follows that two Borel structures coincide (Ref. 10; Theorem 4.2 and Theorem 5.1).

Let $\mathbf{K}$ denote the set of unitary equivalence classes of representations of $\hat{\mathbf{A}}^c_{\rm sp}$ contained in $\mathbf{K}^c$.

For any two representations $\pi_{\lambda_1}, \pi_{\lambda_2}$, we have

$$\begin{aligned}&|\bar{\theta}(\pi_{\lambda_1}(x^*x)) - \bar{\theta}(\pi_{\lambda_2}(x^*x))| \\ &\quad = |\operatorname*{lub}_{\mathbf{a}\in\mathbf{D}} z_{\lambda_1,\mathbf{a}}(x^*x) - \operatorname*{lub}_{\mathbf{a}\in\mathbf{D}} z_{\lambda_2,\mathbf{a}}(x^*x)| \\ &\quad \leqslant \operatorname*{lub}_{\mathbf{a},\mathbf{a}'\in\mathbf{D}} |z_{\lambda_1,\mathbf{a}}(x^*x) - z_{\lambda_2,\mathbf{a}'}(x^*x)|.\end{aligned}$$

We assert that the representations π_{λ_1} and π_{λ_2} are unitary equivalent if and only if for any spacelike direction $\mathbf{a},\mathbf{a}'$ in D

$$|z_{\lambda_1,\mathbf{a}}(x^*x) - z_{\lambda_2,\mathbf{a}'}(x^*x)| \leqslant \delta \quad (x \in \mathbf{A})$$

for a rational number δ. The if part is obvious. To prove the only if part assume that

$$|z_{\lambda_1,\mathbf{a}}(x^*x) - z_{\lambda_2,\mathbf{a}'}(x^*x)| \leqslant \delta_{\mathbf{a},\mathbf{a}'} \quad (x \in \mathbf{A}),$$

where $\{\delta_{\mathbf{a},\mathbf{a}'}\}$ is a null sequence of rational numbers. Hence

$$\begin{aligned}F_{\lambda_1}(x^*x) &= \operatorname*{lub}_{\mathbf{a}\in\mathbf{D}}\{z_{\lambda_1,\mathbf{a}}(x^*x)\mathrm{I}: z_{\lambda_1,\mathbf{a}}(x)\mathrm{I} \in \pi(\mathbf{K}_\mathbf{a}(x))\} \\ &= F_{\lambda_2}(x^*x).\end{aligned}$$

Then by the above construction the vacuum functionals L_{λ_1} and L_{λ_2} coincide and therefore π_{λ_1} is unitary equivalent to π_{λ_2}.

Let $\tilde{\theta}$ be the one-to-one map defined by $\bar{\theta}$ from $\mathbf{K}$ onto the set $\tilde{X}^2$ of Δ^2-equivalence classes of X^2, and let $\tilde{X}^2$ have the quotient Borel structure derived from X^2. We shall show that $\tilde{\theta}$ is a Borel isomorphism with respect to the quotient Borel structure $\mathbf{B}_q$ on $\mathbf{K}$ as derived from $\mathbf{K}^c$. Let $\tilde{E}$ be a subset of $\tilde{X}^2$. Then $\tilde{E}$ is a Borel set if and only if the set E of elements of $\tilde{E}$ is a Borel set and this is a Borel set if and only if $\bar{\theta}^{-1}(E)$ is a Borel set. However, $\bar{\theta}^{-1}(E)$ contains each unitary equivalence in $\mathbf{K}^c$, that it meets, and so $\bar{\theta}^{-1}(E)^{\sim}$ of unitary equivalance classes of elements of $\bar{\theta}^{-1}(E)$ is in $\mathbf{B}_{\rm q}$. Since $\bar{\theta}^{-1}(E)^{\sim} = \tilde{\theta}^{-1}(E)$, $\tilde{\theta}$ is a Borel isomorphism.

X^2 is a compact group, Δ^2 is a dense subgroup, and $\tilde{X}^2 = X^2/\Delta^2$. It follows from Ref. 10 (Theorem 7.2) that $\tilde{X}^2$ is not metrically countably separated. This implies that $\mathbf{K}$ with the Borel structure $\mathbf{B}_q$ is not metrically countably separated. Let $\mathbf{B}_s$ be the Borel structure which $\mathbf{K}$ inherits as a subspace of $\hat{\mathbf{A}}_{\rm sp}$. Then from a similar argument as in Lemma 3.5 it follows that $\mathbf{B}_s \subset \mathbf{B}_q$. This implies that $\mathbf{K}$ with the Borel structure $\mathbf{B}_s$ is not metrically countably separated and therefore $\hat{\mathbf{A}}_{\rm sp}$ is not metrically countably separated. This completes the proof of Theorem 3.1.

ACKNOWLEDGMENTS

I wish to thank Professor D. Buchholz, Professor R. F. Streater, and Professor J. Yngvason for several discussions. I am also grateful to Professor H. J. Borchers and Professor S. Doplicher for helpful comments.

[1]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[2]H. J. Borchers, "Locality and Spectrum Condition for the Translations," preprint, 1982.
[3]D. Buchholz and K. Fredenhagen, Commun. Math. Phys. **84**, 1 (1982).
[4]H. J. Borchers, R. Haag, and B. Schroer, Nuovo Cimento **29**, 148 (1963).
[5]S. Doplicher, R. Haag, and J. E. Roberts, Commun. Math. Phys. **13**, 1 (1969).
[6]J. Glimm, Ann. Math. **73**, 572 (1961).
[7]H. J. Borchers, "On the Existence of Vacuum States" (unpublished, 1982).
[8]H. J. Borchers, "Translation Group and Spectrum Condition," preprint, 1982.
[9]H. J. Borchers, Commun. Math. Phys. **1**, 57 (1965).
[10]G. W. Mackey, Trans. Am. Math. Soc. **85**, 134 (1957).

Existence theorem for gauge algebra

I. A. Batalin
Physical Lebedev Institute, Academy of Sciences, Leninsky Prospect 53, Moscow, 117924, USSR
G. A. Vilkovisky
State Committee of Standards, Leninsky Prospect 9, Moscow, 117049, USSR

(Received 30 April 1984; accepted for publication 8 June 1984)

A general gauge action is defined by postulating a minimum of its properties necessary for the existence of loop expansion in the quantum theory. The structure of the general gauge algebra is derived from these postulates. The proof of existence of the structure functions, lacking in previous works, is given. The question of uniqueness is also completely cleared up. Some unsolved problems are discussed in connection with the covariant-quantization conjecture.

I. INTRODUCTION

The transformations of invariance of gauge fields in the natural basis may not form a Lie group. Furthermore, the algebra of infinitesimal transformations may not be closed. This nonclosure is, however, of a special form, and the corresponding algebraic construction is known as "open gauge algebra." The form of gauge algebra determines the quantization rules for gauge fields and is thus of paramount importance in quantum theory.

The open gauge algebra and the corresponding quantization rules were first discovered in the canonical formalism of gravity theory.[1,2] The subject received further development in works[3–7] stimulated by the discovery of open algebra in supergravity.[8] The full structure of open algebra was derived in Refs. 4 and 9 but these derivations were to a certain extent heuristic. In Ref. 10 the generating equation was obtained, containing all structure relations of gauge algebra. The existence theorem for this equation was formulated in Ref. 10, but the proof was omitted. It is the purpose of the present work to fill up these gaps.

In the present paper we consider the general gauge action, postulating a minimum of its properties, necessary for the existence of loop expansion in the quantum theory. We then systematically derive the algebraic consequences of these postulates. The range of questions arising at the level of finite gauge transformations is considered in Refs. 11 and 12. Together with Ref. 12 the present paper may serve as an introduction to the theory of open group. The present construction is, however, applicable only to *irreducible* gauge field theories (see Sec. VI). The generalization to reducible theories can be found in Refs. 13 and 14.

The plan of this paper is the following. The postulates of gauge theory are formulated and discussed in Sec. II. In Sec. III a detailed derivation is given of the lowest-order relations of gauge algebra. Sec. IV introduces the closed description of gauge algebra. In Sec. V we present the full proof of the principal existence theorem. A lemma needed for the proof is given in Appendix A. In Sec. VI we settle the question of uniqueness, discuss the application to relativistic field theories, and point out some unsolved problems. Appendix B contains a classification of theories with a degenerate Hessian of the action. Our postulates single out a special class of such theories. We show that a theory must belong to this class as a necessary condition for the applicability of the standard loop technique.

In some parts of the present consideration we essentially use the ideas contained in the previous works, especially in Ref. 4.

Notation and conventions: We shall be dealing with sets of boson and fermion variables. The Grassmann parity of a quantity A will be denoted by $\epsilon(A)$. Right and left derivatives will be ∂_r and ∂_l.

The rank of a matrix is generally defined as the maximum size of its invertible square minor. In the case of an even-parity matrix M one may speak about two ranks: those of the Bose–Bose and Fermi-Fermi blocks. These ranks will be denoted by $\text{rank}_{\pm} M$. The rank of an even-parity matrix is $\text{rank}\, M = \text{rank}_+ M + \text{rank}_- M$.

We shall use the condensed notation[15] for the gauge field: φ^i, $i = 1,...,n$; $n = n_+ + n_-$, where $n_+(n_-)$ is the number of boson (fermion) components. This means that in fact we shall work with the finite-dimensional model. The results transfer to Euclidean field theory in the usual way.

II. POSTULATES OF GAUGE THEORY

The field φ^i, $i = 1,...,n = n_+ + n_-$, $\epsilon(\varphi^i)\equiv\epsilon_i$, is a gauge field, if its action $\mathscr{S}(\varphi)$ is a boson satisfying the following two postulates.

Postulate 1: There exists at least one stationary point φ_0:

$$\mathscr{S}_i\big|_{\varphi=\varphi_0} = 0, \quad \mathscr{S}_i \equiv \frac{\partial_r \mathscr{S}}{\partial\varphi^i}, \tag{2.1}$$

and $\mathscr{S}(\varphi)$ is regular (infinitely differentiable) in its neighborhood.

Postulate 2: The set of field equations can be divided into two subsets:

$$\mathscr{S}_\alpha = 0, \quad \alpha = 1,...,m; \quad 0\leqslant m\leqslant n \tag{2.2}$$

and

$$\mathscr{S}_A = 0, \quad A = 1,...,n-m, \tag{2.3}$$

in such a way that

$$\text{rank}\, \frac{\partial_l \mathscr{S}_A}{\partial\varphi^i}\bigg|_{\varphi=\varphi_0} = n-m, \tag{2.4}$$

and at least those solutions of (2.3) that lie in some neighborhood of φ_0 satisfy (2.2) identically.

The theory must of course be invariant under arbitrary regular reparametrizations of the field φ. To see the invariance of Postulate 2, we note that it can be equivalently formulated as follows.

Postulate 2′: There exists (at least locally, in a neighborhood of φ_0) a smooth m-dimensional surface Σ ($\varphi_0 \in \Sigma$), on which field equations are fulfilled

$$\mathscr{S}_i|_\Sigma = 0, \tag{2.5}$$

and

$$\operatorname{rank} \mathscr{S}_{ji}|_\Sigma = n - m, \tag{2.6}$$

$$\mathscr{S}_{ji} \equiv \frac{\partial_l \partial_r \mathscr{S}}{\partial\varphi^j\, \partial\varphi^i}. \tag{2.7}$$

(Note, that the rank of $\mathscr{S}_{ji}$ is reparametrization invariant at stationary points of $\mathscr{S}$.) The surface Σ will be called the stationary orbit.

The equivalence of Postulates 2 and 2′ is based on the implicit function theorem. Clearly, (2.3) is the equation of Σ, and (2.4) is the condition that Σ is m-dimensional. Equation (2.5) follows from the last requirement of Postulate 2.

Let θ^α, $\alpha = 1,\ldots,m = m_+ + m_-$, $\epsilon(\theta^\alpha) \equiv \epsilon_\alpha$, be parameters on Σ, where $m_+ (m_-)$ is the number of boson (fermion) parameters. Let

$$\Sigma\colon\ \varphi^i = f^i(\theta), \tag{2.8}$$

$$\operatorname{rank} \frac{\partial_r f^i(\theta)}{\partial\theta^\alpha} \equiv m, \quad f^i|_{\theta=0} = \varphi_0^i. \tag{2.9}$$

Then

$$\mathscr{S}_i(f(\theta)) \equiv 0, \quad \mathscr{S}_{ji}(f(\theta)) \frac{\partial_r f^i(\theta)}{\partial\theta^\alpha} \equiv 0. \tag{2.10}$$

Thus tangent vectors to Σ are zero-eigenvalue eigenvectors of $\mathscr{S}_{ji}|_\Sigma$. We arrive at one more equivalent formulation of Postulate 2.

Postulate 2″: The Hessian of $\mathscr{S}(\varphi)$ is degenerate on an m-dimensional surface, which passes through φ_0 and whose tangent space at each point coincides with the null space of the Hessian.

Gauge theory is thus defined by the properties of its classical solutions. The purpose of the present work is the derivation of properties of gauge theory off the classical solutions. We shall show that at least in some (n-dimensional) neighborhood of φ_0 there exists a sequence of quantities (structure functions) which form *the gauge algebra*, a construction generalizing the Lie algebra. The zeroth-order structure function is the gauge action $\mathscr{S}(\varphi)$ itself. All higher-order structure functions can be explicitly expressed through $\mathscr{S}(\varphi)$.

The first-order relations of the gauge algebra are Noether identities. The following statement is equivalent to Postulate 2 (or 2′ or 2″).

Postulate 2‴: At least in some neighborhood of φ_0 the gauge action satisfies the Noether identities

$$\mathscr{S}_i R^i_\alpha \equiv 0, \quad \alpha = 1,\ldots,m = m_+ + m_-, \tag{2.11}$$

where R^i_α are regular functions, such that

$$\operatorname{rank}_\pm R^i_\alpha|_{\varphi=\varphi_0} = m_\pm, \quad \epsilon(R^i_\alpha) = \epsilon_i + \epsilon_\alpha, \tag{2.12}$$

and

$$\operatorname{rank}_\pm \mathscr{S}_{ji}|_{\varphi=\varphi_0} = n_\pm - m_\pm. \tag{2.13}$$

Let us show that 2‴ follows from Postulate 2. According to (2.4), Eqs. (2.3) are solvable in a neighborhood of φ_0 with respect to some $n - m$ variables φ'^A. Let φ'^α be the remaining m field variables. Then the following reparametrization

$$\varphi^i \equiv (\varphi'^A, \varphi'^\alpha) \leftrightarrow (J_A, \varphi'^\alpha), \qquad J_A = \mathscr{S}_A(\varphi) \tag{2.14}$$

is regular in a neighborhood of φ_0. Consider the functions $\mathscr{S}_\alpha$ in the new parametrization and define

$$R^A_\alpha = -\frac{\partial_l}{\partial J_A} \int_0^1 \frac{dx}{x} (\mathscr{S}_\beta R^\beta_\alpha)|_{J\to xJ}, \tag{2.15}$$

where R^β_α is an arbitrary regular invertible matrix. The convergence of the above integral and the regularity of R^A_α in a neighborhood of φ_0 are guaranteed by Postulates 1 and 2. Indeed, $\mathscr{S}_\beta$ is analytic in J near $J = 0$ according to Postulate 1, and $\mathscr{S}_\beta|_{J=0} = 0$ in consequence of the last requirement of Postulate 2. Multiplying (2.15) by $\mathscr{S}_A = J_A$, we obtain the relations

$$\mathscr{S}_A R^A_\alpha + \mathscr{S}_\beta R^\beta_\alpha = (\mathscr{S}_\beta R^\beta_\alpha)|_{J=0} = 0, \tag{2.16}$$

which are the Noether identities (2.11) with

$$R^i_\alpha = (R^A_\alpha, R^\beta_\alpha). \tag{2.17}$$

The condition (2.12) is satisfied by virtue of the above choice of R^β_α. We must still prove (2.13). Differentiating the Noether identities, we find

$$(\mathscr{S}_{ji} R^i_\alpha)_{\varphi=\varphi_0} = 0. \tag{2.18}$$

It follows from (2.18) that the matrix $\mathscr{S}_{ji}|_{\varphi=\varphi_0}$ has m zero-eigenvalue eigenvectors $R^i_\alpha|_{\varphi=\varphi_0}$ satisfying condition (2.12). Consequently,

$$\operatorname{rank} \mathscr{S}_{ji}|_{\varphi=\varphi_0} \leqslant n - m. \tag{2.19}$$

On the other hand,

$$\operatorname{rank} \mathscr{S}_{ji}|_{\varphi=\varphi_0} \geqslant n - m \tag{2.20}$$

from (2.4). This gives (2.13).

Conversely, let us show that Postulate 2 follows from 2‴. According to (2.13), there exists a set of $n - m$ field variables φ^A, such that

$$\operatorname{rank} \mathscr{S}_{jA}|_{\varphi=\varphi_0} = n - m. \tag{2.21}$$

Let φ^α be the remaining m variables. Then the Noether identities (2.11) take the form

$$\mathscr{S}_A R^A_\beta + \mathscr{S}_\alpha R^\alpha_\beta = 0. \tag{2.22}$$

We have to show that R^α_β is invertible, i.e.,

$$\operatorname{rank} R^\alpha_\beta = m \tag{2.23}$$

in a neighborhood of φ_0, because then equations $\mathscr{S}_\alpha = 0$ and $\mathscr{S}_A = 0$ are just Eqs. (2.2) and (2.3) of Postulate 2. To prove (2.23), we differentiate (2.22):

$$(\mathscr{S}_{jA} R^A_\beta + \mathscr{S}_{j\alpha} R^\alpha_\beta)|_{\varphi=\varphi_0} = 0, \tag{2.24}$$

and suppose that there exist such $\lambda^\beta \neq 0$ that

$$R^\alpha_\beta|_{\varphi=\varphi_0} \lambda^\beta = 0. \tag{2.25}$$

From (2.25) and (2.24) we find

$$(\mathscr{S}_{,jA} R^{A}_{\beta})|_{\varphi=\varphi_0} \lambda^{\beta} = 0, \tag{2.26}$$

whence

$$R^{A}_{\beta}|_{\varphi=\varphi_0} \lambda^{\beta} = 0 \tag{2.27}$$

due to (2.21). Equations (2.27) and (2.25) give

$$R^{i}_{\beta}|_{\varphi=\varphi_0} \lambda^{\beta} = 0, \tag{2.28}$$

which contradicts (2.12). Hence, the supposition (2.25) is wrong, and

$$\operatorname{rank} R^{\alpha}_{\beta}|_{\varphi=\varphi_0} = m. \tag{2.29}$$

By regularity the equality (2.23) holds also in a neighborhood of φ_0.

The Noether identities with properties (2.11)–(2.13) are usually considered as the definition of gauge theory. From the present standpoint R^{i}_{α} are the off-shell structure functions which can be expressed through the gauge action as shown above.

The vectors R^{i}_{α} may be interpreted as the generators of infinitesimal transformations

$$\delta\varphi^{i} = R^{i}_{\alpha}\delta\theta^{\alpha}, \quad \epsilon(\delta\theta^{\alpha}) = \epsilon_{\alpha} \tag{2.30}$$

with parameters $\delta\theta^{\alpha}$, leaving the gauge action invariant. However, the existence of finite gauge transformations is not obvious. It does not follow from anywhere that R^{i}_{α} generate a Lie group. In fact the group properties of R^{i}_{α}, which follow from the above postulates, are generally more complicated. At the algebraic level these properties are considered in the present paper (and previous works). At the group level they are considered in Refs. 11 and 12. Here we shall make only the following remark.

Since tangent vectors to the surface (2.8) belong to the null space of $\mathscr{S}_{,ji}|_{\Sigma}$, they must be linear combinations of $R^{i}_{\alpha}|_{\Sigma}$. As a result we obtain the differential equations of the stationary orbit

$$\Sigma: \quad \frac{\partial_r f^{i}(\theta)}{\partial\theta^{\alpha}} = R^{i}_{\beta}(f(\theta))\Lambda^{\beta}_{\alpha}(\theta), \tag{2.31}$$

$$f^{i}|_{\theta=0} = \varphi^{i}_{0}, \tag{2.32}$$

where $\Lambda^{\beta}_{\alpha}(\theta)$ is some nonsingular matrix. Equations (2.31) are the Lie equations. Thus the integrability of the Lie equations with initial data satisfying (2.1) is guaranteed by the postulates. The Lie equations (2.31) with arbitrary initial data are generally nonintegrable. Nevertheless, as shown in Ref. 12, the nonstationary orbits, i.e., finite gauge transformations, always exist and satisfy some generalized Lie equations.

The present postulates can be justified. In Appendix B we show that they are the necessary conditions for the existence of loop expansion in quantum theory.

Our consideration of gauge theory is purely local. It is confined to a neighborhood of one stationary point: φ_0. (If the action has several stationary points not belonging to one and the same orbit, then the neighborhood of each of them must be considered separately.) Such a local consideration suffices if in the quantum theory we confine ourselves to loop expansion.

III. LOWEST-ORDER RELATIONS OF GAUGE ALGEBRA

Let us go on to deriving the consequences of the above postulates. First of all they permit us to find the general solutions of equations

$$R^{i}_{\alpha}x^{\alpha} = 0 \tag{3.1}$$

and

$$\mathscr{S}_{,i}\, y^{i} = 0. \tag{3.2}$$

According to (2.12) the general solution of Eq. (3.1) is

$$x^{\alpha} \equiv 0. \tag{3.3}$$

We shall show that the general regular solution of Eq. (3.2) is of the form

$$y^{i} = R^{i}_{\mu} A^{\mu} + \mathscr{S}_{,k} B^{ik}, \tag{3.4}$$

where A^{μ} and B^{ik} are arbitrary regular functions, and B^{ik} possesses the antisymmetry property

$$B^{ik} = -(-1)^{\epsilon_i\epsilon_k} B^{ki}. \tag{3.5}$$

Evidently, (3.4) is the solution of Eq. (3.2). To prove that it is the general regular solution, we rewrite Eq. (3.2) as

$$\mathscr{S}_{,A}\, y^{A} + \mathscr{S}_{,\alpha}\, y^{\alpha} = 0, \tag{3.6}$$

where $\mathscr{S}_{,A}$ and $\mathscr{S}_{,\alpha}$ are the subsets from Postulate 2, and use the Noether identities (2.22). Eq. (3.6) takes the form

$$\mathscr{S}_{,A} z^{A} = 0, \tag{3.7}$$

$$z^{A} \equiv y^{A} - R^{A}_{\beta}(R^{-1})^{\beta}_{\alpha}\, y^{\alpha}. \tag{3.8}$$

Next we make the reparametrization (2.14) and differentiate (3.7) with respect to J_A,

$$z^{A} + J_B \frac{\partial_l z^{A}}{\partial J_B} = J_B P^{AB}, \tag{3.9}$$

$$P^{AB} \equiv \frac{\partial_l z^{A}}{\partial J_B} - (-1)^{\epsilon_A\epsilon_B} \frac{\partial_l z^{B}}{\partial J_A}. \tag{3.10}$$

Making the replacement in Eq. (3.9)

$$J_A \to xJ_A, \tag{3.11}$$

where x is a numerical parameter, we find

$$\frac{d}{dx} xz^{A}(xJ) = xJ_B P^{AB}(xJ), \tag{3.12}$$

$$z^{A}(J) - \lim_{x\to 0} xz^{A}(xJ) = J_B \int_0^1 xP^{AB}(xJ)dx. \tag{3.13}$$

Since the solution y^{i} is supposed to be regular near $J = 0$, the integral on the right-hand side of (3.13) converges. Also

$$\lim_{x\to 0} xz^{A}(xJ) = 0. \tag{3.14}$$

As a result we obtain representation (3.4) with

$$A^{\mu} = (R^{-1})^{\mu}_{\alpha}\, y^{\alpha}, \quad B^{\alpha\beta} = 0, \quad B^{\alpha A} = B^{A\alpha} = 0, \tag{3.15}$$

$$B^{AB} = \int_0^1 xP^{AB}(xJ)dx.$$

Expression (3.4) shows that any infinitesimal transformation, leaving the gauge action invariant, is a combination of transformations (2.30) and trivial transformations

$$\delta^{\text{triv}}\varphi^{i} = \mathscr{S}_{,k}\delta\theta^{ik}, \quad \delta\theta^{ik} = -(-1)^{\epsilon_i\epsilon_k}\delta\theta^{ki},$$

which do not move stationary points. The existence of trivial transformations causes the complicacy of the general off-shell gauge algebra.

As mentioned above, the first-order relations of the gauge algebra are the Noether identities (2.11). To derive the further relations we define the operator

$$\hat{\Gamma}_\beta(\cdots) = \frac{\partial_r(\cdots)}{\partial\varphi^i} R^i_\beta \tag{3.16}$$

and apply it to (2.11) with the subsequent antisymmetrization in group indices. The result is

$$\mathscr{S}_i\, y^i_{\alpha\beta} = 0, \tag{3.17}$$

where

$$y^i_{\alpha\beta} \equiv \frac{\partial_r R^i_\alpha}{\partial\varphi^k} R^k_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta} \frac{\partial_r R^i_\beta}{\partial\varphi^k} R^k_\alpha. \tag{3.18}$$

Equation (3.17) is of the type (3.2). Hence, there exist such functions $T^\mu_{\alpha\beta}$ and $E^{ik}_{\alpha\beta}$ that

$$\frac{\partial_r R^i_\alpha}{\partial\varphi^k} R^k_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta} \frac{\partial_r R^i_\beta}{\partial\varphi^k} R^k_\alpha = -R^i_\mu T^\mu_{\alpha\beta} - \mathscr{S}_k E^{ik}_{\alpha\beta}. \tag{3.19}$$

Equations (3.19) are the second-order relations of the gauge algebra, and $T^\mu_{\alpha\beta}, E^{ik}_{\alpha\beta}$ are the new structure functions. The formulas of the type (3.15) give explicit expressions of these new functions through R^i_α. It follows from these expressions that $T^\mu_{\alpha\beta}$ and $E^{ik}_{\alpha\beta}$ are regular and possess the antisymmetry properties

$$T^\mu_{\alpha\beta} = -(-1)^{\epsilon_\alpha\epsilon_\beta} T^\mu_{\beta\alpha}, \tag{3.20}$$

$$E^{ik}_{\alpha\beta} = -(-1)^{\epsilon_\alpha\epsilon_\beta} E^{ik}_{\beta\alpha} = -(-1)^{\epsilon_i\epsilon_k} E^{ki}_{\alpha\beta}. \tag{3.21}$$

Expression (3.19) gives the general form of the commutator of gauge transformations. In the particular case, when $E^{ik}_{\alpha\beta} = 0$, the algebra is closed. If also $T^\mu_{\alpha\beta} = \text{const}$, we have a Lie algebra. However, generally the off-shell gauge algebra is open because of the admixture of trivial transformations.

The third-order relations of the gauge algebra are the generalized Jacobi identities. Applying the operator (3.16) to (3.19) with the subsequent cyclic permutation of group indices, also using Eqs. (2.11) and (3.19), we obtain

$$R^i_\mu X^\mu_{\alpha\beta\delta} + \mathscr{S}_k Y^{ik}_{\alpha\beta\delta} = 0, \tag{3.22}$$

where

$$X^\mu_{\alpha\beta\delta} \equiv (-1)^{\epsilon_\alpha\epsilon_\delta}\left(\frac{\partial_r T^\mu_{\alpha\beta}}{\partial\varphi^i} R^i_\delta + T^\mu_{\alpha\gamma} T^\gamma_{\beta\delta}\right) + \text{cycl. perm. } (\alpha,\beta,\delta), \tag{3.23}$$

$$\begin{aligned} Y^{ik}_{\alpha\beta\delta} \equiv (-1)^{\epsilon_\alpha\epsilon_\delta}\Bigg(&\frac{\partial_r E^{ik}_{\alpha\beta}}{\partial\varphi^j} R^j_\delta + E^{ik}_{\alpha\gamma} T^\gamma_{\beta\delta} \\ &- (-1)^{\epsilon_\alpha\epsilon_i} \frac{\partial_r R^k_\alpha}{\partial\varphi^j} E^{ij}_{\beta\delta} \\ &+ (-1)^{(\epsilon_i+\epsilon_\alpha)\epsilon_k} \frac{\partial_r R^i_\alpha}{\partial\varphi^j} E^{kj}_{\beta\delta}\Bigg) \\ &+ \text{cycl. perm. } (\alpha,\beta,\delta). \end{aligned} \tag{3.24}$$

We shall prove that there exist such new structure functions $F^{\alpha k}_{\gamma\beta\delta}$ and $D^{ikj}_{\gamma\beta\delta}$ that

$$X^\mu_{\alpha\beta\delta} = -\mathscr{S}_k F^{\mu k}_{\alpha\beta\delta} \tag{3.25}$$

and

$$\begin{aligned} Y^{ik}_{\alpha\beta\delta} &+ (-1)^{\epsilon_i\epsilon_\mu} R^k_\mu F^{\mu i}_{\alpha\beta\delta} \\ &- (-1)^{(\epsilon_i+\epsilon_\mu)\epsilon_k} R^i_\mu F^{\mu k}_{\alpha\beta\delta} = -\mathscr{S}_j D^{ikj}_{\alpha\beta\delta}. \end{aligned} \tag{3.26}$$

The new structure functions are regular and possess the following antisymmetry properties:

$$F^{\mu i}_{\alpha\beta\delta} = -(-1)^{\epsilon_{\alpha\beta\delta}} F^{\mu i}_{\beta\alpha\delta} = -(-1)^{\epsilon_{\alpha\beta\delta}} F^{\mu i}_{\alpha\delta\beta}, \tag{3.27}$$

$$D^{ikj}_{\alpha\beta\delta} = -(-1)^{\epsilon_{\alpha\beta\delta}} D^{ikj}_{\beta\alpha\delta} = -(-1)^{\epsilon_{\alpha\beta\delta}} D^{ikj}_{\alpha\delta\beta}, \tag{3.28}$$

$$\epsilon_{\alpha\beta\delta} \equiv \epsilon_\alpha\epsilon_\beta + \epsilon_\beta\epsilon_\delta + \epsilon_\delta\epsilon_\alpha, \tag{3.29}$$

$$D^{ikj}_{\alpha\beta\delta} = -(-1)^{\epsilon_i\epsilon_k} D^{kij}_{\alpha\beta\delta} = -(-1)^{\epsilon_k\epsilon_j} D^{ijk}_{\alpha\beta\delta}. \tag{3.30}$$

Equations (3.25) are the generalized Jacobi identities. Equations (3.26) are the fourth-order relations of the gauge algebra, which have no analogy in the Lie algebra.

The authors of Ref. 9 obtained all structure relations of the gauge algebra operating only with the conditions of "irreducibility" and "completeness" in the form (3.3) and (3.4). In fact these conditions are insufficient to prove already the fourth-order structure relations (3.26) with the antisymmetry properties (3.30). The correct proof requires the full use of postulates explicitly formulated above. The proof is the following.

At first one considers Eq. (3.22) at $i = \sigma$, where R^σ_μ is the invertible minor of R^i_μ. This gives

$$X^\mu_{\alpha\beta\delta} = -\mathscr{S}_k (R^{-1})^\mu_\sigma Y^{\sigma k}_{\alpha\beta\delta}(-1)^{(\epsilon_\mu+\epsilon_\sigma)\epsilon_k}, \tag{3.31}$$

which is relation (3.25). Next one uses (3.25) in (3.22) to obtain

$$\mathscr{S}_k Z^{ik}_{\alpha\beta\delta} = 0, \tag{3.32}$$

where

$$\begin{aligned} Z^{ik}_{\alpha\beta\delta} \equiv Y^{ik}_{\alpha\beta\delta} &+ (-1)^{\epsilon_i\epsilon_\mu} R^k_\mu F^{\mu i}_{\alpha\beta\delta} \\ &- (-1)^{(\epsilon_i+\epsilon_\mu)\epsilon_k} R^i_\mu F^{\mu k}_{\alpha\beta\delta}. \end{aligned} \tag{3.33}$$

The second term on the right-hand side of (3.33) can be added by virtue of the Noether identities.

Equation (3.32) is of the type (3.2). However, its general solution in the form (3.4) is not yet the structure relation (3.26). Even the use of explicit expressions (3.15) is insufficient to prove the antisymmetry properties (3.30).

To derive (3.26), note that Eq. (3.31) does not uniquely determine $F^{\mu k}_{\alpha\beta\delta}$ in (3.25). In particular, one may put

$$\begin{aligned} F^{\mu k}_{\alpha\beta\delta} = (R^{-1})^\mu_\sigma Y^{\sigma k}_{\alpha\beta\delta}(-1)^{(\epsilon_\mu+\epsilon_\sigma)\epsilon_k} \\ + R^k_\gamma M^{\gamma\mu}_{\alpha\beta\delta}, \end{aligned} \tag{3.34}$$

where $M^{\gamma\mu}_{\alpha\beta\delta}$ are arbitrary regular functions possessing the cyclic antisymmetry (3.27). This arbitrariness can be used to make the components of $Z^{ik}_{\alpha\beta\delta}$ with $k = \nu$ (or $i = \nu$) vanish. Indeed, the quantity $Z^{ik}_{\alpha\beta\delta}$ at $k = \nu$ equals

$$Z^{iv}_{\alpha\beta\delta} = Y^{iv}_{\alpha\beta\delta} + Y^{vi}_{\alpha\beta\delta}(-1)^{\epsilon_v\epsilon_i}$$
$$- R^{\,i}_{\ \sigma}(R^{-1})^{\sigma}_{\mu} Y^{\mu v}_{\alpha\beta\delta}(-1)^{\epsilon_v(\epsilon_i+\epsilon_\mu)}$$
$$+(-1)^{\epsilon_i\epsilon_\mu} R^{\,v}_{\ \mu} R^{\,i}_{\ \gamma} M^{\gamma\mu}_{\alpha\beta\delta}$$
$$-(-1)^{(\epsilon_i+\epsilon_\mu)\epsilon_v} R^{\,i}_{\ \mu} R^{\,v}_{\ \gamma} M^{\gamma\mu}_{\alpha\beta\delta}, \qquad (3.35)$$

as follows from (3.33) and (3.34). Choosing $M^{\gamma\mu}_{\alpha\beta\delta}$ as

$$M^{\gamma\sigma}_{\alpha\beta\delta} = \tfrac{1}{2}(R^{-1})^{\gamma}_{v}(R^{-1})^{\sigma}_{\mu} Y^{v\mu}_{\alpha\beta\delta}(-1)^{\epsilon_v\epsilon_\sigma} \qquad (3.36)$$

and taking into consideration the fact that the quantity $Y^{ik}_{\alpha\beta\delta}$ is antisymmetric in the upper indices, we obtain

$$Z^{iv}_{\alpha\beta\delta} = 0. \qquad (3.37)$$

Thus the structure functions $F^{\mu k}_{\alpha\beta\delta}$ can be defined in such a way that only the components $Z^{AB}_{\alpha\beta\delta}$ of $Z^{ik}_{\alpha\beta\delta}$ survive. For these components Eq. (3.32) takes the form

$$\mathscr{S}_B Z^{AB}_{\alpha\beta\delta} = 0. \qquad (3.38)$$

Using the reparametrization (2.14) and differentiating (3.38) with respect to $J_B = \mathscr{S}_B$, we obtain

$$Z^{AB}_{\alpha\beta\delta} = -J_C \frac{\partial_l Z^{AC}_{\alpha\beta\delta}}{\partial J_B}(-1)^{\epsilon_B\epsilon_C}. \qquad (3.39)$$

Since $Z^{AB}_{\alpha\beta\delta}$ is antisymmetric in the upper indices, this can be rewritten as

$$Z^{AB}_{\alpha\beta\delta} = -J_C \frac{1}{2}\left[\frac{\partial_l Z^{AC}_{\alpha\beta\delta}}{\partial J_B}(-1)^{\epsilon_B\epsilon_C} - \frac{\partial_l Z^{BC}_{\alpha\beta\delta}}{\partial J_A}(-1)^{\epsilon_A(\epsilon_B+\epsilon_C)}\right] \qquad (3.40)$$

and finally as

$$Z^{AB}_{\alpha\beta\delta} + \frac{1}{2} J_C \frac{\partial_l Z^{AB}_{\alpha\beta\delta}}{\partial J_C} = -\frac{1}{2} J_C U^{ABC}_{\alpha\beta\delta}, \qquad (3.41)$$

where

$$U^{ABC}_{\alpha\beta\delta} = \frac{\partial_l Z^{AC}_{\alpha\beta\delta}}{\partial J_B}(-1)^{\epsilon_B\epsilon_C} - \frac{\partial_l Z^{BC}_{\alpha\beta\delta}}{\partial J_A}(-1)^{\epsilon_A(\epsilon_B+\epsilon_C)} - \frac{\partial_l Z^{AB}_{\alpha\beta\delta}}{\partial J_C}. \qquad (3.42)$$

Making the replacement

$$J \to xJ \qquad (3.43)$$

in Eq. (3.41), we obtain

$$\frac{d}{dx}(x^2 Z^{AB}_{\alpha\beta\delta}) = -x^2 J_C U^{ABC}_{\alpha\beta\delta}, \qquad (3.44)$$

whence

$$Z^{AB}_{\alpha\beta\delta} = -J_C D^{ABC}_{\alpha\beta\delta}, \qquad (3.45)$$

$$D^{ABC}_{\alpha\beta\delta} \equiv \int_0^1 x^2 U^{ABC}_{\alpha\beta\delta}(xJ)dx. \qquad (3.46)$$

Equations (3.45) and (3.37) give the fourth-order structure relation (3.26), in which only the components $D^{ABC}_{\alpha\beta\delta}$ of $D^{ikj}_{\alpha\beta\delta}$ survive. The antisymmetry properties (3.28)–(3.30) are seen from (3.42).

The further structure relations of the algebra are more and more complicated. Clearly, a more powerful technique is needed to handle them. Such a technique is developed below. In the next section a unique equation is formulated, generating all structure relations of gauge algebra.

IV. THE GENERATING EQUATION FOR GAUGE ALGEBRA

For the closed description of gauge algebra one introduces a phase space in which conjugate variables have the opposite statistics.[10]

Let Φ^A, $A = 1,...,N$, $\epsilon(\Phi^A) = \epsilon_A$, be some set of boson and fermion variables (fields). To each Φ^A one puts into correspondence a new variable (antifield) Φ^*_A of the opposite statistics

$$\epsilon(\Phi^*_A) = \epsilon(\Phi^A) + 1. \qquad (4.1)$$

For functions on the phase space of fields and antifields one defines a binary operation called antibrackets:

$$(X,Y) \equiv \frac{\partial_r X}{\partial \Phi^A}\frac{\partial_l Y}{\partial \Phi^*_A} - \frac{\partial_r X}{\partial \Phi^*_A}\frac{\partial_l Y}{\partial \Phi^A}. \qquad (4.2)$$

The properties of antibrackets are in a sense opposite to the properties of the usual Poisson brackets. One has

$$\epsilon((X,Y)) = \epsilon_X + \epsilon_Y + 1, \quad \epsilon_X \equiv \epsilon(X), \qquad (4.3)$$

$$(X,Y) = -(-1)^{(\epsilon_X+1)(\epsilon_Y+1)}(Y,X), \qquad (4.4)$$

$$(X,YZ) = (X,Y)Z + (-1)^{\epsilon_Y\epsilon_Z}(X,Z)Y, \qquad (4.5)$$

$$(-1)^{(\epsilon_X+1)(\epsilon_Z+1)}(X,(Y,Z)) + \text{cycl. perm. } X,Y,Z = 0. \qquad (4.6)$$

For any fermion F

$$(F,F) \equiv 0, \quad \epsilon(F) = 1, \qquad (4.7)$$

while for a boson B

$$(B,B) = 2\frac{\partial_r B}{\partial \Phi^A}\frac{\partial_l B}{\partial \Phi^*_A} \not\equiv 0, \quad \epsilon(B) = 0. \qquad (4.8)$$

For any X

$$((X,X),X) = (X,(X,X)) \equiv 0. \qquad (4.9)$$

The properties of canonical transformations in the space of fields and antifields are considered in Ref. 12.

The equation

$$(S,S) = 0 \qquad (4.10)$$

for a boson S is called the master equation. It plays the central role in the formulation of quantization rules for gauge theories.[10,13,14]

To analyze the properties of the master equation it is convenient to introduce the collective notation for fields and antifields

$$z^a = \{\Phi^A, \Phi^*_A\}; \quad a = 1,...,2N, \qquad (4.11)$$

and rewrite the antibrackets (4.2) as

$$(X,Y) = \frac{\partial_r X}{\partial z^a}\zeta^{ab}\frac{\partial_l Y}{\partial z^b}, \quad \zeta^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \qquad (4.12)$$

Then Eq. (4.10) takes the form

$$\frac{\partial_r S}{\partial z^a}\zeta^{ab}\frac{\partial_l S}{\partial z^b}=0. \tag{4.13}$$

We shall be interested in the solutions of the master equation, which have a stationary point (where $\partial S/\partial z=0$) and are regular in its neighborhood. Differentiating (4.13) in a neighborhood of the stationary point, we find

$$\frac{\partial_r S}{\partial z^a}\mathscr{R}^a_b=0, \tag{4.14}$$

where

$$\mathscr{R}^a_b\equiv\zeta^{ac}\frac{\partial_l\partial_r S}{\partial z^c\partial z^b}. \tag{4.15}$$

$S(z)$ may be regarded as some action function. Then Eqs. (4.14) are "Noether identities," and the columns of the Hessian (4.15) serve as "generators of gauge transformations." Differentiating (4.14), we find that the matrix of generators is nilpotent at the stationary point:

$$\mathscr{R}^a_b\mathscr{R}^b_c|_{\partial S/\partial z=0}=0. \tag{4.16}$$

Let $r_\pm$ be the rank of the Hessian of S at the stationary point:

$$r_\pm=\operatorname{rank}_\pm\left.\frac{\partial_l\partial_r S}{\partial z^a\,\partial z^b}\right|_{\partial S/\partial z=0},\quad r_++r_-=r. \tag{4.17}$$

Then from (4.16) we have

$$N-r_\pm\geqslant r_\mp, \tag{4.18}$$

or

$$r\leqslant N. \tag{4.19}$$

The solution S of the master equation is called proper if

$$r=N. \tag{4.20}$$

If the solution is proper, then its Hessian (4.15) at the stationary point has no other zero-eigenvalue eigenvectors except those contained in itself. Only proper solutions are of interest in gauge theory.[10,13,14]

Let us include the initial gauge field φ^i into the set Φ^A:

$$\varphi^i\subset\Phi^A \tag{4.21}$$

and require that

$$S(\Phi;\Phi^*)|_{\Phi^*=0}=\mathscr{S}(\varphi), \tag{4.22}$$

where $\mathscr{S}(\varphi)$ is the initial gauge action. The proper solution of the master equation satisfying the boundary condition (4.22) serves as the action in the functional integral of the quantum gauge theory.[10]

It is nontrivial to combine the boundary condition (4.22) and the condition that the solution be proper. The difficulty lies in the fact that if the boundary value of S in (4.22) is the gauge action, then there are initially m zero-eigenvalue eignevectors R^i_α which are not contained in the Hessian. To include R^i_α in the Hessian of S one introduces m new fields (ghosts) c^α and requires[10]

$$c^\alpha\subset\Phi^A, \tag{4.23}$$

$$\left.\frac{\partial_l\partial_r S(\Phi;\Phi^*)}{\partial\varphi^*_i\,\partial c^\alpha}\right|_{\Phi^*=0}=R^i_\alpha(\varphi). \tag{4.24}$$

This defines the statistics of ghosts as opposite to the statistics of the parameters of gauge transformations:

$$\epsilon(c^\alpha)=\epsilon_\alpha+1. \tag{4.25}$$

One also introduces the notion of ghost number and ascribes the following values of this number to the introduced fields and their antifields:

$$\mathrm{gh}(\varphi^i)=0,\quad \mathrm{gh}(c^\alpha)=1, \tag{4.26}$$

$$\mathrm{gh}(\varphi^*_i)=-1,\quad \mathrm{gh}(c^*_\alpha)=-2, \tag{4.27}$$

$$\mathrm{gh}(XY)=\mathrm{gh}(X)+\mathrm{gh}(Y).$$

One then looks for the solution of the master equation as an expansion in powers of the auxiliary fields c,c^*,φ^* with the conserved ghost number equal to zero

$$\mathrm{gh}(S)=0. \tag{4.28}$$

The generic monomial in this expansion is proportional to

$$(c^*)^p(\varphi^*)^s(c)^t \tag{4.29}$$

with

$$t=s+2p. \tag{4.30}$$

We shall see below that the bosonic solution of the master equation, satisfying conditions (4.21)–(4.24) and (4.28), exists and is proper.

One may ask, what relation does it all have to the gauge algebra? The answer is the following. Let us fix the content of the set Φ^A as

$$\Phi^A=\{\varphi^i,c^\alpha\}. \tag{4.31}$$

Correspondingly,

$$\Phi^*_A=\{\varphi^*_i,c^*_\alpha\}, \tag{4.32}$$

and the master equation (4.10) takes the form

$$\frac{\partial_r S}{\partial\varphi^i}\frac{\partial_l S}{\partial\varphi^*_i}+\frac{\partial_r S}{\partial c^\alpha}\frac{\partial_l S}{\partial c^*_\alpha}=0. \tag{4.33}$$

Let us expand the solution

$$S(\Phi;\Phi^*)=S(\varphi,c;\varphi^*,c^*) \tag{4.34}$$

in powers of the auxiliary fields c,c^*,φ^* taking into account ghost-number conservation and the boundary conditions (4.22) and (4.24). The lowest-order terms of this expansion, allowed by the requirement that the ghost number vanish, are

$$\begin{aligned}S(\varphi,c;\varphi^*,c^*)=U&+\varphi^*_iU^i_\alpha c^\alpha\\&+(\varphi^*_i\varphi^*_kU^{ki}_{\alpha\beta}+c^*_\gamma U^\gamma_{\alpha\beta})c^\beta c^\alpha\\&+(\varphi^*_i\varphi^*_k\varphi^*_jU^{jki}_{\alpha\beta\delta}+2\varphi^*_ic^*_\gamma U^{\gamma i}_{\alpha\beta\delta})c^\delta c^\beta c^\alpha\\&+o(c^4),\end{aligned} \tag{4.35}$$

where we kept terms at most cubic in c^α and introduced the notation $U^{\dots}_{\dots}$ for coefficients. These coefficients are functions of φ^i and possess the symmetry properties, which are obvious from (4.35). According to the boundary conditions (4.22) and (4.24), the two lowest-order coefficients are

$$U=\mathscr{S}(\varphi), \tag{4.36}$$

$$U^i_\alpha=R^i_\alpha(\varphi). \tag{4.37}$$

The master equation (4.33) is equivalent to a sequence of relations upon the coefficients $U^{\dots}_{\dots}$. The lowest-order relations are found to be

$$\frac{\partial_r U}{\partial\varphi^i} U^i_{\ \alpha} c^\alpha = 0, \tag{4.38}$$

$$\frac{\partial_r U^i_{\ \alpha} c^\alpha}{\partial\varphi^k} U^k_\beta c^\beta + U^i_\mu U^\mu_{\alpha\beta} c^\beta c^\alpha + 2U^{ik}_{\alpha\beta} c^\beta c^\alpha \frac{\partial_r U}{\partial\varphi^k} = 0, \tag{4.39}$$

$$\frac{\partial_r U^\mu_{\alpha\beta} c^\beta c^\alpha}{\partial\varphi^i} U^i_\gamma c^\gamma + 2U^\mu_{\gamma\beta} c^\beta U^\gamma_{\sigma\alpha} c^\alpha c^\sigma + 2U^{\mu i}_{\alpha\beta\delta} c^\delta c^\beta c^\alpha \frac{\partial_r U}{\partial\varphi^i} = 0, \tag{4.40}$$

$$\begin{aligned}
&\frac{\partial_r U^{ik}_{\alpha\beta} c^\beta c^\alpha}{\partial\varphi^j} U^j_\mu c^\mu + 2U^{ik}_{\gamma\beta} c^\beta U^\gamma_{\mu\nu} c^\nu c^\mu \\
&+ 3U^{ikj}_{\alpha\beta\delta} c^\delta c^\beta c^\alpha \frac{\partial_r U}{\partial\varphi^j} \\
&- (-1)^{\epsilon_k}\left(\frac{\partial_r U^i_{\ \alpha} c^\alpha}{\partial\varphi^j} U^{jk}_{\beta\delta} c^\delta c^\beta \right. \\
&\left. + U^i_\mu U^{\mu k}_{\alpha\beta\delta} c^\delta c^\beta c^\alpha\right) \\
&- (-1)^{[\epsilon_i + (\epsilon_i + 1)(\epsilon_k + 1)]}\left(\frac{\partial_r U^k_{\ \alpha} c^\alpha}{\partial\varphi^j} U^{ji}_{\beta\delta} c^\delta c^\beta \right. \\
&\left. + U^k_\mu U^{\mu i}_{\alpha\beta\delta} c^\delta c^\beta c^\alpha\right) = 0,
\end{aligned} \tag{4.41}$$

and so on. In the above relations all c's can be differentiated away (at the expense of appearance of complicated sign factors).

We see now that relations (4.38) with identifications (4.36)–(4.37) are just the first-order relations of the gauge algebra: the Noether identities (2.11). Redenoting the other $U^{\cdots}_{\cdots}$ as follows:

$$U^\gamma_{\alpha\beta} = -\tfrac{1}{2} T^\gamma_{\alpha\beta}(-1)^{\epsilon_\alpha}, \tag{4.42}$$

$$U^{ik}_{\alpha\beta} = -\tfrac{1}{4} E^{ik}_{\alpha\beta}(-1)^{[\epsilon_\alpha + \epsilon_k(\epsilon_i + 1)]}, \tag{4.43}$$

$$U^{\mu i}_{\alpha\beta\delta} = -\tfrac{1}{12} F^{\mu i}_{\alpha\beta\delta}(-1)^{[\epsilon_\beta + \epsilon_\alpha\epsilon_\delta + \epsilon_i\epsilon_\mu]}, \tag{4.44}$$

$$U^{ikj}_{\alpha\beta\delta} = -\tfrac{1}{36} D^{ikj}_{\alpha\beta\delta}(-1)^{[\epsilon_\beta + \epsilon_\alpha\epsilon_\delta + \epsilon_k + \epsilon_{ikj}]}, \tag{4.45}$$

$$\epsilon_{ikj} \equiv \epsilon_i\epsilon_k + \epsilon_k\epsilon_j + \epsilon_j\epsilon_i, \tag{4.46}$$

we find that relations (4.39) coincide with the commutation relations (3.19), relations (4.40) are just the Jacobi identities (3.25), and relations (4.41) are the fourth-order structure relations (3.26) of the gauge algebra. The symmetry properties (3.20)–(3.21) and (3.27)–(3.30) of the structure functions are fulfilled in consequence of the symmetry properties of the $U^{\cdots}_{\cdots}$.

In the same way all higher-order structure relations of the algebra can be obtained as the relations which the master equation imposes upon the coefficients $U^{\cdots}_{\cdots}$, and the coefficients $U^{\cdots}_{\cdots}$ can be identified with the structure functions of the gauge algebra.[10] Conservation of ghost number plays the role of a selection rule for the sets of indices which the structure functions can have.

Thus the derivation of the gauge algebra reduces to the proof of existence of the corresponding solution of the master equation.

V. EXISTENCE THEOREM

The existence of the structure functions will be proved by induction given only the zeroth-order structure function: the gauge action $\mathscr{S}(\varphi)$. Note that under the condition of ghost-number conservation the expansion of S in powers of the auxiliary fields c, c^*, φ^* is in fact the expansion in terms of c:

$$S(\varphi, c; \varphi^*, c^*) = S|_{c=0} + \sum_{n=1}^{\infty} V_{\alpha_n \cdots \alpha_1}(\varphi, \varphi^*, c^*) c^{\alpha_1} \cdots c^{\alpha_n}, \tag{5.1}$$

in which the coefficients are finite polynomials in φ^*, c^* [cf. (4.35)]. Therefore, it suffices to carry out the induction with respect to the number n in (5.1), i.e., with respect to the number of lower indices of the structure functions.

Note also that conservation of ghost number allows to rewrite the boundary condition (4.22) as

$$S|_{c=0} = \mathscr{S}(\varphi). \tag{5.2}$$

The second boundary condition (4.24) was introduced to make the solution proper and involved the gauge generators. However, according to Sec. II, the gauge action $\mathscr{S}(\varphi)$ is the only initially given quantity, while the generators are already the first-order structure functions, whose existence should be a part of the general theorem. Therefore, it should be possible to avoid the introduction of the generators in the boundary conditions. Indeed, suffice it to require that

$$\operatorname{rank}_\pm \left.\frac{\partial_l \partial_r S}{\partial\varphi^*_i \, \partial c^\alpha}\right|_{\substack{c=0 \\ \varphi=\varphi_0}} = m_\pm . \tag{5.3}$$

The content of the present work can be now packed into the following theorem.

Theorem: Let $\mathscr{S}(\varphi)$ be the gauge action satisfying Postulates 1 and 2 of Sec. II. Then a bosonic solution of the master equation (4.33), satisfying the boundary conditions (5.2) and (5.3), exists as expansion (5.1) with the conserved ghost number equal to zero. The coefficients of this expansion are finite polynomials in φ^*, c^* and are regular (infinitely differentiable) functions of φ in some neighborhood of the stationary point φ_0. The solution is proper.

Proof: First of all we shall prove that if the above solution exists, then it is proper. Indeed, as a consequence of ghost-number conservation and the boundary condition (5.2) the following values of fields and antifields correspond to a stationary point of S:

$$\varphi^i = \varphi^i_0, \quad c^\alpha = 0, \quad \varphi^*_i = 0, \quad c^*_\alpha = 0. \tag{5.4}$$

As a consequence of the same reasons, the Hessian of S at this stationary point has only the following nonvanishing elements:

$$\left.\frac{\partial_l \partial_r S}{\partial\varphi^i \, \partial\varphi^k}\right|_{\substack{c=0 \\ \varphi=\varphi_0}} = \left.\frac{\partial_l \partial_r \mathscr{S}(\varphi)}{\partial\varphi^i \, \partial\varphi^k}\right|_{\varphi=\varphi_0}, \qquad \left.\frac{\partial_l \partial_r S}{\partial\varphi^*_i \, \partial c^\alpha}\right|_{\substack{c=0 \\ \varphi=\varphi_0}}, \quad \left.\frac{\partial_l \partial_r S}{\partial c^\alpha \, \partial\varphi^*_i}\right|_{\substack{c=0 \\ \varphi=\varphi_0}}. \tag{5.5}$$

The rank of the Hessian at the stationary point is, therefore,

$$r = \operatorname{rank} \left.\frac{\partial_l \partial_r \mathscr{S}(\varphi)}{\partial\varphi^i \, \partial\varphi^k}\right|_{\varphi=\varphi_0} + 2 \operatorname{rank} \left.\frac{\partial_l \partial_r S}{\partial\varphi^*_i \, \partial c^\alpha}\right|_{\substack{c=0 \\ \varphi=\varphi_0}}, \tag{5.6}$$

which makes

$$r=(n-m)+2m=n+m \tag{5.7}$$

as a consequence of Postulate 2 [Eq. (2.13)] and the boundary condition (5.3). But the number of fields (φ^i and c^α) is also $n+m$. Therefore, equality (4.20) holds, and the solution is proper.

Let us now formulate the requirements of the theorem to the coefficients of expansion (5.1). First of all the coefficients $V_{\alpha_n\cdots\alpha_1}$ must have the statistics and the symmetry of the monomial $c^{\alpha_1}\cdots c^{\alpha_n}$:

$$\epsilon(V_{\alpha_n\cdots\alpha_1})=\sum_{i=1}^{n}(\epsilon_{\alpha_i}+1), \tag{5.8}$$

$$V_{\alpha_n\cdots\alpha_1}=(V^{\text{sym}})_{\alpha_n\cdots\alpha_1}. \tag{5.9}$$

[For any quantity $X_{\alpha_n\cdots\alpha_1}$ this symmetry is defined as

$$(X^{\text{sym}})_{\alpha_n\cdots\alpha_1}=\eta^{\beta_n\cdots\beta_1}_{\alpha_n\cdots\alpha_1}X_{\beta_n\cdots\beta_1},$$

$$n!\ \ \eta^{\beta_n\cdots\beta_1}_{\alpha_n\cdots\alpha_1}\equiv\frac{\partial_r}{\partial c^{\alpha_1}}\cdots\frac{\partial_r}{\partial c^{\alpha_n}}(c^{\beta_1}\cdots c^{\beta_n}).\Big]$$

Further, the boundary condition (5.2) fixes the coefficient at $c=0$, and the boundary condition (5.3) restricts the coefficient with $n=1$:

$$\text{rank}_{\pm}\left.\frac{\partial_l V_\alpha}{\partial\varphi^*_i}\right|_{\varphi=\varphi_0}=m_\pm. \tag{5.10}$$

The coefficients $V_{\alpha_n\cdots\alpha_1}(\varphi,\varphi^*,c^*)$ must be finite polynomials in φ^* and c^*, satisfying the requirement

$$\text{gh}(V_{\alpha_n\cdots\alpha_1})=-n. \tag{5.11}$$

They must also be regular functions of φ in some neighborhood of φ_0. Finally, under conditions (5.2), (5.8), and (5.9) the master equation (4.33) is equivalent to the following sequence of relations:

$$\Omega V_{\alpha_1}=0, \tag{5.12}$$

$$\Omega V_{\alpha_n\cdots\alpha_1}=(B^{\text{sym}})_{\alpha_n\cdots\alpha_1},\quad n\geqslant 2. \tag{5.13}$$

Here Ω is the operator

$$\Omega=\frac{\partial_r\mathscr{S}}{\partial\varphi^i}\frac{\partial_l}{\partial\varphi^*_i}+V_\alpha\frac{\partial_l}{\partial c^*_\alpha}, \tag{5.14}$$

and

$$\begin{aligned}B_{\alpha_n\cdots\alpha_1}=&-\sum_{k=1}^{n-1}(-1)^{\epsilon_{nk}}\\&\times\Big[\tfrac{1}{2}(V_{\alpha_n\cdots\alpha_{k+1}},V_{\alpha_k\cdots\alpha_1})\\&+(n-k+1)V_{\alpha_n\cdots\alpha_{k+1}\alpha}\frac{\partial_l V_{\alpha_k\cdots\alpha_1}}{\partial c^*_\alpha}\Big],\end{aligned} \tag{5.15}$$

$$\epsilon_{nk}=n-k+\sum_{i=k+1}^{n}\epsilon_{\alpha_i}.$$

The antibrackets in (5.15) concern the dependence of $V_{\dots}$ on φ,φ^*.

The proof of existence will consist of three steps. We shall prove the following.

(1) There exists V_{α_1} with all the required properties.

(2) If for all $1\leqslant n\leqslant N-1$ there exist functions $V_{\alpha_n\cdots\alpha_1}$, satisfying the requirements of the theorem, then the quantity $(B^{\text{sym}})_{\alpha_N\cdots\alpha_1}$ is constructed of these functions only and satisfies the identity

$$\Omega(B^{\text{sym}})_{\alpha_N\cdots\alpha_1}=0. \tag{5.16}$$

(3) Identity (5.16) is necessary and sufficient for the existence of the Nth-order function $V_{\alpha_N\cdots\alpha_1}$ with all the required properties, if the lower-order functions $V_{\alpha_n\cdots\alpha_1}$, $1\leqslant n\leqslant N-1$ possess these properties.

Let us turn to the proof.

(1) As shown in Sec. II, Postulate 2 is equivalent to 2‴. Thus in some neighborhood of φ_0 there exist regular functions $R^i_\alpha(\varphi)$ satisfying relations (2.11) and (2.12). Define

$$V_{\alpha_1}=\varphi^*_iR^i_{\alpha_1}(\varphi). \tag{5.17}$$

It is easy to verify that this V_{α_1} possesses all the required properties. In particular, (5.12) follows from (2.11), and (5.10) follows from (2.12).

(2) The first-order function V_{α_1}, satisfying requirement (5.11), cannot depend on c^*:

$$\frac{\partial_l V_{\alpha_1}}{\partial c^*_\beta}=0. \tag{5.18}$$

[This is of course fulfilled in the solution (5.17).] Therefore, the term containing (5.18) in the sum (5.15) vanishes. As a result the quantity $B_{\alpha_N\cdots\alpha_1}$ contains only $V_{\alpha_n\cdots\alpha_1}$ with $1\leqslant n\leqslant N-1$.

Given $V_{\alpha_n\cdots\alpha_1}$ for $1\leqslant n\leqslant N-1$, we may construct a partial sum of the series (5.1):

$$S_{N-1}=\mathscr{S}(\varphi)+\sum_{n=1}^{N-1}V_{\alpha_n\cdots\alpha_1}c^{\alpha_1}\cdots c^{\alpha_n},\quad N\geqslant 2. \tag{5.19}$$

By direct computation we find

$$\begin{aligned}-\tfrac{1}{2}(S_{N-1},S_{N-1})=&-\Omega V_{\alpha_1}c^{\alpha_1}\\&+\sum_{n=2}^{N-1}[B_{\alpha_n\cdots\alpha_1}-\Omega V_{\alpha_n\cdots\alpha_1}]c^{\alpha_1}\cdots c^{\alpha_n}\\&+\sum_{n=N}^{2(N-1)}B_{\alpha_n\cdots\alpha_1}c^{\alpha_1}\cdots c^{\alpha_n}.\end{aligned} \tag{5.20}$$

Since, by assumption, the functions $V_{\dots}$, entering the partial sum (5.19), satisfy relations (5.12)–(5.13), we have

$$-\tfrac{1}{2}(S_{N-1},S_{N-1})=B_{\alpha_N\cdots\alpha_1}c^{\alpha_1}\cdots c^{\alpha_N}+o(c^{N+1}). \tag{5.21}$$

Let us now use the cyclic identity (4.9),

$$((S_{N-1},S_{N-1}),S_{N-1})\equiv 0. \tag{5.22}$$

Note that by assumption,

$$\text{gh}(S_{N-1})=0. \tag{5.23}$$

Therefore,

$$\frac{\partial_l S_{N-1}}{\partial\varphi^*_i}=o(c),\quad\frac{\partial_l S_{N-1}}{\partial c^*_\alpha}=o(c^2). \tag{5.24}$$

Inserting (5.21) and (5.19) into (5.22) and using (5.24), we obtain

$$\Omega B_{\alpha_N\cdots\alpha_1}c^{\alpha_1}\cdots c^{\alpha_N}+o(c^{N+1})=0. \tag{5.25}$$

Hence identity (5.16).

(3) The operator Ω, defined in (5.14), contains the first-order function V_α whose existence is proved already. Using equation (5.12) for this function, we find that the operator Ω is nilpotent

$$\Omega^2 = 0. \tag{5.26}$$

Since $(B^{\rm sym})_{\alpha_N \cdots \alpha_1}$ contains only $V_{\alpha_n \cdots \alpha_1}$ with $1 \leqslant n \leqslant N-1$, relation (5.13) becomes a linear inhomogeneous equation for the Nth-order function $V_{\alpha_N \cdots \alpha_1}$. Applying the operator Ω to this equation and using (5.26) we prove that identity (5.16) is necessary for the existence of a solution.

To prove that identity (5.16) is sufficient for the existence of a solution, consider a function $B(\varphi,\varphi^*,c^*)$, which is regular in some neighborhood of the point

$$\varphi = \varphi_0, \quad \varphi^* = 0, \quad c^* = 0, \tag{5.27}$$

satisfies the equation

$$\Omega B(\varphi,\varphi^*,c^*) = 0, \tag{5.28}$$

and vanishes on the surface

$$\varphi \in \Sigma, \quad \varphi^* = 0, \quad c^* = 0, \tag{5.29}$$

where Σ is the stationary orbit (see Sec. II). We shall prove that any such $B(\varphi,\varphi^*,c^*)$ has the form

$$B(\varphi,\varphi^*,c^*) = \Omega V(\varphi,\varphi^*,c^*), \tag{5.30}$$

where $V(\varphi,\varphi^*,c^*)$ is a regular function, for which we shall obtain an explicit representation.

To solve Eq. (5.28), rewrite expression (5.14) for the operator Ω as

$$\Omega = \mathscr{S}_A \frac{\partial_l}{\partial \varphi^*_A} + \mathscr{S}_\alpha \frac{\partial_l}{\partial \varphi^*_\alpha} + V_\alpha \frac{\partial_l}{\partial C^*_\alpha}, \tag{5.31}$$

where $\mathscr{S}_A$ and $\mathscr{S}_\alpha$ are the subsets from Postulate 2. This defines the division of the set of variables φ^*_i into two subsets

$$\varphi^*_i = (\varphi^*_A, \varphi^*_\alpha). \tag{5.32}$$

For the subset φ^*_α the following condition is true:

$$\operatorname{rank} \left. \frac{\partial_l V_\beta}{\partial \varphi^*_\alpha} \right|_{\varphi = \varphi_0} = m. \tag{5.33}$$

The proof uses relations (5.10)–(5.12) and exactly repeats the proof of condition (2.29) in Sec. II.

Due to condition (5.33) we may introduce V_α as independent variables instead of φ^*_α in Eq. (5.28). Simultaneously we shall introduce the reparametrization (2.14) of the field φ^i. Thus we make the following replacement of variables in Eq. (5.28):

$$\begin{aligned}(\varphi,\varphi^*,c^*) &= (\varphi'^A, \varphi'^\alpha, \varphi^*_A, \varphi^*_\alpha, c^*_\alpha)\\ &\to (J_A, \varphi'^\alpha, \varphi^*_A, V_\alpha, c^*_\alpha).\end{aligned} \tag{5.34}$$

Introducing the collective notation

$$G_i = (J_A, V_\alpha), \quad P_i = (\varphi^*_A, c^*_\alpha), \tag{5.35}$$

we find that operator Ω and Eq. (5.28) in the new variables take the form [the derivation uses relation (5.12)]

$$\Omega = G_i \frac{\partial_l}{\partial P_i}, \tag{5.36}$$

$$G_i \frac{\partial_l}{\partial P_i} \widetilde{B}(G,P,\varphi'^\alpha) = 0, \tag{5.37}$$

where

$$\widetilde{B}(G,P,\varphi'^\alpha) \equiv B(\varphi,\varphi^*,c^*). \tag{5.38}$$

Note, that

$$\epsilon(G_i) = \epsilon(P_i) + 1. \tag{5.39}$$

From the conditions imposed upon $B(\varphi,\varphi^*,c^*)$ it follows that the function $\widetilde{B}(G,P,\varphi'^\alpha)$ is regular near the surface

$$G = 0, \quad P = 0 \tag{5.40}$$

and vanishes on this surface. Thus $\widetilde{B}(G,P,\varphi'^\alpha)$ satisfies all the conditions of the lemma proved in Appendix A. Hence,

$$\begin{aligned}\widetilde{B}(G,P,\varphi'^\alpha) = \left(G_i \frac{\partial_l}{\partial P_i}\right)\left(P_k \frac{\partial_l}{\partial G_k}\right)\\ \times \int_0^1 \widetilde{B}(xG,xP,\varphi'^\alpha)\frac{dx}{x}.\end{aligned} \tag{5.41}$$

This gives us representation (5.30) with

$$V(\varphi,\varphi^*,c^*) = P_k \frac{\partial_l}{\partial G_k} \int_0^1 \widetilde{B}(xG,xP,\varphi'^\alpha)\frac{dx}{x}. \tag{5.42}$$

Now we must identify $B(\varphi,\varphi^*,c^*)$ with the quantity $(B^{\rm sym})_{\alpha_N \cdots \alpha_1}$ and $V(\varphi,\varphi^*,c^*)$ with the Nth-order function $V_{\alpha_N \cdots \alpha_1}$. For this purpose consider expression (5.15). If all $V_{\alpha_n \cdots \alpha_1}$, $1 \leqslant n \leqslant N-1$, entering this expression are regular functions of φ and finite polynomials in φ^*, c^* with the ghost numbers (5.11), then $B_{\alpha_N \cdots \alpha_1}$ is also a regular function of φ and a finite polynomial in φ^*, c^* with the ghost number

$$\operatorname{gh}(B_{\alpha_N \cdots \alpha_1}) = -N+1, \quad N \geqslant 2. \tag{5.43}$$

By virtue of (5.43), $B_{\alpha_N \cdots \alpha_1}$ vanishes when $\varphi^* = c^* = 0$. In particular, it vanishes on the surface (5.29). Finally, the symmetrized $B_{\alpha_N \cdots \alpha_1}$ satisifes Eq. (5.28). Therefore, representation (5.30) is valid for

$$B(\varphi,\varphi^*,c^*) = (B^{\rm sym})_{\alpha_N \cdots \alpha_1}. \tag{5.44}$$

Thus we proved that there exists a Nth-order function

$$V(\varphi,\varphi^*,c^*) = V_{\alpha_N \cdots \alpha_1}, \tag{5.45}$$

which satisfies Eq. (5.13). Let us prove that this function satisfies all the other requirements of the theorem. For this purpose consider representation (5.42) for quantities (5.45) and (5.44). From this representation we conclude that $V_{\alpha_N \cdots \alpha_1}$ has the symmetry of $(B^{\rm sym})_{\alpha_N \cdots \alpha_1}$. Hence, property (5.9). We also conclude that

$$\epsilon(V_{\alpha_N \cdots \alpha_1}) = \epsilon(B_{\alpha_N \cdots \alpha_1}) + 1, \tag{5.46}$$

because

$$\epsilon\left(P_k \frac{\partial}{\partial G_k}\right) = 1. \tag{5.47}$$

If the lower-order functions $V_{\alpha_n \cdots \alpha_1}$, $1 \leqslant n \leqslant N-1$, possess property (5.8), then from (5.15)

$$\epsilon(B_{\alpha_N \cdots \alpha_1}) = \sum_{i=1}^N (\epsilon_{\alpha_i} + 1) + 1. \tag{5.48}$$

Hence, property (5.8) for $n = N$.

To trace conservation of the ghost number, note that replacement (5.34) is linear in antifields, and replacement

$$G \to xG, \quad P \to xP \tag{5.49}$$

is equivalent to

$$\varphi^* \to x\varphi^*, \quad c^* \to xc^*, \quad J(\varphi) \to xJ(\varphi). \tag{5.50}$$

Therefore, if $B(\varphi,\varphi^*,c^*)$ is a finite polynomial in φ^*,c^* with a conserved ghost number, then the integral in (5.42) is also a finite polynomial in φ^*,c^* with the same ghost number. Noting that operator $P_k\, \partial/\partial G_k$ preserves polynomiality in φ^*,c^*, and

$$\mathrm{gh}\left(P_k \frac{\partial}{\partial G_k}\right) = -1, \tag{5.51}$$

we conclude from (5.42) that

$$\mathrm{gh}(V_{\alpha_N\cdots\alpha_1}) = \mathrm{gh}(B_{\alpha_N\cdots\alpha_1}) - 1. \tag{5.52}$$

Together with (5.43) this proves property (5.11) of the Nth-order function.

Finally, from representation (5.42) we conclude that $V_{\alpha_N\cdots\alpha_1}$ is a regular function of φ since $B_{\alpha_N\cdots\alpha_1}$, constructed of the lower-order functions, possesses this property. This completes the proof of the theorem.

VI. ON TRANSFORMATIONS OF THE BASIS OF GAUGE ALGEBRA

A solution of the master equation, satisfying the requirements of the above theorem, is not unique. Indeed, given the functions $V_{\alpha_n\cdots\alpha_1}$ of the first $N-1$ orders, we have the linear inhomogeneous equation (5.13) for the Nth-order function. Let $V_{\alpha_N\cdots\alpha_1}$ and $\overline{V}_{\alpha_N\cdots\alpha_1}$ be two solutions of this equation. Then their difference satisfies the corresponding homogeneous equation:

$$\Omega\,(\overline{V}_{\alpha_N\cdots\alpha_1} - V_{\alpha_N\cdots\alpha_1}) = 0. \tag{6.1}$$

This is just Eq. (5.28). Moreover, the difference $\overline{V}_{\cdots} - V_{\cdots}$ satisfies all the conditions imposed on $B(\varphi,\varphi^*,c^*)$ in (5.28). Hence, we find the general solution

$$\overline{V}_{\alpha_N\cdots\alpha_1} - V_{\alpha_N\cdots\alpha_1} = \Omega X_{\alpha_N\cdots\alpha_1}. \tag{6.2}$$

Here $X_{\alpha_N\cdots\alpha_1}$ is a regular function of φ and a finite polynomial in φ^*,c^*, possessing the properties

$$X_{\alpha_N\cdots\alpha_1} = (X^{\mathrm{sym}})_{\alpha_N\cdots\alpha_1}, \tag{6.3}$$

$$\epsilon(X_{\alpha_N\cdots\alpha_1}) = \sum_{i=1}^{N} (\epsilon_{\alpha_i} + 1) + 1, \tag{6.4}$$

$$\mathrm{gh}(X_{\alpha_N\cdots\alpha_1}) = -N - 1. \tag{6.5}$$

At $N=1$ one must also require (5.10) for both V_{α_1} and $\overline{V}_{\alpha_1}$. Otherwise $X_{\alpha_N\cdots\alpha_1}$ is arbitrary.

Thus, given the gauge action $\mathscr{S}(\varphi)$, the first-order function V_α is defined up to a transformation (6.2) at $N=1$. The second-order function $V_{\alpha_2\alpha_1}$ possesses the extra arbitrariness in a transformation (6.2) at $N=2$, and so on. The total arbitrariness of a solution is described by a set of functions

$$\{X_{\alpha_n\cdots\alpha_1}(\varphi,\varphi^*,c^*);\ 1 \leqslant n < \infty\} \tag{6.6}$$

possessing the above properties. Such a set is equivalent to one fermion function of all variables:

$$F(\varphi,c;\varphi^*,c^*) = \sum_{n=1}^{\infty} X_{\alpha_n\cdots\alpha_1} c^{\alpha_1}\cdots c^{\alpha_n}, \tag{6.7}$$

which can be used as generator of a canonical transformation in the space of fields and antifields.[12,16] It can be shown that such a canonical transformation of $S(\varphi,c;\varphi^*,c^*)$ is equivalent to the sequence of transformations (6.2). Thus a solution of the master equation, possessing the required properties, is unique up to a smooth canonical transformation preserving the boundary conditions (5.2), (5.3), and the ghost number of S.[12]

The structure functions of gauge algebra are the coefficients of expansion of $V_{\cdots}(\varphi,\varphi^*,c^*)$ in powers of φ^*,c^*. A transformation (6.2) of $V_{\cdots}$ (or a canonical transformation of S) is equivalent to some transformation of the structure functions. This transformation may be regarded as *a change of the basis of gauge algebra*. In particular, transformation (6.2) at $N=1$ is equivalent to the following change of the basis of generators:

$$R^{\,i}_{\alpha}(\varphi) \equiv \frac{\partial_l V_\alpha}{\partial \varphi^*_i}, \tag{6.8}$$

$$\overline{R}^{\,i}_{\alpha}(\varphi) = R^{\,i}_{\beta}(\varphi)\Lambda^{\,\beta}_{\alpha}(\varphi) + \frac{\partial_r \mathscr{S}(\varphi)}{\partial \varphi^n} K^{\,ni}_{\alpha}(\varphi). \tag{6.9}$$

Here $\Lambda^{\,\beta}_{\alpha}(\varphi)$ and $K^{\,ni}_{\alpha}(\varphi)$ are regular functions, such that

$$\mathrm{rank}\,\Lambda^{\,\beta}_{\alpha}|_{\varphi=\varphi_0} = m, \quad K^{\,ni}_{\alpha} = -K^{\,in}_{\alpha}(-1)^{\epsilon_n\epsilon_i}. \tag{6.10}$$

One can verify independently, that, given the gauge action $\mathscr{S}(\varphi)$, the generators $R^{\,i}_{\alpha}(\varphi)$ are defined by Eqs. (2.11)–(2.12) up to a transformation (6.9).

In specific bases some structure functions may vanish, and the structure relations of the algebra may look simpler. We shall say that the basis is of rank s, if the functions $V_{\alpha_n\cdots\alpha_1}$ of order $n > s$ vanish. [In view of condition (5.10), rank cannot be smaller than 1.]

The following simple theorem gives the possibility to establish the rank of a basis, given its functions $V_{\cdots}$ of the first s orders.

Theorem: A given set of functions of the first s orders

$$V_{\alpha_n\cdots\alpha_1}, \quad 1 \leqslant n \leqslant s \tag{6.11}$$

can be continued as the basis of rank s

$$V_{\alpha_n\cdots\alpha_1} = 0, \quad n > s \tag{6.12}$$

if and only if the functions (6.11)–(6.12) satisfy s "strong" identities

$$(B^{\mathrm{sym}})_{\alpha_n\cdots\alpha_1} \equiv 0, \quad s+1 \leqslant n \leqslant 2s. \tag{6.13}$$

The proof is based on Eq. (5.20).

In Refs. 11 and 12 a proof has been given, that for any gauge theory there exists a basis of rank 1, i.e., a basis in which the gauge algebra is abelian. (See also Ref. 16.) Using boldface for quantities in the abelian basis, we have

$$\mathbf{S}(\varphi,c;\varphi^*,c^*) = \mathscr{S}(\varphi) + \varphi^*_i \mathbf{R}^i_\alpha(\varphi) c^\alpha, \tag{6.14}$$

and the master equation $(\mathbf{S},\mathbf{S}) = 0$ is exhausted by relations

$$\frac{\partial_r \mathscr{S}}{\partial \varphi^i} \mathbf{R}^i_\alpha = 0, \tag{6.15}$$

$$\frac{\partial_r \mathbf{R}^i_\alpha}{\partial \varphi^k} \mathbf{R}^k_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta} \frac{\partial_r \mathbf{R}^i_\beta}{\partial \varphi^k} \mathbf{R}^k_\alpha = 0. \tag{6.16}$$

The abelian generators $\mathbf{R}^i_\alpha$ are connected with generators in any other basis by a transformation (6.9), and all higher-order structure functions in the abelian basis vanish.

The above result of Refs. 11 and 12 makes the fact of the existence of a solution for S obvious. What is not obvious, is the existence of Nth-order structure functions when the lower-order functions are given *in an arbitrary basis*. Just this is proved in the present paper. In particular, we prove that for generators in *any* basis the commutation relations are of the form (3.19). In other words, the present existence theorem guarantees that the most general form of structure relations is obtained.

The significance of the general basis is connected with the covariant-quantization conjecture in field theory.[17] Additional requirements, existing in field theory, such as locality and Lorentz covariance, destroy the democracy of the bases. It is supposed, that for a local and covariant gauge action there exists a basis of the gauge algebra, in which all structure functions are local and covariant. This is the first part of the covariant-quantization conjecture. In the known examples the local and covariant basis exists and is generally nonabelian and open.[8,15,18,19] (On the other hand, examples are unknown, where the rank of the local and covariant basis would exceed 2.) Here a reservation is needed, however. In the present paper the gauge albegra is constructed in the irreducible basis [condition (2.12) or (5.10)]. There are, however, field-theoretic examples, in which a local, covariant, and irreducible basis does not exist. The local and covariant generators in these theories are linearly dependent. Therefore, the construction of gauge algebra should be generalized to include reducible bases. Such a generalization is done in Refs. 13 and 14. The covariant-quantization conjecture should be understood in the sense of this generalization.

The covariant-quantization conjecture concerns Feynman rules for gauge theories. These rules are described by the functional integral, containing the action $S(\Phi;\Phi^*)$ and a measure.[10] The measure ensures the independence of the functional integral of the choice of the basis for $S(\Phi;\Phi^*)$. The problem is, however, that the measure cannot be completly determined without an appeal to canonical quantization.[10] The conjecture, on which all methods of covariant quantization (Refs. 6, 9–14, 20, and 21) are based, is that if the local basis is used for the construction of $S(\Phi;\Phi^*)$, then the measure may be ignored.[17]

If we add the condition of locality to the postulates of Sec. II, then $\mathscr{S}(\varphi)$ will become the action of a dynamical system with first-class constraints. The Feynman rules for such systems, obtained by canonical quantization, are well known.[1–4,22,23] To prove the above conjecture, we must compare the phase-space and configuration-space descriptions. For many examples such a comparison has been carried out (Refs. 2,5,22,24,25). It is almost evident, that the algebra of first-class constraints[4] generates a local basis of the Lagrangian gauge algebra. However, in the general case the detailed correspondence between the two descriptions is not established.

APPENDIX A: LEMMA

Here we shall prove the following lemma.

Let G_i and P_i, $i=1,\dots,n$, be independent variables of the opposite statistics

$$\epsilon(G_i)=\epsilon(P_i)+1. \tag{A1}$$

Consider the equation

$$G_i\frac{\partial_l}{\partial P_i}B(G,P)=0. \tag{A2}$$

Lemma: Any solution of Eq. (A.2), which is regular near the point $G=P=0$ and vanishes at this point

$$B(0,0)=0, \tag{A3}$$

can be represented as

$$B(G,P)=\left(G_i\frac{\partial_l}{\partial P_i}\right)\left(P_k\frac{\partial_l}{\partial G_k}\right)\times\int_0^1 B(xG,xP)\frac{dx}{x}. \tag{A4}$$

Proof: Operating with $P_k\partial_l/\partial G_k$ on Eq. (A2) and using the identities

$$\left(G_i\frac{\partial_l}{\partial P_i}\right)^2=\left(P_i\frac{\partial_l}{\partial G_i}\right)^2=0, \tag{A5}$$

$$\left(G_i\frac{\partial_l}{\partial P_i}\right)\left(P_k\frac{\partial_l}{\partial G_k}\right)+\left(P_k\frac{\partial_l}{\partial G_k}\right)\left(G_i\frac{\partial_l}{\partial P_i}\right)=G_i\frac{\partial_l}{\partial G_i}+P_i\frac{\partial_l}{\partial P_i}, \tag{A6}$$

one obtains

$$\left(G_i\frac{\partial_l}{\partial G_i}+P_i\frac{\partial_l}{\partial P_i}\right)B(G,P)=\left(G_i\frac{\partial_l}{\partial P_i}\right)\left(P_k\frac{\partial_l}{\partial G_k}\right)B(G,P). \tag{A7}$$

Replace here

$$G_i\to xG_i,\quad P_i\to xP_i, \tag{A8}$$

where x is a numerical parameter. This gives

$$x\frac{d}{dx}B(xG,xP)=\left(G_i\frac{\partial_l}{\partial P_i}\right)\left(P_k\frac{\partial_l}{\partial G_k}\right)B(xG,xP). \tag{A9}$$

Thus one obtains

$$B(G,P)=B(x_0G,x_0P)+\left(G_i\frac{\partial_l}{\partial P_i}\right)\left(P_k\frac{\partial_l}{\partial G_k}\right)\times\int_{x_0}^1 B(xG,xP)\frac{dx}{x},\quad x_0>0. \tag{A10}$$

The existence of the limit $x_0\to 0$ is guaranteed by the regularity of $B(G,P)$ and by condition (A3). Hence, representation (A4).

APPENDIX B: CLASSIFICATION OF THEORIES WITH A DEGENERATE HESSIAN OF THE ACTION

Here we shall show that the postulates, adopted in the present paper, are the necessary conditions for the existence of loop expansion in quantum theory. For this purpose we shall consider theories which do not satisfy Postulate 2. As a preliminary we shall give one more (fifth already) equivalent formulation of Postulate 2.

Theorem: Postulate 2 is equivalent to the (local) existence of a regular reparametrization

$$\varphi^i \leftrightarrow (\xi^A, \theta^\mu), \tag{B1}$$

$$i = 1,...,n; \quad A = 1,...,(n-m); \quad \mu = 1,...,m;$$

$$\operatorname{rank} \frac{\partial(\xi,\theta)}{\partial\varphi^i}\bigg|_{\varphi=\varphi_0} = n, \tag{B2}$$

such that in the new variables the action does not depend on m field components

$$\frac{\partial_r \mathscr{S}}{\partial\theta^\mu} \equiv 0 \tag{B3}$$

and satisfies the relation

$$\operatorname{rank} \frac{\partial_l \partial_r \mathscr{S}}{\partial\xi^A \partial\xi^B}\bigg|_{\varphi=\varphi_0} = n - m \tag{B4}$$

with respect to the remaining $(n-m)$ components.

Proof: It is easy to derive Postulate 2 from the existence of reparametrization (B1)–(B4). Indeed, from (B3) we have

$$\frac{\partial_r \mathscr{S}(\varphi)}{\partial\varphi^i} \mathbf{R}^i_\mu(\varphi) \equiv 0, \tag{B5}$$

where

$$\mathbf{R}^i_\mu(\varphi) = \frac{\partial_r \varphi^i}{\partial\theta^\mu}. \tag{B6}$$

Relations

$$\operatorname{rank} \mathbf{R}^i_\mu |_{\varphi=\varphi_0} = m, \quad \epsilon(\mathbf{R}^i_\mu) = \epsilon_i + \epsilon_\mu \tag{B7}$$

follow from (B6) and (B2), and relation (2.13) follows from (B4) and (B3).

It is not easy to derive the existence of reparametrization (B1)–(B4) from Postulate 2. This derivation is the main result of Ref. 12. Note that expression (B6) defines the Noether generators in the abelian basis (6.14).

Let us now justify our postulates. The possibility to expand the action at a stationary point

$$\mathscr{S}(\varphi) = \mathscr{S}(\varphi_0) + \frac{1}{2} \frac{\partial_r \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^k}\bigg|_{\varphi_0} (\varphi - \varphi_0)^i (\varphi - \varphi_0)^k + \frac{1}{3!} \frac{\partial_r \partial_r \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^k \partial\varphi^m}\bigg|_{\varphi_0} \times (\varphi - \varphi_0)^i (\varphi - \varphi_0)^k (\varphi - \varphi_0)^m + \cdots \tag{B8}$$

is the necessary condition for the applicability of the standard loop technique. Hence, Postulate 1.

Consider next the Hessian of $\mathscr{S}$ at the stationary point. Generally

$$\operatorname{rank} \frac{\partial_l \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^k}\bigg|_{\varphi_0} = n - m, \tag{B9}$$

$$0 \leqslant m \leqslant n. \tag{B10}$$

If $m \neq 0$, the Hessian has zero-eigenvalue eigenvectors

$$\frac{\partial_l \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^j}\bigg|_{\varphi_0} \Lambda^j = 0, \quad \Lambda^j \not\equiv 0. \tag{B11}$$

Let Λ^j_α, $\alpha = 1,...,$ be such a set of zero-eigenvalue eigenvectors, that any Λ^j from (B11) is a linear combination of Λ^j_α. Let

$$r = \operatorname{rank} \Lambda^j_\alpha. \tag{B12}$$

Generally,

$$0 \leqslant r \leqslant m. \tag{B13}$$

The case $r < m$ takes place if and only if there exist nonvanishing zero-eigenvalue eigenvectors Λ^j which consist only of noninvertible components.

There are only three different types of theories according to the following classification. Consider a regular reparametrization in a neighborhood of φ_0

$$\varphi^i = \varphi^i_0 + \frac{\partial_r \varphi^i}{\partial\bar\varphi^k}\bigg|_0 (\bar\varphi - \bar\varphi_0)^k + \frac{1}{2} \frac{\partial_r \partial_r \varphi^i}{\partial\bar\varphi^k \partial\bar\varphi^n}\bigg|_0 (\bar\varphi - \bar\varphi_0)^k (\bar\varphi - \bar\varphi_0)^n + \cdots. \tag{B14}$$

If $m \neq 0$, we may try to choose the linear term of this replacement so as to eliminate m variables $(\bar\varphi - \bar\varphi_0)^\alpha$ from the quadratic term of (B8). This is equivalent to requiring

$$\frac{\partial_l \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^j}\bigg|_{\varphi_0} \frac{\partial_r \varphi^j}{\partial\bar\varphi^\alpha}\bigg|_0 = 0, \qquad \operatorname{rank} \frac{\partial_r \varphi^j}{\partial\bar\varphi^\alpha}\bigg|_0 = m. \tag{B15}$$

From (B11)–(B13) we conclude that the elimination of m variables from the quadratic term of the action is possible if and only if $r = \mathrm{m}$. If it is impossible to eliminate m variables from the quadratic term of the action, we have a theory of *the first type*.

If m variables $(\bar\varphi - \bar\varphi_0)^\alpha$ do not enter the quadratic term of the action, we may try to eliminate them also from the cubic term of the action by the choice of the quadratic term of replacement (B14). This can be done if and only if the three-point vertex in (B8) satisfies the constraint

$$\frac{\partial_r \partial_r \partial_r \mathscr{S}}{\partial\varphi^i \partial\varphi^k \partial\varphi^m}\bigg|_{\varphi_0} \Lambda^i_\alpha \Lambda^k_\beta \Lambda^m_\gamma (-1)^{\epsilon_\alpha(\epsilon_k + \epsilon_m) + \epsilon_\beta \epsilon_m} = 0, \tag{B16}$$

$$\epsilon_\alpha \equiv \epsilon(\bar\varphi^\alpha), \quad \operatorname{rank} \Lambda^i_\alpha = m.$$

If m variables $(\bar\varphi - \bar\varphi_0)^\alpha$ do not enter the quadratic and cubic terms of the action, then we may try to eliminate them also from the quartic term. This requires the fulfillment of a certain constraint for the four-point vertex, and so on. If it is possible to eliminate m variables from the first $N-1$ terms of the action (B8) and impossible to eliminate them from the Nth term ($N \geqslant 3$), we have a theory of *the second type*.

There remains the case when m variables can be completely eliminated from the action by a regular reparametrization. In this case we have a theory of *the third type*. The case $m = 0$ is the particular case of this type.

Theories of the first type are most unusual. The space of solutions of their linearized equations has a nontrivial zero-dimensional subspace. Theories of the second type are essentially nonlinear at the stationary point. Theories of the third type are normal theories with $(n-m)$ field components and a nondegenerate Hessian.

According to the theorem, formulated at the beginning of this Appendix, Postulate 2 is equivalent to the condition that the theory belongs to the third type.

Because of the degeneration of the Hessian the standard loop technique is evidently inapplicable to theories of the first and second types. This leaves us with Postulate 2.

The following examples illustrate the above classification:

$$(1)\quad \mathscr{S}(\varphi) = a\varphi^2, \quad \epsilon(\varphi) = 0.$$

Here a is a boson parameter. If a is invertible, this is the third-type theory with $n - m = 1$. If $a = 0$, this is the third-type theory with $n - m = 0$. If a is noninvertible and nonzero, this is the theory of the first type.

$$(2)\ \mathscr{S}(\varphi) = \varphi^4, \quad \epsilon(\varphi) = 0.$$

This is the theory of the second type.

$$(3)\ \mathscr{S}(\bar{\psi},\psi,\varphi) = \bar{\psi}\psi + \bar{\psi}\psi\varphi^2 + \bar{\psi}\eta + \bar{\eta}\psi,$$
$$\epsilon(\bar{\psi}) = \epsilon(\psi) = 1, \quad \epsilon(\varphi) = 0.$$

Here $\bar{\psi},\psi$, and φ are independent field variables, and $\bar{\eta},\eta$ are fermion parameters (sources). If $\bar{\eta}\eta = 0$, this is the normal (third-type) gauge theory. If $\bar{\eta}\eta \neq 0$, this is the theory of the first type.

$$(4)\ \mathscr{S}(\psi_i) = \psi_1\psi_2\psi_3\psi_4,$$
$$\epsilon(\psi_i) = 1, \quad i = 1,2,3,4.$$

Depending on the choice of the stationary point, this theory is either essentially nonlinear (the second type) or belongs to the first type.

[1] E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975).
[2] E. S. Fradkin and G. A. Vilkovisky, CERN preprint TH. 2332-CERN (1977).
[3] I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).
[4] E. S. Fradkin and T. E. Fradkina, Phys. Lett. B **72**, 343 (1978).
[5] E. S. Fradkin and M. A. Vasiliev, Phys. Lett. B **72**, 70 (1977).
[6] R. E. Kallosh, Pis'ma Zh. Eksp. Teor. Fiz. **26**, 575 (1977) [JETP Lett. **26**, 427 (1977)]; Nucl. Phys. B **141**, 141 (1978).
[7] G. Sterman, P. K. Townsend, and P. van Nieuwenhuizen, Phys. Rev. D **17**, 1501 (1978).
[8] D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D **14**, 912 (1976).
[9] B. de Wit and J. W. van Holten, Phys. Lett. B **79**, 389 (1979).
[10] I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **102**, 27 (1981).
[11] I. A. Batalin and G. A. Vilkovisky, in *Quantum Gravity*, edited by M. A. Markov (Plenum, New York, 1983).
[12] I. A. Batalin and G. A. Vilkovisky, Nucl. Phys. B **234**, 106 (1984).
[13] I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **120**, 166 (1983).
[14] I. A. Batalin and G. A. Vilkovisky, Phys. Rev. D **28**, 2567 (1983); Phys. Rev. D **30**, 508 (1984).
[15] B. S. DeWitt, Phys. Rev. **162**, 1195 (1967).
[16] B. L. Voronov and I. V. Tyutin, Teor. Mat. Fiz. **50**, 333 (1982).
[17] G. A. Vilkovisky, in *Quantum Theory of Gravity, Essays in Honor of the 60th Birthday of Bryce S. DeWitt*, edited by S. M. Christensen (Hilger, Bristol, 1984).
[18] L. D. Faddeev and A. A. Slavnov, *Gauge Fields: Introduction to Quantum Theory* (Addison-Wesley, Reading, MA, 1980).
[19] P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).
[20] L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967); Usp. Fiz. Nauk **111**, 427 (1973) [Sov. Phys. Usp. **16**, 777 (1974)].
[21] C. Becchi, A. Rouet, and A. Stora, Commun. Math. Phys. **42**, 127 (1975).
[22] L. D. Faddeev, Teor. Mat. Fiz. **1**, 3 (1969) [Theor. Math. Phys. **1**, 1 (1970)].
[23] I. A. Batalin and E. S. Fradkin, Phys. Lett. B **122**, 157 (1983).
[24] E. S. Fradkin and G. A. Vilkovisky, Lett. Nuovo Cimento **13**, 187 (1975).
[25] T. E. Fradkina, "Quantization of Fermi–Bose systems with general constraints," preprint N 5, P. N. Lebedev Physical Institute, Moscow (1979).

Effective action on the hyperbolic plane in a constant external field

A. Comtet[a)]
Division de Physique Theorique,[b)] *Institut de Physique Nucléaire, Université Paris-Sud, B. P. No. 1, F-91406 Orsay Cedex, France*

P. J. Houston
Division de Physique Theorique,[b)] *Institut de Physique Nucléaire, Université Paris-Sud, B. P. No. 1, F-91406 Orsay, Cedex, France and School of Mathematics, 39 Trinity College, Dublin 2, Ireland*[c), d)]

(Received 12 April 1984; accepted for publication 1 June 1984)

Massive scalar and spin $\frac{1}{2}$ fields are considered on a two-dimensional space-time with constant background curvature and in the presence of an external constant field. For each case, the Euler–Heisenberg effective action and the pair creation rate and their dependence on both the curvature and external field is determined exactly.

I. INTRODUCTION

One approach to the problem of gravitational quantum field theory that has developed consists of treating the gravitational field as a background external classical field in a quadratic quantum Lagrangian giving the quantum fluctuations about the background field.[1,2] The object is then to determine the dependence of the ground state energy on this external field. Thus the essential problem that is being considered, which has a history going back to the work of Casimir,[3] Euler and Heisenberg,[4] and Schwinger,[5] is the evaluation of the effective action and pair creation rate in an external field. The solution to this problem, at least at a formal level, is well known, as the quantum Lagrangian is quadratic. That is, the effective action, etc. is given by the logarithmic determinant of a second-order differential operator depending on the external field.[1,6,7] In spite of this formal solution, there are but a few examples for which the effective action can be determined explicitly. In this paper we consider a two-dimensional space-time on which there is a background metric defined and in addition we have an external field acting. The spatial fluctuations of the background fields are kept minimal by making those fields constant and the space-time is such that the Wick-rotation trick may be employed. We determine the effective action and pair creation rate and their exact dependence on two parameters, one for the metric and the other for the external field.

Thus we consider here the problem of a massive scalar-Klein–Gordon field and of a massive Fermi–Dirac field which are on two-dimensional space-time with an external background metric whose curvature is constant and the fields are also in the presence of a constant external electromagnetic field (we will say what we mean by the term constant later on). In each case we determine, exactly, the expressions for the Euler–Heisenberg effective action and for the rate of pair creation.

The background space-time metric we have can be consistently Wick-rotated[8,9] and in fact is the Wick rotation of the Riemannian metric of the Poincaré half-plane[10] (the hyperbolic plane) which has constant negative curvature.[11] On this Riemannian space, for the Klein–Gordon and Dirac fields we need only consider elliptic differential operators to compute the effective action, i.e., the logarithmic determinant. We do this by obtaining the discrete and continuous eigenspectra of these elliptic operators exactly and we reduce the expression for the trace of the heat kernel and the zeta function to a simple contour integral. Taking the Wick rotation of this log determinant we obtain from the real and imaginary parts the effective action and the rate of pair creation, respectively. The result of these computations are in Eqs. (4.6) and (4.9).

The paper has been organized as follows: In Sec. II we consider the Klein–Gordon and Dirac fields on the Poincaré half-plane in the presence of a constant electromagnetic field. In Sec. III we compute the effective action on this Riemannian space. In Sec. IV we carry out the Wick rotation to the two-dimensional space-time. Finally, the Appendix contains details pertaining to Sec. III.

II. KLEIN–GORDON AND DIRAC OPERATORS ON THE POINCARÉ HALF-PLANE

For any two-dimensional Riemannian manifold M we may choose isothermal (or conformal) coordinates (x_1,x_2) so that the metric takes the form

$$ds^2 = \lambda(x_1,x_2)^2(dx_1^2 + dx_2^2). \tag{2.1}$$

The scalar curvature on M is given by

$$S = -\lambda^{-2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)\ln\lambda^2. \tag{2.2}$$

For the Poincaré half-plane the metric is given as

$$\lambda(x_1,x_2) = a/x_1, \quad x_1 > 0, \tag{2.3}$$

in which case the curvature S is a negative constant

$$S = -2/a^2. \tag{2.4}$$

An electromagnetic field is obtained on M by specifying a one-form $A = A_i(x_1,x_2)dx^i$. The field strength is then given by

$$B = dA = \left(\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2}\right)dx_1 \wedge dx_2. \tag{2.5}$$

We are interested in a field strength which is constant on M and by this we mean that B is given, up to a constant of proportionality, by the volume form on M. Thus, in terms of

---

[a)] Present address: Center for Nonlinear Studies, Los Alamas National Laboratory, Los Alamos, New Mexico 87545.
[b)] Laboratoire associe au CNRS.
[c)] Irish Department of Education postdoctoral fellow.
[d)] Present address.

 0022-2488/85/010185-07$02.50

isothermal coordinates B is constant on M if, for some constant k,

$$B = k\lambda\,(x_1,x_2)^2 dx_1 \wedge dx_2. \tag{2.6}$$

This field strength is obtained from the gauge field A, in the $A_1 = 0$ gauge, by

$$\frac{\partial A_2}{\partial x_1}(x_1,x_2) = k\lambda\,(x_1,x_2)^2. \tag{2.7}$$

For the Poincaré half-plane we take

$$A = A_2\, dx_2 = -(b/x_1)dx_2, \tag{2.8}$$

to get a constant field strength

$$B = (b/x_1^2)dx_1 \wedge dx_2 = (b/a^2)\lambda\,(x_1,x_2)^2\, dx_1 \wedge dx_2. \tag{2.9}$$

In terms of isothermal coordinates, the Klein–Gordon operator for a scalar particle of mass m in an external gauge field A is

$$\mathscr{K} = -\Delta_A + m^2, \tag{2.10}$$

where

$$\Delta_A = \lambda^{-2}(D_1^2 + D_2^2), \quad D_i = \frac{\partial}{\partial x_i} - iA_i. \tag{2.11}$$

Now on the Poincaré half-plane using Eqs. (2.3) and (2.8) the Klein–Gordon operator is given by Eq. (2.10) where

$$\Delta_A = \frac{x_1^2}{a^2}\left(\frac{\partial^2}{\partial x_1^2} + \left(\frac{\partial}{\partial x_2} + \frac{ib}{x_1}\right)^2\right). \tag{2.12}$$

For a spin $\frac{1}{2}$ particle of mass m in an external gauge field A we have the following operator (with respect to isothermal coordinates):

$$\Theta_1 = \mathscr{D}_A + m, \tag{2.13}$$

where

$$\mathscr{D}_A = \lambda^{-1}\gamma^i\left(D_i + \frac{1}{2}\frac{\partial}{\partial x_i}(\ln\lambda)\right). \tag{2.14}$$

Here, $\{\gamma^1,\gamma^2\}$ define a Clifford algebra

$$\gamma^i\gamma^j + \gamma^j\gamma^i = 2\delta_{ij}. \tag{2.15}$$

We will find it more convenient to consider, instead of the Dirac operator, the following operator:

$$\Theta_2 = -\mathscr{D}_A^2 + m^2. \tag{2.16}$$

Moreover, the operator $\mathscr{D}_A^2$ may be written

$$\mathscr{D}_A^2 = \left(\Delta_A - \frac{1}{4}S - \frac{1}{4}\lambda^{-2}\left(\frac{\partial}{\partial x_i}(\ln\lambda)\frac{\partial}{\partial x_j}(\ln\lambda)\right)\right) + \gamma^1\gamma^2\left(F - \epsilon_{ij}\frac{\partial}{\partial x_i}(\ln\lambda)D_j\right), \tag{2.17}$$

where

$$F = \epsilon_{ij}D_iD_j = -i\left(\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2}\right), \tag{2.18}$$

and $\epsilon_{ij} = -\epsilon_{ji}$, $\epsilon_{12} = 1$. For the Poincaré half-plane with A given by Eq. (2.8)

$$\mathscr{D}_A^2 = \frac{x_1^2}{a^2}\left(\frac{\partial^2}{\partial x_1^2} + \left(\frac{\partial}{\partial x_2} + \frac{ib}{x_1}\right)^2 + \frac{1}{4x_1^2} + \gamma^1\gamma^2\frac{1}{x_1}\frac{\partial}{\partial x_2}\right). \tag{2.19}$$

III. EVALUATION OF LOG DETERMINANTS

We now obtain the log determinant, on the Poincaré half-plane, of the Klein–Gordon operator and the Dirac operator for scalar and spinor particles of mass m in a constant external field. For the Dirac operator because of the charge conjugation symmetry of its log determinant it follows that

$$\ln\det\,\Theta_1 = \tfrac{1}{2}\ln\det\Theta_2. \tag{3.1}$$

Thus we need only consider the elliptic operators $\mathscr{K}$ and Θ_2 in the Klein–Gordon and Dirac cases, respectively, and the computation of the log determinant proceeds via the heat kernel and zeta function.[1,6,7]

A. The Klein–Gordon operator

The continuous and discrete eigenspectra of the operator $\mathscr{K}$, given by Eqs. (2.10) and (2.12), and denoted $\{w_\nu^{(c)}, \phi_{\nu,k}^{(c)}\}$ and $\{w_n^{(d)}, \phi_{n,k}^{(c)}\}$, respectively, are as follows: $\mathscr{K}\phi^{(\cdot)}_{\cdot} = w^{(\cdot)}_{\cdot}\phi$, where

$$\begin{aligned} w_\nu^{(c)} &= m^2 + (\tfrac{1}{4} + b^2 + \nu^2)/a^2, \quad 0 \leqslant \nu < \infty, \\ w_n^{(d)} &= m^2 + (\tfrac{1}{4} + b^2 - (n + \tfrac{1}{2} - |b|)^2)/a^2, \\ &\quad n = 0,1,\ldots, \quad 0 \leqslant n < |b| - \tfrac{1}{2}, \end{aligned} \tag{3.2}$$

and the eigenfunctions are

$$\begin{aligned} \phi_{\nu,k}^{(c)}(x_1,x_2) &= a^{-1}f(\nu,k,\pm|b|;\ x_1,x_2), \\ &\quad \pm kb > 0,\ 0 \leqslant \nu < \infty, \\ \phi_{n,k}^{(d)}(x_1,x_2) &= a^{-1}f_n(k,|b|;\ x_1,x_2), \quad kb < 0, \\ &\quad n = 0,1,\ldots, \quad 0 \leqslant n < |b| - \tfrac{1}{2}. \end{aligned} \tag{3.3}$$

Here the functions f and f_n are given in the Appendix. The eigenfunctions satisfy the following normalization conditions:

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_0^{\infty}\frac{a^2}{x_1^2}dx_1\,dx_2\,\phi_{\nu,k}^{(c)*}(x_1,x_2)\phi_{\nu',k'}^{(c)}(x_1,x_2) \\ &\quad = \delta(\nu - \nu')\delta(k - k'), \\ &\int_{-\infty}^{\infty}\int_0^{\infty}\frac{a^2}{x_1^2}dx_1\,dx_2\,\phi_{n,k}^{(d)}(x_1,x_2)\phi_{n',k'}^{(d)}(x_1,x_2) \\ &\quad = \delta_{n,n'}\delta(k - k'), \end{aligned} \tag{3.4}$$

where all other inner products are zero. The heat kernel is given by

$$\begin{aligned} &H_{\mathscr{K}}(s;x_1,x_2;x_1',x_2') \\ &\quad = \int_{kb<0}dk\sum_{n=0}^{|b|-\frac{1}{2}}e^{-sw_n^{(d)}}\phi_{n,k}^{(d)}(x_1',x_2')\phi_{n,k}^{(d)*}(x_1',x_2') \\ &\quad + \int_{-\infty}^{\infty}dk\int_0^{\infty}d\nu\,e^{-sw_\nu^{(c)}}\phi_{\nu,k}^{(c)}(x_1,x_2)\phi_{\nu,k}^{(c)*}(x_1,x_2). \end{aligned} \tag{3.5}$$

However, we cannot compute the trace of $H_{\mathscr{K}}$ directly, as the volume of the Poincaré half-plane is infinite. Because $H_{\mathscr{K}}(s;x_1,x_2;x_1,x_2)$ is independent of (x_1,x_2) we may proceed in the same way as for Euclidean space by factoring out the volume. Thus we have (see the Appendix)

$$y_{\mathscr{K}}(s) = H_{\mathscr{K}}(s;x_1,x_2;x_1,x_2)$$

$$= \frac{1}{2\pi a^2}\left\{\sum_{n=0}^{|b|-\frac{1}{2}}(|b| - (n+\tfrac{1}{2}))e^{-sw_n^{(d)}} + \frac{1}{\pi}\int_0^\infty d\nu\, \nu \operatorname{Im}\left(\Psi\left(\frac{1}{2} + i\nu - b\right) + \Psi\left(\frac{1}{2} + i\nu + b\right)\right)e^{-sw_\nu^{(c)}}\right\}, \tag{3.6}$$

and letting

$$V = \int_{-L_2}^{L_2}\int_{\epsilon_1}^{L_1}\frac{a^2}{x_1^2}dx_1\,dx_2$$

the trace of $H_{\mathscr{K}}$ may be written

$$Y_{\mathscr{K}}(s) = Vy_{\mathscr{K}}(s). \tag{3.7}$$

A less cumbersome expression than Eq. (3.6) and (3.7) for the trace of the heat kernel can be obtained by integrating over a suitable contour in the complex ν plane. Thus

$$Y_{\mathscr{K}}(s) = \int_C d\nu\, \rho_{\mathscr{K}}(\nu)e^{-sw(\nu)}, \tag{3.8}$$

where

$$\rho_{\mathscr{K}}(\nu) = (V/4\pi^2a^2i)\nu(\Psi(\tfrac{1}{2} + i\nu - b) + \Psi(\tfrac{1}{2} + i\nu + b)),$$
$$w(\nu) = m^2 + (\tfrac{1}{4} + b^2 + \nu^2)/a^2, \tag{3.9}$$

and the contour C is shown in Fig. 1(a). The zeta function for $\mathscr{K}$ is given by the Mellin transform of $Y_{\mathscr{K}}(s)$, i.e.,

$$\zeta_{\mathscr{K}}(z) = (\Gamma(z))^{-1}\int_0^\infty ds\, s^{z-1}Y_{\mathscr{K}}(s).$$

Thus

$$\zeta_{\mathscr{K}}(z) = \int_C d\nu\, \rho_{\mathscr{K}}(\nu)w(\nu)^{-z}. \tag{3.10}$$

However, this converges and defines an analytic function only for Re $z > 1$. To obtain ln det $\mathscr{K}$ we need to analytically continue Eq. (3.10) to the origin. This may be done by isolating and removing from the integrand in Eq. (3.10) the terms which diverge most as $\nu \to \pm\infty$. The integral in the expression for the zeta function

$$\zeta_{\mathscr{K}}(z) = \int_C d\nu\, \tilde{\rho}_{\mathscr{K}}(\nu)w(\nu)^{-z} + \frac{V}{4\pi}\frac{w(o)^{1-z}}{z-1}, \tag{3.11}$$

where C is as in Fig. 1(b) and

$$\tilde{\rho}_{\mathscr{K}}(\nu) = \rho_{\mathscr{K}}(\nu) - (V/2\pi^2a^2i)V\ln\nu, \tag{3.12}$$

converges for Re $z > -\frac{1}{2}$ and $\zeta_{\mathscr{K}}$ is now analytic in a neighborhood of the origin [see equation (A9) for the asymptotic expansion for Ψ]. Computing $-\zeta'_{\mathscr{K}}(0)$ gives [C the same as for Eq. (3.11)]

$$\ln\det\mathscr{K} = \int_C d\nu\, \tilde{\rho}_{\mathscr{K}}(\nu)\ln w(\nu) + \frac{V}{4\pi}w(0)(1 - \ln w(0)). \tag{3.13}$$

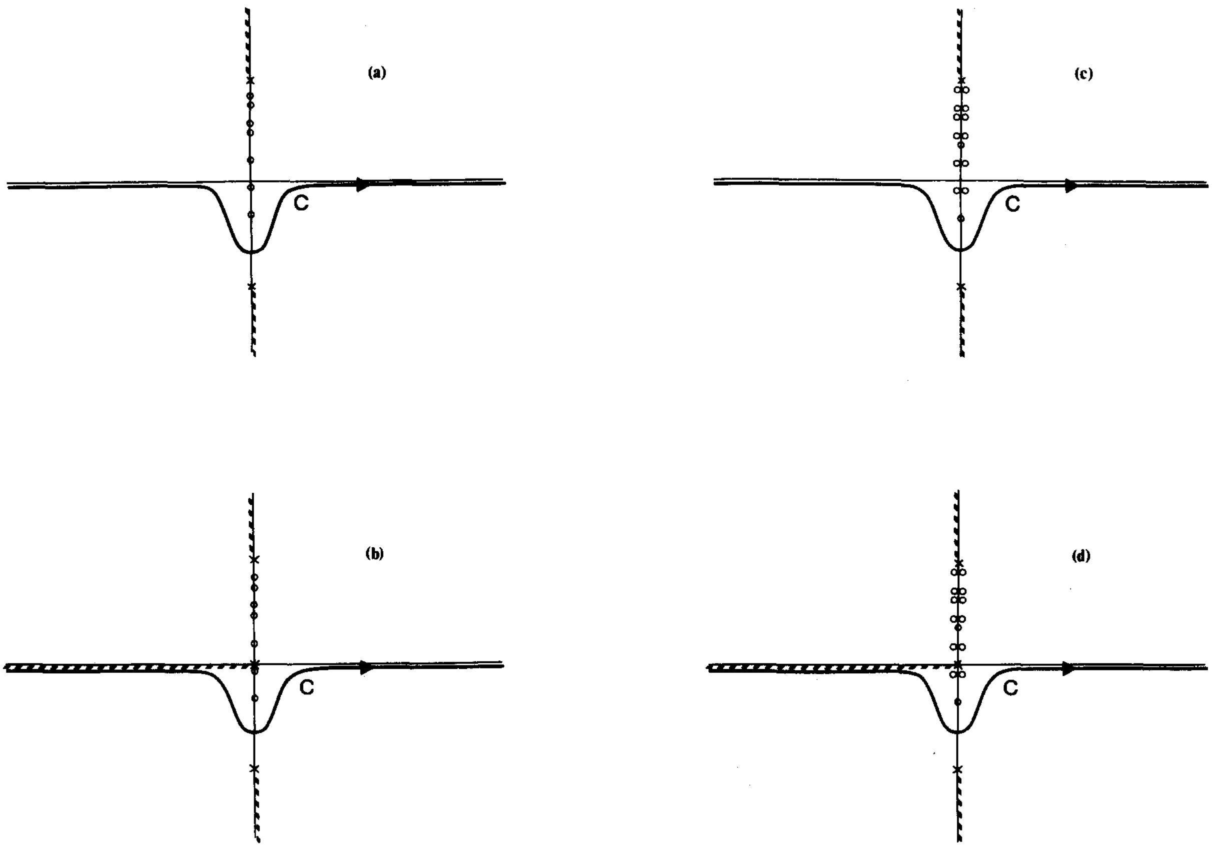


FIG. 1. The poles (denoted by O) in (a) and (b) are at $(\pm b + n + \frac{1}{2})i$ and in (c) and (d) they are at $(\pm b + n)i$, $(\pm b + n + 1)i (n = 0,1,\ldots)$. There are branch points (denoted by $\times$) in (a) and (b) at $\pm i\sqrt{(\frac{1}{4} + m^2a^2 + b^2)}$ and in (c) and (d) at $\pm i\sqrt{(m^2a^2 + b^2)}$. The only other branch point is at the origin in (b) and (d). (- - - - denotes a cut.)

B. The Dirac operator

As a result of the comments at the beginning of this section we consider the operator Θ_2 given by Eqs. (2.16) and (2.19) in order to compute the log determinant of the Dirac operator Θ_1. The continuous and discrete eigenspectra of Θ_2 denoted $\{\chi_\nu^{(c)},\psi_{\pm 1,\nu,k}^{(c)}\}$ and $\{\chi_0^{(d)},\psi_{0,k}^{(d)};\chi_b^{(d)},\psi_{\pm 1,n,k}^{(d)}\}_{n>0}$, respectively, are as follows: $\Theta_2\psi_{\cdot,k}^{(\cdot)} = \chi^{(\cdot)}_{\cdot}\psi_{\cdot,k}^{(\cdot)}$, where

$$\chi_\nu^{(c)} = m^2 + (b^2+\nu^2)/a^2, \quad 0\leqslant\nu<\infty,$$
$$\chi_n^{(d)} = m^2 + (b^2-(n-|b|)^2)/a^2, \quad n=0,1,\ldots, \quad 0\leqslant n<|b|, \tag{3.14}$$

and the eigenfunctions are

$$\psi_{\alpha,\nu,k}^{(c)}(x_1,x_2) = e_\alpha a^{-1} f(\nu,k,\pm|b-\alpha/2|;x_1,x_2),$$
$$\pm k(b-\alpha/2)>0, \quad 0\leqslant\nu<\infty,$$
$$\psi_{0,k}^{(d)}(x_1,x_2) = e_{\mp 1}a^{-1}f_0(k,|b|+\tfrac{1}{2};x_1,x_2),$$
$$\pm b>0, \quad kb<0, \quad |b|>0,$$
$$\psi_{\alpha,n,k}^{(d)}(x_1,x_2) = e_\alpha a^{-1} f_{n-(\frac{1}{2})(1\mp\alpha)}(k,|b|\mp\alpha/2;x_1,x_2),$$
$$\pm b>0, \quad kb<0, \quad n=1,2,\ldots, \quad 1\leqslant n<|b|. \tag{3.15}$$

Here α takes the values $+1$ and -1 and $\{e_{+1}, e_{-1}\}$ consists of orthonormal constant eigenvectors of $\gamma^1\gamma^2$: $\gamma^1\gamma^2 e_\pm = \pm ie_{\pm 1}$. The functions f, and f_n are given in the Appendix. The eigenfunctions satisfy the following normalization conditions:

$$\int_{-\infty}^{\infty}\int_0^\infty \frac{a^2}{x_1^2}\,dx_1\,dx_2\,\psi_{\alpha,\nu,k}^{(c)\dagger}(x_1,x_2)\psi_{\alpha',\nu',k'}^{(c)}(x_1,x_2) = \delta_{\alpha,\alpha'}\delta(\nu-\nu')\delta(k-k'),$$
$$\int_{-\infty}^{\infty}\int_0^\infty \frac{a^2}{x_1^2}\,dx_1\,dx_2\,\psi_{\alpha,n,k}^{(d)\dagger}(x_1,x_2)\psi_{\alpha',n'k'}^{(d)}(x_1,x_2) = \delta_{\alpha,\alpha'}\delta_{n,n'}\delta(k-k'),$$
$$\int_{-\infty}^{\infty}\int_0^\infty \frac{a^2}{x_1^2}\,dx_1\,dx_2\,\psi_{0,k}^{(d)\dagger}(x_1,x_2)\psi_{0,k'}^{(d)}(x_1,x_2) = \delta(k-k'), \tag{3.16}$$

where all other inner products are zero. The heat kernel is given by

$$\begin{aligned}H_{\Theta_2}(s;x_1,x_2;x_1',x_1') &= \int_{kb<0} dk\,\Big\{e^{-s\chi_0^{(d)}}\psi_{0,k}^{(d)}(x_1,x_2)\psi_{0,k}^{(d)\dagger}(x_1',x_2')\\
&\quad + \sum_{\alpha=\pm 1,n=1}^{|b|} e^{-s\chi_n^{(d)}}\psi_{\alpha,n,k}^{(d)}(x_1,x_2)\psi_{\alpha,n,k}^{(d)\dagger}(x_1',x_2')\\
&\quad + \int_{-\infty}^{\infty}dk\int_0^\infty d\nu\sum_{\alpha=\pm 1}e^{-s\chi_\nu^{(c)}}\\
&\quad \times\psi_{\alpha,\nu,k}^{(c)}(x_1,x_2)\psi_{\alpha,\nu,k}^{(c)\dagger}(x_1',x_2').\end{aligned} \tag{3.17}$$

To compute the trace of H_{Θ_2} we proceed in the same way as for the Klein–Gordon case as we have that $\operatorname{tr} H_{\Theta_2}\times(s;x_1,x_2;x_1,x_2)$ is independent of (x_1,x_2). We obtain

$$\begin{aligned}Y_{\Theta_2}(s) &= V\operatorname{tr} H_{\Theta_2}(s;x_1,x_2;x_1,x_2)\\
&= \frac{V}{2\pi a^2}\Big\{|b|e^{-s\chi_0^{(d)}} + 2\sum_{n=1}^{|b|}(|b|-n)e^{-s\chi_n^{(d)}}\\
&\quad + \frac{1}{\pi}\int_0^\infty d\nu\,\nu\,\mathrm{Im}(\Psi(i\nu-b)+\Psi(i\nu-b+1)\\
&\quad + \Psi(i\nu+b+1)+\Psi(i\nu+b))e^{-s\chi_\nu^{(c)}}\Big\}.\end{aligned} \tag{3.18}$$

Again by going to the complex ν plane we can write

$$Y_{\Theta_2}(s) = \int_C d\nu\,\rho_{\Theta_2}(\nu)e^{-s\chi(\nu)}, \tag{3.19}$$

where

$$\rho_{\Theta_2}(\nu) = (V/4\pi^2a^2i)\nu(\Psi(i\nu-b)+\Psi(i\nu-b+1) + \Psi(i\nu+b+1)+\Psi(i\nu+b)),$$
$$\chi(\nu) = m^2 + (b^2+\nu^2)/a^2, \tag{3.20}$$

and the contour C is shown in Fig. 1(c). The zeta function of Θ_2 is again given by the Mellin transform of Y_{Θ_2}, i.e.,

$$\zeta_{\Theta_2}(z) = \int_C d\nu\,\rho_{\Theta_2}(\nu)\chi(\nu)^{-z}. \tag{3.21}$$

Convergence and analyticity of this function only occurs for $\mathrm{Re}\,z>1$. Again we must analytically continue Eq. (3.21) to the origin and this is done in the same way as was done for Eq. (3.11). We obtain

$$\zeta_{\Theta_2}(z) = \int_C d\nu\,\tilde\rho_{\Theta_2}(\nu)\chi(\nu)^{-z} + \frac{V}{2\pi}\,\frac{\chi(0)^{-z}}{z-1}, \tag{3.22}$$

where C is as in Fig. 1(d) and

$$\tilde\rho_{\Theta_2}(\nu) = \rho_{\Theta_2}(\nu) - (V/\pi^2a^2i)\nu\ln\nu. \tag{3.23}$$

Now $\zeta_{\Theta_2}(z)$ is given by a convergent integral for $\mathrm{Re}\,z>-\frac{1}{2}$ and is analytic in a neighborhood of the origin. The log determinant is given by $-\zeta'_{\Theta_2}(0)$, i.e., [C as for Eq. (3.33)]

$$\ln\det\Theta_2 = \int_C d\nu\,\tilde\rho_{\Theta_2}(\nu)\ln\chi(\nu) + \frac{V}{2\pi}\chi(0)(1-\ln\chi(0)). \tag{3.24}$$

IV. EFFECTIVE ACTION AND PAIR CREATION

We now compute the rate of pair creation and the effective action for the Klein–Gordon and Dirac fields of mass m in the presence of an external constant electromagnetic field and on a two-dimensional space-time with line element

$$ds^2 = (a^2/t^2)(-dt^2+dx^2), \tag{4.1}$$

which has constant curvature. We do this by taking the computations of the previous section, which were for the Riemannian metric (2.1) and (2.3), and performing a Wick rotation. This is possible for the above line element (4.1) as it admits a smooth timelike vector field whose integral curves have infinite proper length.[8,9]

For our case, the Wick rotation is implemented by considering the following family of Riemannian metrics:

$$ds^2 = (a^2/y_1^2)(\gamma\,dy_1^2+dy_2^2), \quad \gamma>0. \tag{4.2}$$

By a change of coordinates, $x_1=\gamma^{1/2}y_1$, $x_2=y_2$, the metric (4.2) becomes

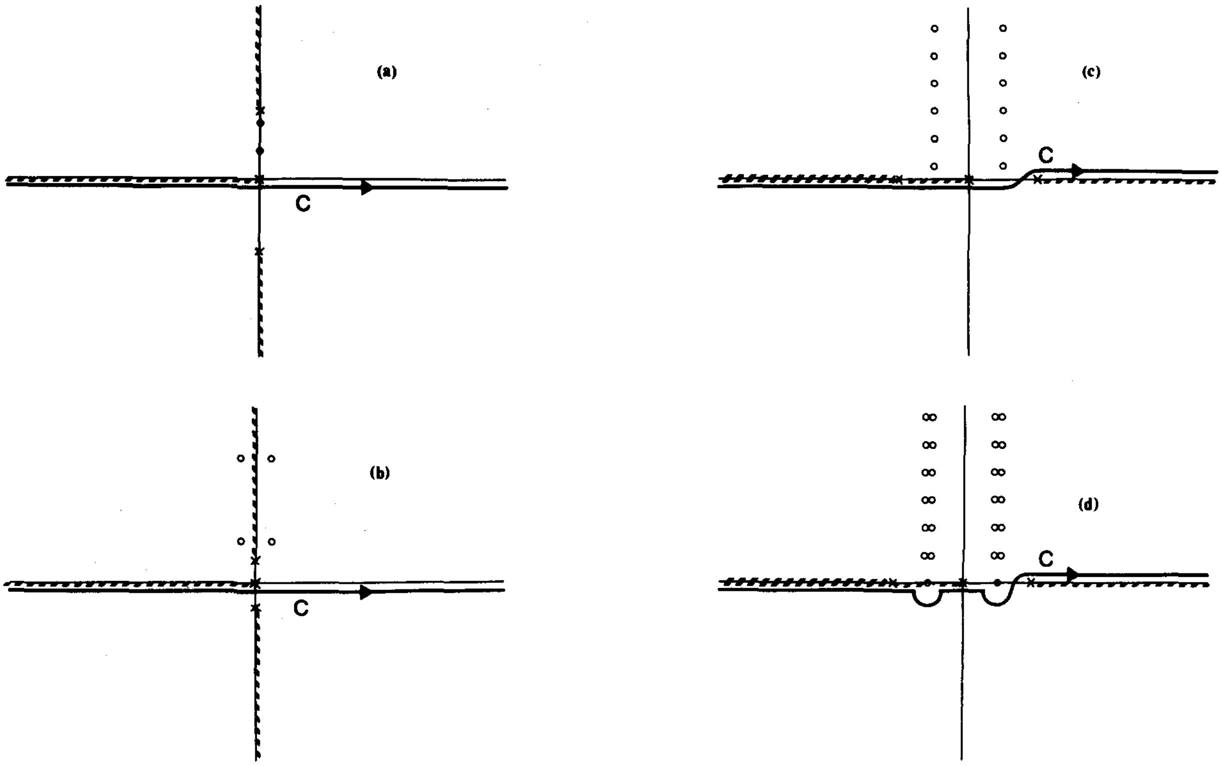


FIG. 2. The poles in (b) and (c) are at $\pm b + (n+\frac{1}{2})i$ and in (d) they are at $\pm b + ni$, $\pm b + (n+1)i$ $(n = 0,1,\ldots)$. In (b), (c), and (d) the branch points are at 0, $\pm i\sqrt{(\frac{1}{4} - m^2a^2 - b^2)}$; 0, $\pm\sqrt{m^2a^2 + b^2 - \frac{1}{4}}$; and 0, $\pm\sqrt{(m^2a^2+b^2)}$, respectively. The situation in (a) is identical to that of Fig. 1(b) except that in (a) $m^2a^2 + b^2 < \frac{1}{4}$. (The symbols used in this figure are the same as in Fig. 1.)

$$ds^2 = (a^2\gamma/x_1^2)(dx_1^2 + dx_2^2). \tag{4.3}$$

Making explicit the a and b dependence of the Klein–Gordon, Dirac, and Dirac-squared operators as given by equations (2.10)–(2.12), (2.13), and (2.14), and (2.16) and (2.17), respectively, by writing $\mathscr{K} = \mathscr{K}(a,b)$, $\Theta_1 = \Theta_1(a,b)$, and $\Theta_2 = \Theta_2(a,b)$, then the Klein–Gordon, Dirac, and Dirac-squared operators for a field of mass m in the presence of the electromagnetic field (2.8) and on the two-dimensional space with metric (4.3) are $\mathscr{K}(\gamma^{1/2}a, \gamma^{1/2}b)$, $\Theta_1(\gamma^{1/2}a, \gamma^{1/2}b)$, and $\Theta_2(\gamma^{1/2}a, \gamma^{1/2}b)$, as can be readily checked. The corresponding zeta functions $\zeta_{\mathscr{K}(\gamma^{1/2}a,\gamma^{1/2}b)}(z)$, $\zeta_{\Theta_2(\gamma^{1/2}a,\gamma^{1/2}b)}(z)$ are as given in the previous section with the replacement of (a,b) by $(\gamma^{1/2}a, \gamma^{1/2}b)$.

The next step is to consider is that of analytically continuing in the γ plane to $\gamma = e^{(\pi - 0)i}$. The Klein–Gordon–Dirac and Dirac-squared operators become operators with respect to the space-time metric (4.1) denoted $\widetilde{\mathscr{K}}(a,b)$, $\widetilde{\Theta}_1(a,b)$, and $\widetilde{\Theta}_2(a,b)$, respectively. The functions $\zeta_{\mathscr{K}(\gamma^{1/2}a,\gamma^{1/2}b)}(z)$, $\zeta_{\Theta_2(\gamma^{1/2}a,\gamma^{1/2}b)}(z)$ are analytically continued in the γ plane to $\gamma = e^{(\pi-0)i}$ while maintaining analyticity in a neighborhood of the origin in the z plane. The resulting expressions we take to be the zeta functions for $\widetilde{\mathscr{K}}(a,b)$ and $\widetilde{\Theta}_2(a,b)$ denoted $\zeta_{\widetilde{\mathscr{K}}(a,b)}(z)$ and $\zeta_{\widetilde{\Theta}_2(a,b)}(z)$, respectively. For the logarithmic determinants $\ln\det\widetilde{\mathscr{K}}(a,b)$, $\ln\det\widetilde{\Theta}_2(a,b)$, we compute $-(d/dz)\zeta_{\widetilde{\mathscr{K}}(a,b)}(0)$, $-(d/dz)\zeta_{\widetilde{\Theta}_2(a,b)}(0)$ and then $\ln\det\widetilde{\Theta}_1(a,b) = \frac{1}{2}\ln\det\widetilde{\Theta}_2(a,b)$. The real and imaginary parts of the logarithmic determinant give the Euler–Heisenberg effective action and the rate of pair creation, respectively.

A. The Klein–Gordon operator

Substituting $a \to \gamma^{1/2}a$, $b \to \gamma^{1/2}b$ in equations (3.9), (3.11), and (3.12) for the z-analytically continued zeta function we consider its analytic continuation to $\gamma = e^{(\pi-0)i}$ in the γ plane. The analytic structure and the contour of integration C in the z-plane gets deformed from the situation as shown in Fig. 2(a) to that of Fig. 2(b) when $\frac{1}{4} > b^2 + m^2a^2$ and from Fig. 1(b) to Fig. 2(c) when $\frac{1}{4} \leqslant b^2 + m^2a^2$. The zeta function $\zeta_{\widetilde{\mathscr{K}}(a,b)}(z)$ so obtained may be written in the form

$$\zeta_{\widetilde{\mathscr{K}}(a,b)}(z) = \frac{V}{4\pi^2} e^{z\pi i}\left\{ \frac{\pi|\tilde{w}(0)|^{1-z}}{1-z} - \frac{2}{a^2}\int_0^\infty \nu\, d\nu\, \eta_b(\nu)|\tilde{w}(\nu)|^{-z}\right\}, \quad \tfrac{1}{4} > b^2 + m^2a^2,$$

$$= \frac{V}{4\pi^2}\left\{ -\frac{\pi|\tilde{w}(0)|^{1-z}}{1-z} - \frac{2}{a^2}\int_0^{\sqrt{(m^2a^2+b^2-1/4)}} \nu\, d\nu\, \eta_b(\nu)|\tilde{w}(\nu)|^{-z} - \frac{2}{a^2}e^{z\pi i}\int_{\sqrt{(m^2a^2+b^2-1/4)}}^\infty \nu\, d\nu\, \eta_b(\nu)|\tilde{w}(\nu)|^{-z}\right\},$$
$$\tfrac{1}{4} \leqslant b^2 + m^2a^2, \tag{4.4}$$

where

$$\eta_b(\nu) = \mathrm{Im}(\Psi(i(\nu+b)+\tfrac{1}{2}) + \Psi(i(\nu-b)+\tfrac{1}{2})) - \pi,$$
$$\tilde{w}(\nu) = m^2 + (b^2 - \tfrac{1}{4} - \nu^2)/a^2. \tag{4.5}$$

The real and imaginary parts of the logarithmic determinant are as follows:

$$\mathrm{Re}(\ln\det \tilde{\mathscr{K}}(a,b)) = (V/4\pi^2)\Big\{ \mp \pi|\tilde{w}(0)|(1-\ln|\tilde{w}(0)|) - \frac{2}{a^2}\int_0^\infty \nu\,d\nu\,\eta_b(\nu)\ln|\tilde{w}(\nu)|\Big\},$$
$$\pm(\tfrac{1}{4} - (b^2 + m^2a^2)) \geqslant 0,$$
$$\mathrm{Im}(\ln\det \tilde{\mathscr{K}}(a,b)) = \frac{V}{4\pi}\Big\{ -\pi|\tilde{w}(0)| + \frac{2}{a^2}\int_0^\infty \nu\,d\nu\,\eta_b(\nu)\Big\},$$
$$\tfrac{1}{4} > b^2 + m^2a^2,$$
$$= \frac{V}{2\pi a^2}\int_{\sqrt{(m^2a^2+b^2-1/4)}}^\infty \nu\,d\nu\,\eta_b(\nu), \quad \tfrac{1}{4} \leqslant b^2 + m^2a^2. \tag{4.6}$$

B. The Dirac operator

Substituting $a \to \gamma^{1/2}a$, $b \to \gamma^{1/2}b$ in Eq. (3.20), (3.22), and (3.23) we again consider the analytic continuation to $\gamma = e^{(\pi-0)i}$. The analytic structure and contour of integration in the z plane gets deformed from the situation in Fig. 1(d) to that in Fig. 2(d) (the contour in the latter figure circumvents the poles which come to lie on the real-z axis). The zeta function so obtained is

$$\zeta_{\tilde{\Theta}_2(a,b)}(z) = \frac{V}{2\pi^2}\Big\{ \frac{\pi|\tilde{\chi}(0)|^{1-z}}{z-1} - \frac{e^{z\pi i}}{a^2}\int_{\sqrt{(m^2a^2+b^2)}}^\infty \nu\,d\nu\,\mu_b(\nu)|\tilde{\chi}(\nu)|^{-z} - \frac{1}{a^2}P\int_0^{\sqrt{(m^2a^2+b^2)}} \nu\,d\nu\,\mu_b(\nu)|\tilde{\chi}(\nu)|^{-z}\Big\}, \tag{4.7}$$

where P denotes the Cauchy principal value of the improper integral and where

$$\tilde{\chi}(\nu) = m^2 + (b^2 - \nu^2)/a^2,$$
$$\mu_b(\nu) = \mathrm{Im}(\Psi(i(\nu+b)) + \Psi(i(\nu-b)) + \Psi(i(\nu+b)+1) + \Psi(i(\nu-b)+1)) - 2\pi. \tag{4.8}$$

The real and imaginary parts of the logarithmic determinant are as follows [$\ln\det\tilde{\Theta}_1(a,b) = \tfrac{1}{2}\ln\det\tilde{\Theta}_2(a,b)$]:

$$\mathrm{Re}(\ln\det\Theta_1(a,b)) = (V/4\pi^2)\Big\{\pi|\tilde{\chi}(0)|(1-\ln|\tilde{\chi}(0)|) - \frac{1}{a^2}P\int_0^\infty \nu\,d\nu\,\mu_b(\nu)\ln|\tilde{\chi}(\nu)|\Big\},$$
$$\mathrm{Im}(\ln\det\tilde{\Theta}_1(a,b)) = \frac{V}{4\pi a^2}\int_{\sqrt{(m^2a^2+b^2)}}^\infty \nu\,d\nu\,\mu_b(\nu). \tag{4.9}$$

APPENDIX: DEFINITION AND PROPERTIES OF THE FUNCTIONS f AND f_n

The functions f and f_n, which were used in the text to give the eigenfunctions of the Klein–Gordon and Dirac operators, are now defined and some of their properties are discussed.

We take f to be the following function:

$$f(\nu,k,\beta;x_1,x_2) = \Big(\frac{\nu\sinh 2\pi\nu}{4\pi^3|k|}\Big)^{1/2}\Big|\Gamma(i\nu-\beta+\tfrac{1}{2})\Big|e^{-ikx_2} \times W_{\beta,i\nu}(2|k|x_1), \tag{A1}$$

where $\Gamma(\cdot)$ denotes the gamma function and $W_{\cdot,\cdot}(\cdot)$ denotes the solution of the Whittaker equation, i.e.,

$$\Big\{\frac{d^2}{dz^2} + \Big(-\frac{1}{4} + \frac{\beta}{z} + \frac{\frac{1}{4}-\mu^2}{z^2}\Big)\Big\}W_{\beta,\mu}(z) = 0, \tag{A2}$$

which can be expressed in terms of the confluent hypergeometric function

$$M(a,b;z) = 1 + \frac{a}{b}z + \frac{a(a+1)}{b(b+1)}z^2 + \cdots$$

in the following way:

$$W_{\beta,\mu}(z) = z^{1/2}e^{-z/2} \times\Big\{\frac{\Gamma(2\mu)}{\Gamma(\frac{1}{2}+\mu-\beta)}z^{-\mu}M(\tfrac{1}{2}-\mu-\beta, 1-2\mu;z) + \frac{\Gamma(-2\mu)}{\Gamma(\frac{1}{2}-\mu-\beta)}z^{\mu}M(\tfrac{1}{2}+\mu-\beta,1+2\mu;Z)\Big\}. \tag{A3}$$

We take f_n to be the following function:

$$f_n(k,\beta;x_1,x_2) = \Big(\frac{n!(2\beta-2n-1)}{4\pi|k|\Gamma(2\beta-n)}\Big)^{1/2}e^{-ikx_2}e^{-|k|x_1} \times(2|k|x_1)^{\beta-n}L_n^{(2\beta-2n-1)}(2|k|x_1), \tag{A4}$$

with $\beta>0$ and n is an integer such that $0\leqslant n\leqslant\beta-\frac{1}{2}$. Here $L_n^{(\alpha)}(z) = (\Gamma(n+\alpha+1)/(n!\Gamma(\alpha+1)))M(-n,\alpha+1;z)$ denotes the associated Laguerre polynomial of degree n.[12]

It follows by the properties of $W_{\cdot,\cdot}(\cdot)$ and $L_\cdot^{(\cdot)}(\cdot)$ that the functions f and f_n satisfy the following eigenvalue problems:

$$i\frac{\partial}{\partial x_2}\begin{cases}f(\nu,k,\beta;x_1,x_2)\\ f_n(k,\beta;x_1,x_2)\end{cases} = \begin{cases}kf(\nu,k,\beta;x_1,x_2),\\ kf_n(k,\beta;x_1,x_2),\end{cases}$$
$$-x_1^2\Big(\frac{\partial^2}{\partial x_1^2} - k^2 + \frac{2|k|\beta}{x_1}\Big)\begin{cases}f(\nu,k,\beta;x_1,x_2)\\ f_n(k,\beta;x_1,x_2)\end{cases} = \begin{cases}(\frac{1}{4}+\nu^2)+f(\nu,k,\beta;x_1,x_2), & 0\leqslant\nu<\infty,\\ (\frac{1}{4}-(\beta-n-\frac{1}{2})^2)f_n(k,\beta;x_1,x_2), & 0\leqslant n-\frac{1}{2}.\end{cases} \tag{A5}$$

Moreover, the multiplicative constants in f and f_n have been chosen in such a way that the following normalization conditions hold:

$$\int_{-\infty}^\infty\int_0^\infty \frac{1}{x_1^2}dx_1\,dx_2 f^*(\nu,k,\beta;x_1,x_2)f(\nu',k',\beta;x_1,x_2) = \delta(\nu-\nu')\delta(k-k'),$$
$$\int_{-\infty}^\infty\int_0^\infty \frac{1}{x_1^2}dx_1\,dx_2 f_n^*(k,\beta;x_1,x_2)f_{n'}(k',\beta;x_1,x_2) = \delta_{n,n'}\delta(k-k') \tag{A6}$$

(the inner product of f and f_n being zero).

We have the following integral identities involving the functions $W_{\cdot,\cdot}(\cdot)$ and $L_{\cdot}^{(\cdot)}(\cdot)$[13]

$$\int_0^\infty dz\, e^{-z} z^\alpha (L_n^{(\alpha)}(z))^2 = \Gamma(\alpha+n+1)/n!$$

$$= \alpha \int_0^\infty dz\, e^{-z} z^{\alpha-1} (L_n^{(\alpha)}(z))^2,$$

$$\int_0^\infty dz \frac{1}{z} (W_{\beta,\mu}(z))^2$$

$$= \frac{\pi(\Psi(\frac{1}{2}+\mu-\beta) - \Psi(\frac{1}{2}-\mu-\beta))}{\sin(2\pi\mu)\Gamma(\frac{1}{2}+\mu-\beta)\Gamma(\frac{1}{2}-\mu-\beta)}, \qquad \text{(A7)}$$

where $\Psi(z) = (d/dz)\ln\Gamma(z)$. From these we obtain for f and f_n:

$$\int_0^\infty d|k|\,|f(\nu,k,\beta;x_1,x_2)|^2 = \frac{1}{2\pi^2} \nu \,\mathrm{Im}\, \Psi(\tfrac{1}{2} + i\nu - \beta),$$

$$\int_0^\infty d|k|\,|f_n(k,\beta;x_1,x_2)|^2 = \frac{1}{2\pi}(\beta - n - \tfrac{1}{2}). \qquad \text{(A8)}$$

Finally, by making use of the asymptotic expansion of the Ψ function[12]

$$\Psi(z) = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} + \cdots, \qquad \text{(A9)}$$

as $|z| \to \infty$, $|\mathrm{Arg}\, z| \leqslant \pi$ we have

$$\mathrm{Im}(\Psi(\tfrac{1}{2} + i\nu - \beta) + \Psi(\tfrac{1}{2} + i\nu + \beta)) = \pi + O(\nu^{-3}), \qquad \text{(A10)}$$

as $\nu \to \infty$. Equation (A10) is used in the computation of the analytic continuation of the zeta function.

[1]S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[2]N. D. Birrell and P. C. W. Davies, *Quantum Field Theory in a Curved Space Time* (Cambridge U. P. Cambridge, 1982), and references therein.
[3]H. B. G. Casimir, Proc. Kon. Ned. Akad. Wet. **51**, 793 (1948).
[4]H. Euler and W. Heisenberg, Z. Phys. **98**, 714 (1936).
[5]J. Schwinger, Phys. Rev. **82**, 664 (1951).
[6]J. Madore, Phys. Rep. **75**, 125 (1981).
[7]N. E. Hurt, *Geometric Quantization in Action* (Reidel, Amsterdam, 1983).
[8]P. Candelas and D. J. Raine, Phys. Rev. D **15**, 1494 (1977).
[9]R. M. Wald, Commun. Math. Phys. **70**, 221 (1979).
[10]See, for example, J. A. Wolf, *Spaces of Constant Curvature* (McGraw–Hill, New York, 1967).
[11]We do not consider the metric of S^2, with constant positive curative, as it is not compatible with the criteria of Ref. 8 for a Wick rotation to be carried out.
[12]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[13]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965).

Symmetries of gauge fields

Moshe Molelekoa
National University of Lesotho, P. O. Roma 180, Lesotho, Africa

(Received 1 February 1984; accepted for publication 22 June 1984)

The concept of homogeneity for a gauge field is introduced as an appropriate geometrical interpretation of the symmetric properties of the gauge field relative to a group of space-time transformations. It is more general than the formulation in terms of an invariant connection, as illustrated by an example on translational symmetries. Sufficient conditions are given for invariance to be equivalent to homogeneity. Both finite and infinitesimal symmetries are treated.

I. INTRODUCTION

The main application of results describing symmetric gauge fields has been to find in a systematic way, solutions of the Yang–Mills equations.

In recent years, a general definition of an infinitesimal space-time symmetry for gauge fields was proposed by Bergmann and Flaherty.[1] Forgacs and Manton[2] extended the definition to include several symmetries. Meanwhile Harnad, Vinet, and Shnider[3] considered a group of finite symmetries that are transformations of space-time, and concluded that the geometric representation of a symmetric gauge field should be an invariant connection. There are other papers[4–6] which deal with space-time symmetries in gauge theories. In all these, the same assumptions are made to formulate the definition. In this paper, weaker assumptions are made and the resulting symmetry condition is termed "homogeneity" as in Brown and Weisberger,[7] where a special group of symmetries is treated from the following point of view.

Let M be a space-time manifold on which a (Lie) group H acts smoothly and P, a principal G-bundle on M. A gauge field on M (with gauge group G) is defined by a connection form ω on P. The field is said to be homogeneous (of class C^k) relative to H if the action of H on M lifts to a map

$$L{:}H \times P \to P$$

(of class C^k) such that for each h in H, the map $L(h)$ defined by $L(h)(p) = L(h,p)$, is an automorphism of ω covering h.

This approach is more general than those described earlier because the map L is not required to define an action on P. An extended example considered here, where H is the group of all translations on R^m, clearly illustrates the need for such a generalization. The symmetries defined above are finite. Since infinitesimal symmetries are more widely used in the literature, a corresponding definition is also given.

Notations and general definitions are given in Sec. II, while homogeneity itself is discussed in Sec. III. It is shown there that in a large class of important cases, homogeneity is equivalent to the invariance of the connection. However, in Sec. IV, a detailed example of translational symmetries, shows the difference between the two kinds of symmetry requirements.

In Sec. V, a corresponding definition of homogeneity is given for infinitesimal symmetries. Again this is contrasted with definitions in existing literature and shown to be more general. The concluding remarks (Sec. VI) show the need for this more general definition of symmetry for gauge fields particularly in application to the well-known case of the electromagnetic field. Other areas of study of homogeneous gauge fields are mentioned.

II. NOTATIONS AND DEFINITIONS

Let M denote an m-dimensional smooth manifold, G a Lie group, π: $P \to M$ a principal G-bundle, Aut(P), the group of all automorphisms of P (which are bundle maps), and ω, a connection form on P.

Each element of Aut(P) which covers the identity map on M is called a gauge transformation. If f is a gauge transformation, then there is a smooth map

$$g{:}\ P \to P$$

such that

$$f(p) = pg(p) \tag{2.1}$$

for all p in P.

A gauge field with gauge group G on M is represented[8] by a connection form ω on a principal G-bundle P over M.

Let H be a Lie group which acts on M from the left. Then each element h of H can be considered as a diffeomorphism of M. Consequently $L(h)$ is an element of Aut(P) which covers h.

Other standard notations and results in the theory of connections (e.g., Kobayashi and Nomizu[9]) will be used in the sequel. Specifically, Ω is the curvature form associated with the connection form ω, θ is the canonical 1-form on G, σ is a (local) cross section of P, $A = \sigma^*\omega$ is a gauge potential, and $F = \sigma^*\Omega$ is the field strength of the gauge field relative to the potential A.

III. FINITE SYMMETRIES AND HOMOGENEITY

Let $\pi{:}P \to M$ be a given principal G-bundle over M and ω a connection form on P. A finite gauge symmetry of ω is a gauge transformation

$$f{:}\ P \to P$$

satisfying

$$f^*\omega = \omega. \tag{3.1}$$

Equation (3.1) may also be written in the form

$$\omega = \mathrm{ad}(g^{-1})\omega + g^*\theta, \tag{3.2}$$

 0022-2488/85/010192-06$02.50

where g is as defined in Eq. (2.1) and θ is the canonical 1-form on G.

Lemma 3.1: Suppose f is a gauge symmetry of ω. Then $g(p)$, as defined in Eq. (2.1), commutes with every element of the holonomy group of ω at p.

Proof: Let k be any element of the holonomy group at p. Then there is a smooth horizontal curve v_t with $0 \leqslant t \leqslant 1$, such that

$$v_0 = p \text{ and } v_1 = pk$$

Since f preserves the connection, the curve $f(v_t)$ is also horizontal.

Following Kobayashi and Nomizu[9] (p. 69), let

$$u_t = f(v_t) = v_t a_t,$$

where

$$a_t = g(v_t) \quad \text{[cf. Eq. (2.1)]}.$$

Differentiation with respect to t yields

$$\dot{u}_t = \dot{v}_t a_t + v_t \dot{a}_t.$$

Since $\dot{u}_t$ and $\dot{v}_t$ are horizontal, the application of ω to the former gives the following result:

$$0 = \omega(\dot{u}_t) = \mathrm{ad}(a_t^{-1})\omega(\dot{v}_t) + a_t^{-1}\dot{a}_t = a_t^{-1}\dot{a}_t.$$

It follows then that

$$\dot{a}_t = 0 \quad \text{and so } a_t = g(p) \quad \text{for all } t.$$

At $t = 1$, this gives

$$u_1 = f(v_1) = v_1 a_1 = pkg(p),$$

because

$$a_1 = g(p).$$

On the other hand, we have

$$u_1 = f(pk) = f(p)k = pg(p)k.$$

So, $g(p)$ commutes with k. ■

The concept of a space-time symmetry for gauge fields is perhaps more appropriately applied when the base space is the four-dimensional space-time manifold. However, the term is used more generally on any smooth manifold M. So, suppose h is a diffeomorphism of M. Then h is said to be a finite space-time symmetry of ω if the equation

$$L(h)^*\omega = \omega$$

holds with $L(h)$ an element of Aut(P) which covers h. Observe that when $L_1(h)$ and $L_2(h)$ are two such maps, then the map

$$L_1(h)^{-1}L_2(h): P \to P$$

is a finite gauge symmetry of ω. Thus if ω has nontrivial gauge symmetries, there are several possible lifts of h which fix ω.

Now, suppose H is a Lie group which acts on M from the left. The action of each h in H on an element x of M will be denoted by $h(x)$ as if h is a diffeomorphism of M. The gauge field defined by ω will be said to be homogeneous (of class C^k) relative to H if the action of H on M lifts to a map

$$L: H \times P \to P \tag{3.3a}$$

(of class C^k) such that

$$\text{(i) } L(h,p) = L(h)(p), \tag{3.3b}$$

where $L(h)$ is an element of Aut(P) covering h and satisfying

$$L(h)^*\omega = \omega, \tag{3.3c}$$

$$\text{(ii) } L(1,p) = p \tag{3.3d}$$

for every p in P, with 1 representing the identity of H.

Condition (i) above merely expresses the requirement that each h in H be a symmetry of ω. When (i) holds, the second equation (ii) can always be achieved by redefining L, if necessary, to be K where

$$K(h,p) = L(1)^{-1}(L(h,p)).$$

If the map L in Eq. (3.3) is smooth, the gauge field is said to be smoothly homogeneous relative to H.

Harnad *et al.*[3] who have defined finite space-time symmetries of gauge fields, require that L define a smooth action of H on P. It will be shown that such a regularity condition may be very restrictive (Sec. IV) and undesirable (see Sec. VI also). If L is chosen to be a smooth action, then ω will be said to be invariant under H. Thus, an invariant gauge field is necessarily homogeneous. The converse, while it does not hold in general (Sec. IV), is valid with certain mild restrictions given in the following propositions.

Proposition 3.1: Suppose G has a discrete center and H is connected. Let ω define an irreducible connection which is smoothly homogeneous relative to H. Then ω is invariant under H.

Proof: Since ω is smoothly homogeneous, there is a smooth map

$$L: H \times P \to P$$

satisfying conditions (i) and (ii) of Eqs. (3.3). We claim that L is a smooth action of H on P. To prove this, the following lemma is used.

Lemma 3.2: Under the hypotheses of Proposition 3.1, the smooth map L of Eqs. (3.3) satisfies the equation

$$L(h^{-1}) = L(h)^{-1} \quad \text{for all } h \text{ in } H.$$

Proof: For each p in P, the map from H to P defined by $L(h^{-1})L(h)(p) = L(h^{-1}, L(h,p))$ is smooth. Hence, the equation

$$L(h^{-1})L(h)(p) = pg(h)$$

holds for some smooth map g from H to G.

Since $L(h^{-1})L(h)$ is a gauge symmetry of ω, $g(h)$ commutes with each element of the holonomy group of ω at p (Lemma 3.1). Since ω is irreducible, this means that $g(h)$ lies in the center $Z(G)$ of G. Now, $Z(G)$ is closed in G, so

$$g: H \to Z(G)$$

is a smooth map. Since H is connected, $g(1) = 1$, and $Z(G)$ is discrete, $g(h) = 1$ for every h in H. Hence, the result follows.

Now, to complete the proof of Proposition 3.1, let k be an element of H. Define $K: H \times P \to P$ by

$$K(h,p) \equiv K(h)(p) = L(kh)L(h^{-1})L(k^{-1})(p).$$

Then K is a smooth map and $K(h)$ is a finite gauge symmetry of ω. Hence, by an argument similar to the one used in the proof of Lemma 3.2, $K(h)$ is the identity map.

Therefore, $L(kh) = L(k)L(h)$, and this completes the proof of Proposition 3.1. ■

Proposition 3.2: A gauge field is invariant under a one-parameter group H of transformations of the space-time manifold if and only if the field is smoothly homogeneous relative to H.

Proof: As stated before, invariance always implies smooth homogeneity. Thus we need only prove the converse.

Suppose the gauge field defined by ω is smoothly homogeneous relative to H. Without loss of generality, we may assume that H is given by a smooth action of the group R of real numbers on M,

$h{:}R \times M \to M$ say,

where $h(t,x) = h_t(x)$ and h_t is in H.

By definition of smooth homogeneity, there is a smooth map

$L{:}R \times P \to P$

satisfying Eqs. (3.3). Let $\bar{X}$ be the smooth vector field defined on P by

$$\bar{X} = \frac{dL(t)}{dt}$$

evaluated at $t = 0$, with $L(t)(p) = L(t,p)$ by previous notation.

Since $L(t)$ is an element of Aut(P), it follows that $\bar{X}$ is G-invariant. Moreover, $L_{\bar{X}}\omega$ is a tensorial 1-form. Hence, $L_{\bar{X}}\omega$ vanishes if and only if for every (local) cross section σ of P, $\sigma^*(L_{\bar{X}}\omega) = 0$.

It is easy to show that

$$\sigma^*(L_{\bar{X}}\omega) = i_{\bar{X}} F - D\psi,$$

where $\psi = \sigma^*(\omega(\bar{X}))$, $D\psi = d\psi + [A,\psi]$, $A = \sigma^*\omega$, and F is the corresponding field strength. Equivalently,

$$\sigma^*(L_{\bar{X}}\omega) = L_X A - DW,$$

where W and ψ are related by $W = A(X) + \psi$ and $X = \pi_*(\bar{X})$. So, $L_{\bar{X}}\omega$ vanishes if and only if for every local cross section σ,

$$L_X A = dW + [A,W] \equiv DW \tag{3.4}$$

for some $W = A(X) + \psi$, where ψ is a (local) tensorial function.

Now let $\sigma{:}U \to P$ be a cross section of P. Suppose x is an element of U and t is sufficiently close to 0 so that $h(t,x)$ is in U, then

$$L(t,\sigma(x)) = \sigma(h(t,x))T_t(x)^{-1},$$

where $T_t(x)$ varies smoothly with t and $T_0(x) = 1$ for all x in U. Also, at x and for small enough t,

$$h_t^* A = \mathrm{ad}(T_t^{-1})(A) + T_t^*\theta,$$

where θ is the canonical 1-form on G. Hence, differentiation at 0 gives the result

$$L_X A = [A,W] + dW, \tag{3.5}$$

where $W = (dT_t/dt)$ at $t = 0$. Differentiation of $L(t,\sigma(x))$ above at $t = 0$ and the application of ω, shows that $A(X) + W$ is a tensorial function. So, $L_{\bar{X}}\omega = 0$.

Let $K(t)$ be the one-parameter group of automorphisms of P generated by the G-invariant vector field $\bar{X}$, and k_t the diffeomorphism of M induced by $K(t)$. Then each $K(t)$ is an automorphism of ω and

$$\begin{aligned} k_{t+s}\pi &= \pi K(t+s) \\ &= \pi K(t)K(s) \\ &= k_t \pi K(s) \\ &= k_t k_s \pi. \end{aligned}$$

This implies that k_t is a one-parameter group of transformations which generate X. By the uniqueness of the flow of a vector field and the condition after Eq. (3.3b), $k_t = h_t$ for all t. This shows that ω is invariant under the action defined by $K(t,p) = K(t)(p)$ on P. Since this action covers h, ω is invariant under H. ■

Before we consider an example of a homogeneous gauge field with a view of contrasting homogeneity with invariance, let us examine the implications of homogeneity on the base manifold M. For simplicity, first assume that P has a global section σ.

Suppose that h is a diffeomorphism of M and that $L(h)$ is an element of Aut(P) covering h. Relative to σ, $L(h)$ is then completely determined by the smooth map

$T_h{:}M \to G$

defined by

$$L(h)(\sigma(x)) = \sigma(h(x))T_h(x)^{-1}. \tag{3.6}$$

The connection form ω is uniquely determined by the gauge potential $A = \sigma^*\omega$. The corresponding field strength is $F = \sigma^*\Omega$, where Ω is the curvature form.

If $L(h)^*\omega = \omega$, and hence h is a symmetry of ω, then T_h is called a transformation function of A relative to the symmetry h (cf. Harnad *et al.*[3]).

It is clear that a necessary and sufficient condition for h to be a symmetry of ω is that

$$\sigma^*\omega = \sigma^*(L(h)^*\omega)$$

for some $L(h)$ in Aut(P) covering h. In view of Eq. (3.6), h is a symmetry of ω if and only if there is a smooth G-valued map T_h on M such that

$$A = \mathrm{ad}(T_h)(h^*A) + T_h^{-1*}\theta \tag{3.7}$$

or equivalently,

$$h^*A = \mathrm{ad}(T_h^{-1})(A) + T_h^*\theta. \tag{3.8}$$

[cf. Eq. (3.2)], where θ is the canonical 1-form on G.

When Eq. (3.8) holds, it follows that

$$h^*F = \mathrm{ad}(T_h^{-1})(F). \tag{3.9}$$

Equation (3.8) merely states the fact that A and h^*A are equivalent gauge potentials. (3.10)

Suppose H is a Lie group of diffeomorphisms of M, and ω is homogeneous relative to H. Let T_h be a transformation function of A relative to each h in H. Then T_{hk} is related to T_h and T_k by

$$T_{hk}(x) = c(h,k)(x)T_k(x)T_h(k(x)), \tag{3.11}$$

where $c(h,k)(x)$ may not be the identity element of G. However, if the transformation functions can be chosen such that $c \equiv 1$, then we obtain an invariant connection.[3] In general, such a choice may not be possible. This is shown in the next section.

In case P is not a trivial bundle, we can get only local cross sections $\sigma{:}U \to P$. Then Eqs. (3.6)–(3.8) are well defined if $h(U)$ is contained in U. Thus, a sufficient condition for homogeneity relative to H, is that P have a trivialization consisting of open subsets on each of which H restricts to an action. Moreover, Eq. (3.8) must hold on each open subset, and the transformation functions must satisfy certain compatibility conditions (see Ref. 3). If $h(U)$ is not contained in U, two different cross sections are necessary in Eq. (3.6). Then T_h depends on both $L(h)$ and the cross sections used.

IV. HOMOGENEITY RELATIVE TO TRANSLATIONS

In this section, M is the space R^m and $H = R^m$ acts by translations on M. Thus there is a map

$$t:H \times M \to M,$$

$$t(x,y) = y + x \equiv t_x(y) = t_y(x).$$

Suppose ω is a connection form on a trivial principal G-bundle

$$\pi:P \to M,$$

and

$$\sigma:M \to P$$

is a global cross section. Let $A = \sigma^*\omega$ and $F = \sigma^*\Omega$ (where Ω is the curvature form) be the gauge potential and its field strength, respectively.

Proposition 4.1: If ω is smoothly homogeneous relative to H, there is a choice of gauge such that the transformed field strength has constant components relative to the natural coordinates of R^m.

Remark: A choice of gauge means a choice of trivialization of P.

Proof: Let $L:H \times P \to P$ be the smooth map which satisfied (i) and (ii) of Eqs. (3.3). Define a map λ from M to P by

$$\lambda(y) = L(y,\sigma(0)) = \sigma(y)T_y(0)^{-1}, \tag{4.1}$$

where T_y is the transformation function of A relative to y in H [cf. Eqs (3.3) and (3.6)]. The map λ is a global cross section of P and hence defines a choice of gauge.

Relative to this gauge, the transformed field strength $\bar{F}$ is

$$\lambda^*\Omega = \tfrac{1}{2}\bar{F}_{ij}\, dx^i \wedge dx^j,$$

where the summation convention is adopted henceforth, and $1 \le i,j \le m$. Then $\bar{F}$ at x is given by

$$\bar{F}_x = \operatorname{ad}(T_x(0))(F_x) \quad \text{[from Eq. (4.1)]}$$

$$= \operatorname{ad}(T_x(0))(t_x{}^*F)_0$$

$$= \operatorname{ad}(T_x(0))(\operatorname{ad}(T_x(0)^{-1})(F_0)) \quad \text{[from Eq. (3.9)]}$$

$$= F_0. \qquad \blacksquare$$

Proposition 4.2: If G is abelian, the converse of Proposition 4.1 holds.

Proof: Suppose F_{ij} are elements of the Lie algebra of G such that $F_{ij} = -F_{ji}$ for $1 \le i,j \le m$. Let A be the gauge potential whose field strength is

$$F = \tfrac{1}{2}F_{ij}\, dx^i \wedge dx^j.$$

Let $B = B_i\, dx^i$, where $B_i(x) = \tfrac{1}{2}F_{ji}\, x^j$. Then $dB = F = dA$ implies that B is gauge equivalent to A. This follows from the general fact that for abelian G and simply connected M, any two gauge potentials are gauge equivalent if and only if they have the same field strength. Thus, to show that A is homogeneous, we may assume that $A = B$, in view of the fact that t_y^*A and t_y^*B are gauge equivalent for all y in H [cf. line (3.10)].

Define $L:H \times P \to P$ such that

$$L(y,\sigma(x)a) = \sigma(x+y)T_y(x)^{-1}a \quad \text{for all } a \text{ in } G,$$

and

$$T_y(x) = \exp(\tfrac{1}{2}F_{ij}y^i x^j). \tag{4.2}$$

Then L is a smooth map such that

$$L(0,p) = p \quad \text{for all } p \text{ in } P.$$

Moreover,

$$L(y)(\sigma(x)a) \equiv L(y,\sigma(x)a)$$

defines $L(y)$ as an element of $\operatorname{Aut}(P)$ which covers t_y.

The following equation also holds:

$$(T_y^{-1})^*\theta = T_y\, dT_y^{-1} = -A_i(y)dx^i.$$

The last expression is $-A_y$, i.e., $-A$ at the point y of M, and so

$$(t_y^*A)_x = A_i(x+y)dx^i$$

$$= A_i(x)dx^i + A_i(y)dx^i$$

$$= A_x + A_y.$$

Thus, we have proved that

$$A_x = (t_y^*A)_x - A_y,$$

i.e.,

$$A = \operatorname{ad}(T_y)(t_y^*A) + T_y^{-1*}\theta,$$

noting that G is abelian and so $\operatorname{ad}(T_y)$ is the identity map. Hence, by Eq. (3.7), each t_y is a symmetry of A. $\blacksquare$

Proposition 4.3: Suppose G is abelian. Then a connection is invariant under H if and only if it is flat.

Proof: Suppose ω is invariant under H. Then the action

$$t:H \times M \to M$$

lifts to an action

$$L:H \times P \to P$$

such that $L(y)^*\omega = \omega$, where $L(y)(p) \equiv L(y,p)$.

Let p_0 be an element of $\pi^{-1}(0) \subseteq P$, and

$$\sigma(x) = L(x,p_0).$$

Then σ is a global cross section of P, and

$$L(y,\sigma(x)) = L(y,L(x,p_0)) = L(y+x,p_0)$$

$$= \sigma(x+y).$$

So, the corresponding transformation function

$$T_y(x) \equiv 1,$$

and thus, by Eq. (3.7),

$$A = t_y^*A. \tag{4.3}$$

This implies that $dA = 0$, and so the connection is flat.

Conversely, if the connection is flat, then $F \equiv 0$. Let σ be any global cross section of P and $A = \sigma^*\omega$. Then A is equivalent to a gauge potential B with constant components, i.e.,

$B = t_y^* B$ for all y in H. By changing the gauge, if necessary, we may assume that $A = B$. Define $L{:}H \times P \to P$ such that

$$L(y,\sigma(x)a) = \sigma(x+y)a \quad \text{for all } a \text{ in } G.$$

Then L is a smooth action of H on P covering t. Moreover,

$$L(y)(p) \equiv L(y,p) \text{ defines } L(y)$$

as an element of Aut(P) covering t_y and

$$\sigma^*(L(y)^*\omega) = t_y{}^*(\sigma^*\omega) = \sigma^*\omega \quad \text{for all } y \text{ in } H.$$

This implies that

$$L(y)^*\omega = \omega \quad \text{for all } y \text{ in } H. \qquad \blacksquare$$

Remarks: Propositions 4.1 and 4.2 imply that for abelian G, a gauge field is smoothly homogeneous relative to H if and only if the field F has constant components relative to the natural coordinates of R^m. It is well known that F completely determines the connection when G is abelian. Thus, invariance of the connection as a criterion for a gauge field to have H as a group of space-time symmetries (as in Ref. 3) may be rather restrictive, in view of Proposition 4.3.

Note that the converse to Proposition 4.1 does not hold for arbitrary nonabelian groups. To see this, take G to be SU(2). The Lie algebra of G is isomorphic to R^3 with the cross product as the multiplication. The adjoint action of G on its Lie algebra corresponds to rotations in R^3.

Take $F_{ij}(x)$ to be (0,0,0) except for

$$F_{31}(x) = (1,0,0) = -F_{13}(x).$$

Let

$$A_2(x) = (0,0,0),$$
$$A_3(x) = (0,1,0),$$

and

$$A_1(x) = (0,x^1,1).$$

Then $A = A_i\, dx^i$ is a gauge potential with field strength

$$F = \tfrac{1}{2} F_{ij}\, dx^i \wedge dx^j.$$

The connection defined by A is irreducible because the Lie algebra of the infinitesimal holonomy group at $\sigma(x)$, which is generated by $F_{ij}(x)$ and all its covariant derivatives at $\sigma(x)$ contains the element $F_{31} = (1,0,0)$ and

$$D_3F_{31} = A_3 \times F_{31} = (0,0,-1)$$

and hence all of R^3.

Suppose the gauge field were homogeneous relative to H. Then for each y in H, t_y^*A would be gauge equivalent to A [line (3.10)]. So all vectors D_iF_{jk} relative to t_y^*A, would be obtainable from the corresponding ones relative to A by a rotation $r(x)$ which depends (smoothly) only on the points of M, because these vectors are tensorial functions.

Now $D_3F_{31} = (0,0,-1)$ relative to all t_y^*A. So, if the assumption were true, $r(x)$ would fix $(0,0,-1)$ and hence, by linearity, fix the third coordinate of each point of R^3. However,

$$D_1F_{13} = A_1 \times F_{13} = (0,-1,x^1+y^1)$$

relative to t_y^*A. So, the third coordinate is not fixed. This contradiction shows that the converse to Proposition 4.1 fails in this case.

V. INFINITESIMAL SYMMETRIES

Most of the space-time symmetries considered in the literature[1–6] are infinitesimal instead of finite. In this section we show how such symmetries result from the definitions in Sec. III. First, we review gauge symmetries. There are two equivalent ways of defining infinitesimal gauge symmetries.

Let ω be a connection form on a principal G-bundle P over M. A G-invariant vertical vector field $\bar{X}$ on P is said to be an infinitesimal gauge symmetry of ω if $L_{\bar{X}}\omega = 0$, where $L_{\bar{X}}$ is the Lie derivative relative to $\bar{X}$. It is a well-known fact that such a condition holds if and only if there is a tensorial function $\bar{\psi}$ which is covariant constant, i.e., $D\bar{\psi} = 0$, where D is the covariant differentiation relative to ω. The relation between $\bar{\psi}$ and $\bar{X}$ is given by $\bar{\psi} = \omega(\bar{X})$.

Let us now consider space-time symmetries. Suppose that $h{:}R \times M \to M$ is a smooth action of the group R of real numbers on M, and X the vector field on M which generates this action. Let ω define a gauge field on M as before. From the proof of Proposition 3.2 follows that ω is smoothly homogeneous relative to h if and only if X lifts to a G-invariant vector field $\bar{X}$ on P such that $L_{\bar{X}}\omega = 0$. In view of this, we make the following definition.

A vector field X on M is called an infinitesimal space-time symmetry of ω if X lifts to a G-invariant vector field $\bar{X}$ on P such that

$$L_{\bar{X}}\omega = 0.$$

In case $\mathscr{h}$ is a Lie algebra of vector fields on M, ω is said to be homogeneous relative to $\mathscr{h}$ if each element of $\mathscr{h}$ is a symmetry of ω.

Remarks: There is an obvious extension of the definitions of symmetry to locally defined diffeomorphisms. Then it follows also from the proof of Proposition 3.2 that X is a symmetry of ω if and only if ω is invariant relative to the flow of X. Another equivalent condition is that Eq. (3.4) holds locally on M. This latter condition is that adopted as the definition of a (infinitesimal space-time) symmetry of a gauge field in the literature.[1–6]

Let us write $W = W_X$ and call it a symmetry function of A relative to X, provided Eq. (3.4) holds. Then any other symmetry function relative to X differs from W_X by an infinitesimal gauge symmetry. Observe that the linearity of Eq. (3.4) in X and W implies that once a basis is chosen for $\mathscr{h}$, the symmetry functions may be chosen to be linear in X. Assuming that this has been done, let W_y be the symmetry function of A relative to Y. Then, by evaluating L_XL_YA, it can be shown that

$$L_{[X,Y]}A = D(XW_Y - YW_X + [W_X,W_Y]),$$

i.e., $XW_Y - YW_X + [W_X,W_Y]$ is a symmetry function of A relative to $[X,Y]$ and hence differs from the preselected $W_{[X,Y]}$ by

$$C(X,Y) = W_{[X,Y]} - XW_Y + YW_X - [W_X,W_Y].$$

Now, by definition of homogeneity, each X in $\mathscr{h}$ lifts to a G-invariant vector field $\bar{X}$ such that $L_{\bar{X}}\omega$ vanishes. The map which associated $\bar{X}$ with X, is, in general, not a Lie algebra homomorphism. Thus $C(X,Y)$ measures obstruction to obtaining a lift of $\mathscr{h}$ to an infinitesimal action of $\mathscr{h}$ on the bundle space P, preserving the connection.

This obstruction $C(X,Y)$ is assumed to be zero in the literature.[2-6] However, from the example of Sec. IV, it will be shown that C may not vanish.

We showed when ω is smoothly homogeneous relative to translation on R^m, with P trivial, that the transformation functions are given, in some gauge, by Eq. (4.2). From this we deduce that the symmetry functions may then be given by

$$\begin{aligned} W_y(x) &= \frac{d}{dt} T_{ty}(x) \text{ at } t = 0. \\ &= \tfrac{1}{2} F_{ij} y^i x^j \end{aligned}$$

[cf. Eq. (3.5)], where the subscript y is used in place of the vector field whose components are the coordinates of y.

By taking $y = e_i$ (the ith vector of the natural basis of R^m) one obtains $W_i(x) \equiv W_y = \frac{1}{2} F_{ij} x^j$. Hence, the obstruction C is given by

$$\begin{aligned} C(\partial_i,\partial_j) &= \partial_j W_i - \partial_i W_j \\ &= \tfrac{1}{2} F_{ij} - \tfrac{1}{2} F_{ji} \\ &= F_{ij} \end{aligned}$$

which, in general, is not identically equal to zero for all the possible subscripts.

This example illustrates the fact that even for infinitesimal space-time symmetries, homogeneity is more general than the definition used in the literature.[2-6]

VI. CONCLUDING REMARKS

In this paper, we have defined homogeneity of a gauge field, both for finite and infinitesimal (space-time) symmetries. The preceding discussion shows that the concept of homogeneity is more general in its applications to the description of a gauge field with a prescribed group of space-time symmetries than other notions in the existing literature. Specifically, it has been shown that it is applicable to the case where the gauge group is abelian. It is conceivable that for nonabelian groups which have nondiscrete centers, the concept may still be more useful and appropriate than the others mentioned before.

The classical electromagnetic field is an example of an abelian gauge field. It is completely determined by its field strength, the components of which are those of the electric and magnetic fields. It is therefore reasonable to consider this field as having translational symmetries if and only if the components are constant, a result which homogeneity gives, but which invariance does not. The latter requires that these components vanish, a condition which is obviously very stringent for this symmetry.

In a subsequent paper, a study of the space of homogeneous Yang–Mills fields will be made using the techniques and the results of Arms[10] who considered gauge symmetries only. The latter symmetries are related to space-time symmetries in the following sense: A space-time symmetry is given by any automorphism of the connection; whereas a gauge symmetry is an automorphism which covers the identity map of the base space M.

The classification of invariant connections when H acts transitively on M (Wang's theorem) is well known in mathematics.[9] Harnad *et al.*[3] have extended this classification to the case of a "simple" action of H on M. The classification of homogeneous gauge fields has been considered by Westwater (see the Acknowledgment) and the author, and a complete description for the transitive action has been done.[11]

Forgacs and Manton[2] have described a way of constructing fields with a given Lie algebra of (infinitesimal) space-time symmetries. In their construction it is assumed that there is no obstruction to lifting its action on the space-time manifold to one on the bundle manifold such that it fixes the connection. It is well known, in mathematics, that such a lift is not always possible. It would be too restrictive to consider only those actions for which the lift is possible. Thus, it would be interesting to examine the extension of their results to the case of homogeneous fields. The extension may be useful in the study of the space of (homogeneous) Yang–Mills fields.

ACKNOWLEDGMENTS

The author is grateful to Professor G. E. Tanyi (National University of Lesotho, Lesotho) and Professor J. M. Arms (University of Washington, Seattle, WA) for their helpful suggestions on the writing of this paper; and to Professor J. M. Westwater (University of Washington) who suggested this discussion which forms part of a Ph. D. dissertation.[12] Since this paper was written, several other papers which either preceded or are related to Refs. 1–6 have been brought to my attention. A partial list is given in Ref. 13.

[1] P. G. Bergmann and E. J. Flaherty, J. Math. Phys. **19**, 212 (1978).
[2] P. Forgacs and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[3] J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980).
[4] R. Jackiw and N. S. Manton, Ann. Phys. **127**, 257 (1980).
[5] R. Jackiw, Phys. Rev. Lett. **41**, 1635 (1978).
[6] R. Jackiw, Acta Phys. Austriaca, Suppl. **XXII**, 383 (1980).
[7] L. S. Brown and W. I. Weisberger, Nucl. Phys. B **157**, 285 (1979).
[8] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[10] J. M. Arms, Proc. Cambridge Philos. Soc. **90**, 361 (1981).
[11] This work has not yet been published.
[12] M. Molelekoa, "Spacetime symmetries of Yang–Mills fields," Ph. D. thesis, Univ. of Washington, Seattle, Washington (1982).
[13] M. Henneaux, J. Math. Phys. **23**, 830 (1982); A. Trautman, Rep. Math. Phys. (Torun) **1**, 29 (1970); **10**, 297 (1976); Czech. J. Phys. B **29**, 107 (1979); Bull. Acad. Pol. Sci. **27**, 7 (1979); J. Phys. A **13**, L 1 (1980); M. E. Mayer, Ann. Israel Phys. Soc. **3**, 80 (1979); *Lecture Notes in Physics* (Springer, New York, 1979), Vol. 116, 291–295; Hadronic J. **4**, 108 (1980); Acta Phys. Austriaca, Suppl. **XXIII**, 477 (1981). M. E. Mayer and A. Trautman, Acta Phys. Austriaca, Suppl. **XXIII**, 433 (1981); D. Bleecker, *Gauge Theory and Variational Principles* (Addison-Wesley, 1981); P. Schwab, Phys. Lett. B **109**, 47 (1982); P. Cordero and C. Teitelboim, Ann. Phys. (NY) **100**, 607 (1976); R. Benguria, P. Cordero, and C. Teitelboim, Nucl. Phys. B **122**, 61 (1977); V. N. Romanov, A. S. Schwarz, and Yu. S. Tyupkin, Nucl. Phys. B **130**, 209 (1977); A. S. Schwarz, Commun. Math. Phys. **56**, 79 (1977).

Projective corepresentations and invariants of PCT group

P. Rudra
Department of Physics, University of Kalyani, Kalyani, West Bengal, 741235, India

(Received 27 September 1983; accepted for publication 8 June 1984)

The comultiplicator group and the nonassociated factor systems for the internal symmetry group of parity (P), charge conjugation (C), and time reversal (T) operators have been obtained. It is found that there are 16 nonassociated factor systems for this group. The basic invariants for all irreducible corepresentations for these factor systems have been constructed.

I. INTRODUCTION

Elementary particles having strong interactions are described[1] not only by the space-time Lorentz group symmetry but also by the discrete symmetry elements of parity (P), charge conjugation (C), and time reversal (T). Of these the T operator is antilinear[2] in nature and this fact makes use of Wigner's corepresentation theory[2] obligatory[3–5] in the investigations of the internal symmetry of strongly interacting particles.

Wigner showed[6,7] that intrinsic spin of particles can be described on the basis of considerations of projective representations of the Lorentz group belonging to different factor systems. Thus, it appears that a complete description of internal properties of particles would require knowledge of all the nonassociated factor systems. In the case of the Lorentz group there are just two such factor systems. In this paper we have constructed all the nonassociated factor systems of the unitary–antiunitary (or magnetic) group generated by P, C, and T. The procedure for this construction is already known in literature.[5,8–10] We have also given the projective irreducible corepresentation matrices for the group elements.

In analyzing different physical properties the integrity basis of invariants[11–15] forms an essential tool. Any invariant of the group can be written as an algebraic function of these integrity basis. In the last section we have obtained the integrity basis for this group for all the inequivalent irreducible projective corepresentations for the different factor systems.

II. PROJECTIVE FACTOR SYSTEMS AND IRREDUCIBLE COREPRESENTATIONS

The underlying linear symmetry group[9,10] M' is defined by the relation $a_1^2 = a_2^2 = a_3^2 = e$, $a_i a_j = a_j a_i$, $\forall i,j = 1,2,3$. Here a_1 stands for P, a_2 for CT, and a_3 for T. The magnetic group $M(G)$ has the linear part $G = \{a_1, a_2\}$ generated by a_1 and a_2. Thus, $M(G) = G\cup Ga_3$. In the notation of Ref. 10, the matrix group extension $\hat{M}$ is defined by the relations

$$\begin{aligned} &A_1{}^2 = A_2{}^2 = E, \quad A_3 A_3^* = J(3)E, \\ &A_2 A_1 = J(2,1) A_1 A_2, \quad A_3 A_1^* = J(3,1)^* A_1 A_3, \\ &A_3 A_2^* = J(3,2)^* A_2 A_3, \quad A_3 F^* = F^{-1} A_3, \end{aligned} \tag{1}$$

where

$$F = J(2,1), J(3,1)^*, J(3,2)^*,$$

$$CD = DC, \quad \text{where } C = A_1, A_2, J(2,1), J(3,1)^*$$

and

$$D = J(3), J(2,1), J(3,1)^*, J(3,2)^*.$$

From these relations we get

$$J(2,1)^2 = J(3,1)^{*2} = J(3,2)^{*2} = J(3)^2 = E. \tag{2}$$

The comultiplicator group $K(M,G)$, which is the intersection of the derived group $[\hat{M}\hat{M}]$ and the group J generated by the elements $J(2,1), J(3,1)^*, J(3,2)^*, J(3)$ turns out to be $K(M,G) = J = \{J(2,1), J(3,1)^*, J(3,2)^*, J(3)\}$. With the definition of the factor system $\omega(\alpha,\beta)$,

$$D(\alpha) D(\beta)^{[\alpha]} = \omega(\alpha,\beta)^{[\alpha\beta]} D(\alpha\beta), \quad \forall \alpha,\beta \in M(G),$$

where

$$A^{[\alpha]} = A \quad \text{if } \alpha \in G, \ A^* \quad \text{if } \alpha \in M(G) - G. \tag{3}$$

Here A is either a matrix or a scalar. We also have the following relations for all $\alpha, \beta, \gamma \in M(G)$:

$$\omega(\alpha,\beta)^{[\gamma]} \omega(\alpha\beta,\gamma) = \omega(\alpha,\beta\gamma) \omega(\beta,\gamma), \text{ and } |(\alpha,\beta)| = 1. \tag{4}$$

Two factor systems $\omega'(\alpha,\beta)$ and $\omega(\alpha,\beta)$ are associated if there are complex numbers $c(\alpha)$ of modulus unity such that

$$\omega'(\alpha,\beta)^{[\alpha,\beta]} = \omega(\alpha,\beta)^{[\alpha\beta]} \cdot c(\alpha) \cdot c(\beta)^{[\alpha]} / c(\alpha\beta), \quad \forall \alpha,\beta \in M(G). \tag{5}$$

For the $K(M,G)$ obtained above we thus get $\omega(a_3,a_3) = \pm 1$,

$$\omega(a_2,a_1) = \pm 1, \quad \omega(a_3,a_1)^* = \pm 1, \quad \omega(a_3,a_2)^* = \pm 1.$$

The projective irreducible representations of G corresponding to the case $\omega(a_2,a_1) = +1$ (in this case it is actually the vector representation) are Δ^1 to Δ^4, all one dimensional given in Table I, and that for the case $\omega(a_2,a_1) = -1$, is a two-dimensional one Δ^5 and is given in Table II. When we want to build the irreducible projective corepresentation of $M(G)$ we have the following three types.[2]

Type (*a*): $\Delta^{\bar{\mu}}(u) \equiv \omega(u,a_3)^* \omega(a_3, a_3{}^{-1} u a_3) \Delta^{\mu}(a_3{}^{-1} u a_3)^*$ is equivalent to Δ^{μ} for all $u \in G$, i.e., $\Delta^{\bar{\mu}}(u) = P^{-1} \Delta^{\mu}(u) P$ with

TABLE I. Irreducible representations of G corresponding to the projective factor system $\omega(a_2,a_1) = +1$.

Irreducible representation	Irreducible character	Group elements			
		e	a_1	a_2	$a_1 \cdot a_2$
Δ^1	ψ^1	1	1	1	1
Δ^2	ψ^2	1	1	-1	-1
Δ^3	ψ^3	1	-1	1	-1
Δ^4	ψ^4	1	-1	-1	1

TABLE II. Irreducible representation of G corresponding to the projective factor system $\omega(a_2,a_1) = -1$.

Irreducible representation	Irreducible character	Group elements			
		e	a_1	a_2	$a_1 \cdot a_2$
Δ^5		$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$	$\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$	$\begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$

$PP^* = +\omega(a_3,a_3)\Delta^{\mu}(a_3^2)$. In this case $D^{\mu}(u) = \Delta^{\mu}(u)$, $\forall u \in G$, and $D^{\mu}(a) = \pm\omega(aa_3^{-1}, a_3)\, \Delta^{\mu}(aa_3^{-1})P$ for all $a \in M(G) - G$. The necessary and sufficient condition for this case to occur is

$$\mathscr{C} = \sum_{a \in M(G) - G} \omega(a,a)\Psi^{\mu}(a^2) = +|G|,$$

where Ψ^{μ} is the character of Δ^{μ} and $|G|$ is the order of the group G. The conventions used here will be followed later also.

Type (*b*): $\quad \Delta^{\bar{\mu}}(u) = P^{-1}\Delta^{\mu}(u)P, \quad \forall u \in G,\quad$ with $PP^* = -\omega(a_3,a_3)\Delta^{\mu}(a_3^2)$. In this case

$$D^{\mu}(u) = \begin{pmatrix}\Delta^{\mu} & 0\\ 0 & \Delta^{\mu}(u)\end{pmatrix}$$

and

$$D^{\mu}(a) = \begin{pmatrix} 0 & -\omega(aa_3^{-1},a_3)\Delta^{\mu}(aa_3^{-1})P \\ \omega(aa_3^{-1},a_3)\Delta^{\mu}(aa_3^{-1})P & 0\end{pmatrix}$$

and the necessary and sufficient criterion is $\mathscr{C} = -|G|$.

Type (*c*): $\Delta^{\bar{\mu}}(u)$ is not equivalent to $\Delta^{\mu}(u)$, i.e., $\Sigma_u \Psi^{\mu}(u) \times \Psi^{\bar{\mu}}(u)^* = 0$. In this case

$$D^{(\mu,\bar{\mu})}(u) = \begin{pmatrix}\Delta^{\mu}(u) & 0\\ 0 & \Delta^{\bar{\mu}}(u)\end{pmatrix}$$

and

$$D^{(\mu,\bar{\mu})}(a) = \begin{pmatrix} 0 & \omega(aa_3)\Delta^{\mu}(aa_3) \\ \omega(a_3,a_3^{-1}a)\Delta^{\mu}(a_3^{-1}a)^* & 0\end{pmatrix}$$

and the necessary and sufficient condition is $\mathscr{C} = 0$.

Applying these well-known facts we get the following result for the PCT group. When $\omega(a_2,a_1) = +1$, then for either $\omega(a_3,a_1)^*$ or $\omega(a_3,a_2)^*$ equalling -1, $\mathscr{C} = 0$ for all Δ^{μ} and all the irreducible projective corepresentations are of type (c) and are two dimensional. If, on the other hand, $\omega(a_3,a_1)^* = \omega(a_3,a_2)^* = +1$, then $\mathscr{C} = 4\omega(a_3,a_3)$, and the corepresentations are either of type (a) [i.e., one dimensional, when $\omega(a_3,a_3) = +1$, this being the case of vector corepresentation] or of type (b) [i.e., two dimensional when $\omega(a_3,a_3) = -1$]. When $\omega(a_2,a_1) = -1$, then for either $\omega(a_3,a_1)^*$ or $\omega(a_3,a_2)^*$ equalling $+1$, $\mathscr{C} = 4\omega(a_3,a_3)$ and the irreducible projective corepresentations are either of type (a) [i.e., two dimensional when $\omega(a_3,a_3) = +1$] or of type (b) [i.e., four dimensional when $\omega(a_3,a_3) = -1$]. If, on the other hand, $\omega(a_3,a_1)^* = \omega(a_3,a_2)^* = -1$, then $\mathscr{C} = -4\omega(a_3, a_3)$ and the irreducible projective corepresentations are either of type (b) [i.e., four dimensional when $\omega(a_3,a_3) = +1$] or of type (a) [i.e., two dimensional when $\omega(a_3,a_3) = -1$]. We summarize the resulting sixteen cases of projective factor systems arising from the different combinations of values for $\omega(a_2,a_1)$, $\omega(a_3,a_3)$, $\omega(a_3,a_1)^*$, $\omega(a_3,a_2)^*$ in Table III. The corresponding corepresentation matrices can of course be easily constructed.

III. INVARIANTS AND INTEGRITY BASIS

We now describe the method of obtaining the invariants of a magnetic group $M(G)$ formed out of the bases u_i's forming the μth irreducible corepresentation $D^{\omega\mu}$ with character $\chi^{\omega\mu}$ belonging to the factor system $\omega(\alpha,\beta)$. Let $u_1,\ldots,u_n$ be the basis forming D^{μ}, so that the operation of the Wigner operator O_α, $\alpha \in M(G)$, on the product basis $u_{i_1}\cdots u_{i_s}$ will be given by

$$O_\alpha u_{i_1}\cdots u_{i_s} = \sum_{j_1\cdots j_s} \underbrace{[D^{\omega\mu}(\alpha)\otimes\cdots\otimes D^{\omega\mu}(\alpha)]_{j_1\cdots j_s, i_1\cdots i_s}}_{s\ \text{terms}} \cdot u_{j_1}\cdots u_{j_s}. \tag{6}$$

Here the Kronecker direct product corepresentation

$$[\underbrace{D^{\omega\mu}\otimes\cdots\otimes D^{\omega\mu}}_{s\ \text{terms}}]$$

belongs to the factor system $\omega(\alpha,\beta)^s$. An invariant of degree s of the form

$$\sum_{i_1\cdots i_s} C_{i_1\cdots i_s} u_{i_1}\cdots u_{i_s}$$

must then satisfy the linear equation

$$C_{j_1\cdots j_S} = \sum_{i_1\ldots i_1} C^{[\alpha]}_{i_1\cdots i_s} \underbrace{[D^{\omega\mu}(\alpha)\otimes\cdots\otimes D^{\omega\mu}(\alpha)]_{j_1\cdots j_s, i_1\cdots i_s}}_{s\ \text{terms}},$$

$$\forall \alpha \in M(G). \tag{7}$$

Equation (7) will have, in general, more than one solution; in fact the number of independent solutions will be equal to the multiplicity of the identity corepresentation in the Kronecker direct product corepresenation. If χ^{Kron} is the character of this direct product corepresentation then the number of independent solutions of Eq. (7) is evidently[16]

$$\frac{2}{|M|}\sum_{u\in G}\chi^{\mathrm{Kron}}(u),$$

where $|M|$ is the order of the magnetic group $M(G)$. We can, of course, simplify χ^{Kron} in a more tractible form. If we break up s in s cycles $\nu_1 + \nu_2 + \cdots + \nu_s = \lambda_1$, $\nu_2 + \cdots + \nu_s = \lambda_2,\ldots,\nu_s = \lambda_s$, and $\nu_1 + 2\nu_2 + \cdots + s\nu_s = \lambda_1 + \lambda_2 + \cdots + \lambda_s = s$, then it turns out after a short combinatorial calculation that

TABLE III. Sixteen types of projective factor systems and types of irreducible corepresentations, their dimensions, and the generic names of the bases.

Projective factor systems and different cases	Irreducible corepresentation	Type of corepresentation	Dimension of corepresentation	Generic name of bases
$\omega(\alpha,\beta)^{[\alpha\beta]} = +1,\quad \alpha,\beta\in M(G)$, Case 1.	D^1 to D^4	Type (a)	1-D	u_1
$\omega(a_2,a_1)=\omega(a_3,a_1)^*=\omega(a_3,a_2)^*=+1$, $\omega(a_3,a_3)=-1$. Case 2.	D^1 to D^4	Type (b)	2-D	u_1,u_2
$\omega(a_2,a_1)=\omega(a_3,a_1)^*=+1$, $\omega(a_3,a_2)^*=-1$, $\omega(a_3,a_3)=+1$, Case 3. $\omega(a_3,a_3)=-1$, Case 6.	$D^{(1,2)}$, $D^{(3,4)}$	Type (c)	2-D	u_1,u_2
$\omega(a_2,a_1)=\omega(a_3,a_2)^*=+1$, $\omega(a_3,a_1)^*=-1$, $\omega(a_3,a_3)=+1$, Case 4. $\omega(a_3,a_3)=-1$, Case 7.	$D^{(1,3)}$, $D^{(2,4)}$	Type (c)	2-D	u_1,u_2
$\omega(a_2,a_1)=+1$, $\omega(a_3,a_1)^*=\omega(a_3,a_2)^*=-1$, $\omega(a_3,a_3)=+1$, Case 5. $\omega(a_3,a_3)=-1$, Case 8.	$D^{(1,4)}$, $D^{(2,3)}$	Type (c)	2-D	u_1,u_2
$\omega(a_2,a_1)=-1$, $\omega(a_3,a_3)=\omega(a_3,a_1)^*=\omega(a_3,a_2)^*=+1$, Case 9.	D^5	Type (a)	2-D	u_1,u_2
$\omega(a_2,a_1)=\omega(a_3,a_2)^*=-1$, $\omega(a_3,a_3)=\omega(a_3,a_1)^*=+1$, Case 10.	D^5	Type (a)	2-D	u_1,u_2
$\omega(a_2,a_1)=\omega(a_3,a_1)^*=-1$, $\omega(a_3,a_3)=\omega(a_3,a_2)^*=+1$, Case 11.	D^5	Type (a)	2-D	u_1,u_2
$\omega(\alpha,\beta)^{[\alpha\beta]}=-1,\quad \forall\alpha,\beta\in M(G)$, Case 12.	D^5	Type (a)	2-D	u_1,u_2
$\omega(a_2,a_1)=\omega(a_3,a_1)^*=\omega(a_3,a_2)^*=-1$, $\omega(a_3,a_3)=+1$, Case 13.	D^5	Type (b)	4-D	u_1,u_2,u_3,u_4
$\omega(a_2,a_1)=\omega(a_3,a_3)=-1$, $\omega(a_3,a_1)^*=\omega(a_3,a_2)^*=+1$, Case 14.	D^5	Type (b)	4-D	u_1,u_2,u_3,u_4
$\omega(a_2,a_1)=\omega(a_3,a_3)=\omega(a_3,a_2)^*=-1$, $\omega(a_3,a_1)^*=+1$, Case 15.	D^5	Type (b)	4-D	u_1,u_2,u_3,u_4
$\omega(a_2,a_1)=\omega(a_3,a_3)=\omega(a_3,a_1)^*=-1$, $\omega(a_3,a_2)^*=+1$, Case 16.	D^5	Type (b)	4-D	u_1,u_2,u_3,u_4

$$\chi^{\,\mathrm{Kron}}(u)=\frac{1}{s!}\sum_{\substack{\text{cycles}\\ 1^{\nu_1}2^{\nu_2}\cdots s^{\nu_s}}}\frac{s!}{1^{\nu_1}\cdot\nu_1!2^{\nu_2}\cdot\nu_2!\cdots s^{\nu_s}\cdot\nu_s!}\cdot\omega(u,u)^{\lambda_2}\omega(u^2,u)^{\lambda_3}\cdots\omega(u^{s-1},u)^{\lambda_s}\cdot\chi^{\omega\mu}(u)^{\nu_1}\chi^{\omega\mu}(u^2)^{\nu_2}\cdots\chi^{\omega\mu}(u^s)^{\nu_s}. \tag{8}$$

It should be noted that there would be no invariant of degree s unless $\omega(\alpha,\beta)^s=1$, for $\forall\alpha,\beta\in M(G)$. It is mentioned that this result is true for all magnetic groups and not just for the PCT group. Though all these invariants of degree s are linearly independent, some of them may be algebraically expressible with real coefficients in terms of the rest and those of degree lower than s. The minimum number of them form the integrity basis. In Table IV we give the invariants and their integrity basis for all the projective factor systems of the PCT group.

IV. CONCLUSION

Finally we want to indicate an interesting result of our analysis. We know[1] the result of the operations of the PCT group on the wavefunctions of the spin-$\frac{1}{2}$ Dirac fermions. With a change of phase convention from that of Ref. 1, we get

$$\begin{aligned}
a_1=\mathrm{P}&:\ \Psi(t,\mathbf{r})\longrightarrow i\eta_{\mathrm{P}}\gamma^0\Psi(t,-\mathbf{r}),\\
a_2=\mathrm{CT}&:\ \Psi(t,\mathbf{r})\longrightarrow i\eta_{\mathrm{P}}\eta_{\mathrm{PCT}}\gamma^1\gamma^2\gamma^3\Psi(-t,\mathbf{r}),\\
a_1a_2=\mathrm{PCT}&:\ \Psi(t,\mathbf{r})\longrightarrow\eta_{\mathrm{PCT}}\gamma^0\gamma^1\gamma^2\gamma^3\Psi(-t,-\mathbf{r}),\\
a_3=\mathrm{T}&:\ \Psi(t,\mathbf{r})\longrightarrow -i\eta_{\mathrm{T}}\gamma^1\gamma^3\Psi(-t,\mathbf{r})^*, \qquad (9a)\\
a_1a_3=\mathrm{PT}&:\ \Psi(t,\mathbf{r})\longrightarrow\eta_{\mathrm{P}}\eta_{\mathrm{T}}\gamma^0\gamma^1\gamma^3\Psi(-t,-\mathbf{r})^*,\\
a_2a_3=\mathrm{C}&:\ \Psi(t,\mathbf{r})\longrightarrow\eta_{\mathrm{P}}\eta_{\mathrm{PCT}}\eta_{\mathrm{T}}\gamma^2\Psi(t,\mathbf{r})^*,\\
a_1a_2a_3=\mathrm{PC}&:\ \Psi(t,\mathbf{r})\longrightarrow -i\eta_{\mathrm{PCT}}\eta_{\mathrm{T}}\gamma^0\gamma^2\Psi(t,-\mathbf{r})^*,
\end{aligned}$$

with $\eta_{\mathrm{P}}^2=-1$, $\eta_{\mathrm{PCT}}^2=1$, $|\eta_{\mathrm{T}}|^2=1$ and the Dirac matrices γ^α in the standard representation

TABLE IV. Invariants and integrity basis for invariants up to degree four for all the projective factor systems.

Different cases and irreducible corepresentations	Invariants up to degree four	Integrity basis
Case 1. D^1	$1; [+]u_1, [-]iu_1; u_1^2;$ $[+]u_1^3, [-]iu_1^3; u_1^4;$	$1; [+]u_1, [-]iu_1;$
$D^\mu, \mu = 2,3,4$	$1; u_1^2; u_1^4;$	$1; u_1^2;$
Case 2. $D^\mu, \mu = 1,2,3,4$	$1; (u_1^2 + u_2^2), i(u_1^2 - u_2^2), iu_1u_2;$ $(u_1^4 + u_2^4), i(u_1^4 - u_2^4), u_1^2u_2^2,$ $(u_1^3u_2 - u_1u_2^3), i(u_1^3u_2 + u_2u_2^3);$	$1; (u_1^2 + u_2^2), i(u_1^2 - u_2^2),$ $iu_1u_2;$
Cases 3, 6. $D^{(1,2)}, D^{(3,4)}$ Cases 4, 7. $D^{(1,3)}, \Delta^{(2,4)}$ Cases 5, 8. $D^{(1,4)}, D^{(2,3)}$	$1; (u_1^2 + u_2^2), i(u_1^2 - u_2^2); u_1^2u_2^2,$ $(u_1^4 + u_2^4), i(u_1^4 - u_2^4);$	$1; (u_1^2 + u_2^2), i(u_1^2 - u_2^2);$
Cases 9–12. D^5	$1; (u_1^2 + u_2^2); u_1^2u_2^2, (u_1^4 + u_2^4);$	$1;(u_1^2 + u_2^2); u_1^2u_2^2;$
Cases 13–16. D^5	$1; (u_1^2 + u_2^2 + u_3^2 + u_4^2),$ $i(u_1^2 + u_2^2 - u_3^2 - u_4^2), i(u_1u_3 + u_2u_4);$ $u_1u_2u_3u_4, (u_1^4 + u_2^4 + u_3^4 + u_4^4),$ $i(u_1^4 + u_2^4 - u_3^4 - u_4^4),$ $(u_1^3u_3 + u_2^3u_4 - u_1u_3^3 - u_2u_4^3),$ $i(u_1^3u_3 + u_2^3u_4 + u_1u_3^3 + u_2u_4^3),$ $(u_1^2u_3^2 + u_2^2u_4^2), (u_1^2u_4^2 + u_2^2u_3^2),$ $(u_1^2u_2^2 + u_3^2u_4^2), i(u_1^2u_2^2 - u_3^2u_4^2),$ $(u_1^2u_2u_4 + u_1u_2^2u_3 - u_1u_3u_4^2 - u_2u_3^2u_4),$ $i(u_1^2u_2u_4 + u_1u_2^2u_3 + u_1u_3u_4^2 + u_2u_3^2u_4);$	$1; (u_1^2 + u_2^2 + u_3^2 + u_4^2),$ $i(u_1^2 + u_2^2 - u_3^2 - u_4^2),$ $i(u_1u_3 + u_2u_4); u_1u_2u_3u_4,$ $(u_1^4 + u_2^4 + u_3^4 + u_4^4),$ $i(u_1^4 + u_2^4 - u_3^4 - u_4^4),$ $(u_1^3u_3 + u_2^3u_4 - u_1u_3^3 - u_2u_4^3),$ $i(u_1^3u_3 + u_2^3u_4 + u_1u_3^3 + u_2u_4^3);$

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}. \tag{9b}$$

This transformation corresponds to $\omega(a_2,a_1) = -1$, $\omega(a_3,a_3) = -1, \omega(a_3,a_1)^* = 1, \omega(a_3,a_2)^* = -1$, i.e., the type (b) four-dimensional case 15 of our analysis. Since any $\frac{1}{2}$-integral spin field can be built up as an odd number direct product of spin-$\frac{1}{2}$ field, all of them belong to the same factor system. This factor system is just one of the 16 possible factor systems. If we have the belief that fields must exist belonging to all possible factor systems, as do in the case of the Lorentz group, then it will be interesting to investigate these other fields.

Note added in proof: Spin-zero bosons belong to the corepresentations D^1 and D^4 of our case 1. Of these the scalar bosons belong to D^1 and the pseudoscalar bosons to D^4.

[1]V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory, Part I* (Pergamon, London, 1971).
[2]E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[3]L. Michel, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gursey (Gordon and Breach, New York, 1964), Vol. I.
[4]F. Kamber and N. Straumann, Helv. Phys. Acta **37**, 563 (1964).
[5]Russell E. Warren and William H. Klink, J. Math. Phys. **13**, 306 (1972).
[6]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[7]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, MA, 1962).
[8]K. Parthasarathy, Commun. Math. Phys. **15**, 305 (1969).
[9]T. Janssen, J. Math. Phys. **13**, 342 (1972).
[10]P. Rudra, J. Math. Phys. **17**, 509 (1976).
[11]W. Burnside, *Theory of Groups of Finite Order* (Dover, New York, 1955).
[12]H. Weyl, *Classical Groups* (Princeton U. P., Princeton, NJ, 1946).
[13]R. P. Biekerstaff and B. G. Wybourne, J. Phys. A **9**, 1051 (1976).
[14]V. Kopsky, J. Phys. C **8**, 3251 (1975).
[15]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976).
[16]P. Rudra, J. Math. Phys. **15**, 2031 (1974).

A twistor approach to Nahm's equations

P. R. Wainwright
Department of Mathematical Sciences, University of Durham, Science Laboratories, Durham DH1 3LE, United Kingdom

(Received 14 December 1983; accepted for publication 11 May 1984)

It has been shown that Nahm's equations $\mathbf{T}' = \mathbf{T} \wedge \mathbf{T}$ may be regarded as dimensionally reduced self-duality conditions. Ward's twistor method is applied to construct a holomorphic vector bundle corresponding to each solution of these equations, and a simple expression for the transition matrix is derived. Some elementary examples are given for which this procedure may be explicitly reversed to generate solutions of Nahm's equations from given transition matrices.

I. INTRODUCTION

The exact multimonopole solutions of nonabelian gauge theories have received much attention in recent years, and several methods have been applied to construct them. The approach of Atiyah and Ward[1,2] leads to a hierarchy of *Ansätze*[3,4] which suffice to generate all self-dual multimonopoles; however, a set of transcendental constraints must be applied to the parameters of the solution to ensure that it is nonsingular. The method originally developed for instantons by Atiyah, Drinfeld, Hitchin, and Manin (ADHM)[5,6] and modified for the monopole case by Nahm[7,8] leads automatically to a nonsingular potential. However, it presupposes the solution of a set of nonlinear differential equations, known as Nahm's equations, for three matrices $T_i(z)$. Using the observation[9] that these equations are themselves the self-duality equations for a one-dimensional field, we apply Ward's twistor method and find the "transition matrix" (2.16) which acts as a kind of generating function for the solutions $T_i(z)$.

It is well-known that a nonabelian gauge field may be described geometrically by a connection on a certain vector bundle. The gauge groups G which we shall consider here are U(n) and its subgroups O(n), USp(n); in each case the corresponding vector bundle is $\mathbb{R}^4 \times E$, where E is a vector space of dimension n over $\mathbb{C}$. In the ADHM construction[5,6] for a gauge field with k instantons on Euclidean four-space, the curved connection A_μ is induced by embedding this bundle in a *flat* bundle $\mathbb{R}^4 \times V$, where V has dimension $n + 2k$ over $\mathbb{C}$. The embedding v is described by a linear map $v(x): E \to V$ at each point $x \in \mathbb{R}^4$; the range of $v(x)$ is the fiber of the subbundle $v(\mathbb{R}^4 \times E)$ at x. This procedure gives rise to a potential of the form

$$A_\mu(x) = v(x)^\dagger \partial_\mu v(x). \tag{1.1}$$

Choosing bases for E and V, $v(x)$ is represented by an $(n + 2k) \times n$ matrix whose columns form a basis for the subbundle $v(\mathbb{R}^4 \times E)$ at the point x. A different choice of this basis gives rise to a potential differing only by a gauge transformation. Let $\Delta(\mathbb{R}^4 \times E')$ be the subbundle orthogonal to $v(\mathbb{R}^4 \times E)$, so that E' has dimension $2k$. Then the embeddings Δ, v are related by

$$\Delta(x)^\dagger v(x) = 0. \tag{1.2}$$

Choosing bases for E' and V, $\Delta(x)$ is represented by an $(n + 2k) \times 2k$ matrix. In what follows we shall use the standard representation of quaternions by 2×2 matrices $e_0 = 1$, $e_i = -i\sigma_i$ $(i = 1,2,3)$; thus the quaternions have a natural action upon $\mathbb{C}^2$ and upon $E' = \mathbb{C}^k \times \mathbb{C}^2$. Points of $\mathbb{R}^4$ correspond to quaternions via $x = x^\mu e_\mu$. ADHM found[5] that the k-instanton solutions can be derived from an embedding Δ of the form

$$\Delta(x) = a + bx, \tag{1.3}$$

where a, b are constant maps $E' \to V$. Let the components of $u \in E'$ in some basis be denoted by u_{sA} $(s = 1,2,\ldots,k;\ A = 1,2)$, and let those of $v \in V$ be $v_r^{B'}$ $(r = 1,2,\ldots,k;\ B' = 1,2)$ and v_i $(i = 1,2,\ldots,n)$. It is possible to choose bases in such a way that $b_r^{B'sA} = \delta_r{}^s \delta^{B'A}$; then the equation $v^\dagger \Delta = 0$ (cf. 1.2) becomes

$$v_{B'}^{\dagger r}(x^{B'A}\delta_r{}^s + e_\mu^{B'A} a_r^{\mu s}) + v^{\dagger i} a_i^{sA} = 0. \tag{1.4}$$

It is easy to prove[6] that the connection A_μ is nonsingular if $\Delta^\dagger \Delta$ is nonsingular; it is *self-dual* if this matrix commutes with the quaternions, that is, if there exists a Hermitian matrix $H_r{}^s(x)$ such that

$$[\Delta(x)^\dagger \Delta(x)]_{rB}{}^{sA} = H_r{}^s(x) \delta_B{}^A. \tag{1.5}$$

The self-dual instanton solutions constructed above are localized in the coordinate x^0 as well as in x^i; but a multimonopole solution is a self-dual gauge field *independent* of x^0. Such a solution may however be regarded as a limiting case of a multiinstanton when $k \to \infty$ (see Ref. 10). These considerations led Nahm to modify the ADHM construction to treat monopoles.[7,8] Now E' and V are *Hilbert spaces*, and $\Delta(x)$ is a differential operator $E' \to V$ such that $\Delta^\dagger(x)$ has a kernel of finite dimension n.

Suppose we seek solutions for which the Higgs field A_0 takes the asymptotic form at spatial infinity, where $r = |\mathbf{x}|$

$$iA_0^{ij} = \delta^{ij}\left(z_i - \frac{k_i}{2r}\right) + O\left(\frac{1}{r^2}\right), \tag{1.6}$$

where the k_i are integers and $z_i \leqslant z_{i+1}$. Let

$$I = \{i : k_i = 0\}, \tag{1.7}$$

$$k(z) = \sum_i k_i \theta(z_i - z), \quad z \in \mathbb{R}. \tag{1.8}$$

Then an element v of the Hilbert space V has components $v_r^{B'}(z)$, where $r = 1,2,\ldots,k(z)$, and s_i, where $i \in I$. The equation $v^\dagger \Delta = 0$ in Nahm's construction is a differential equation

$$v_{B'}^{\dagger r}(z)\left\{x^{B'A}\delta_r{}^s + ie_m^{B'A} T_r^{ms}(z) - i\frac{\overleftarrow{d}}{dz}\delta^{B'A}\delta_r{}^s\right\} + \sum_{i \in I} s_i^\dagger a_i^{sA} \delta(z - z_i) = 0, \tag{1.9}$$

 0022-2488/85/010202-06$02.50

where the functions $T_r^{ms}(z)$ are yet to be determined. In fact this equation has n linearly independent solutions $(v^{(i)}, s^{(i)})$ at each point x, and the connection A_μ is derived from them by a formula analogous to (1.1):

$$A_\mu^{ij}(x) = \langle v^{(i)}, s^{(i)} | \partial_\mu | v^{(j)}, s^{(j)} \rangle, \tag{1.10}$$

where

$$\langle \bar{v}, \bar{s} | v, s \rangle = \int \sum_{r=1}^{k(z)} \bar{v}_{A'r}^{\dagger} v_r^{A'} \, dz + \sum_{i \in I} \bar{s}_i^{\dagger} s_i \tag{1.11}$$

is the inner product of the Hilbert space V. The conditions which ensure the nonsingularity and self-duality of the connection A_μ are as before that the operator $\Delta^\dagger \Delta$ should be nonsingular and commute with the quaternions. The former statement is easily verified for the operator Δ of (1.9); the second is true if and only if

$$T^{m\dagger} = -T^m \tag{1.12}$$

and

$$\frac{dT^m}{dz} = \epsilon^{mpq} T^p T^q + \sum_{i \in I} \alpha_i^m \delta(z - z_i), \tag{1.13}$$

where

$$\alpha_i e_\mu = a_i^\dagger a_i,$$

that is,

$$\alpha_{ir}^{\mu\, s} e_\mu^{B'A} = a_{ir}^{\dagger B'} a_i^{sA}. \tag{1.14}$$

The relations (1.13) are known as *Nahm's equations*; between the "jumping points" z_i they take the form

$$\frac{dT^m}{dz} = \epsilon^{mpq} T^p T^q. \tag{1.15}$$

If the embeddings $v(x)$ are normalized by $v(x)^\dagger v(x) = 1$, then the formulas (1.1) and (1.10) automatically give rise to U(n) potentials. By imposing suitable constraints on Δ as described elsewhere[8] we may construct O(n) or USp(n) gauge fields.

II. THE TRANSITION MATRIX FOR NAHM'S EQUATIONS

In this section we shall consider in detail the equations

$$\frac{dT_i}{dz} = \epsilon_{ijk} T_j T_k \tag{2.1}$$

for the three $k \times k$ skew-Hermitian matrices T_i. Let $T_\mu(x)$ be a U(k) gauge field dependent only on one variable, say x^0, and such that $T_0 = 0$. It is easy to see[9] that the corresponding curvature is self-dual, that is,

$$F_{oi} = \tfrac{1}{2} \epsilon_{ijk} F_{jk} \tag{2.2}$$

if and only if the functions $T_i(x^0)$ satisfy

$$\frac{dT_i}{dx^0} = \frac{1}{2} \epsilon_{ijk} [T_j, T_k]. \tag{2.3}$$

But these equations are identical to Nahm's equations (2.1) with z relabeled as x^0.

Now recall that to each self-dual connection there corresponds a certain holomorphic vector bundle via the following construction due to Ward.[1,2] Let $T_\mu(x)$ be analytically continued to $\mathbb{C}^4$, and the points of $\mathbb{C}^4$ be represented in quaternionic notation

$$x = x^0 - i\mathbf{x}\cdot\boldsymbol{\sigma} = \begin{pmatrix} x^0 - ix^3 & -x^2 - ix^1 \\ x^2 - ix & x^0 + ix^3 \end{pmatrix} \equiv \begin{pmatrix} y & -\bar{z} \\ z & \bar{y} \end{pmatrix}. \tag{2.4}$$

There are two classes of null planes in $\mathbb{C}^4$, which may be distinguished by means of the skew-symmetric tensor $G^{\mu\nu} = V^\mu W^\nu - W^\mu V^\nu$ (unique up to a factor) formed from two independent vectors V, W in the plane. In one class of null planes $G^{\mu\nu}$ is self-dual; in the other class it is anti-self-dual. An anti-self-dual plane θ is described by an equation of the form

$$x\pi = \omega, \tag{2.5}$$

where x is the quaternion of (2.4) and π, ω are two-dimensional complex column vectors and $\pi \neq 0$. Let us write $\mu = \omega_2/\pi_2$, $\nu = \omega_1/\pi_1$, $\zeta = \pi_1/\pi_2$; in terms of these parameters the plane θ is given by the equations

$$\bar{y} + z\zeta = \mu, \quad y - \bar{z}/\zeta = \nu. \tag{2.6}$$

From (2.5) we see that the pairs (π,ω), (π',ω') describe the same plane if and only if $\pi' = \lambda\pi$, $\omega' = \lambda\omega$ for some $\lambda \in \mathbb{C} \setminus \{0\}$. Thus the set of anti-self-dual null planes is the set of equivalence classes of pairs (π,ω) under this relation, namely $\mathbb{CP}^3 \setminus \mathbb{CP}^1$ (the deleted $\mathbb{CP}^1$ consists of those pairs in which $\pi = 0$). We give the set of planes the standard differentiable structure of $\mathbb{CP}^3$. Since $\pi \neq 0$ for each plane, the manifold $\mathbb{CP}^3 \setminus \mathbb{CP}^1$ can be covered by two coordinate patches $U_1 = \{\theta = (\pi,\omega): \pi_1 \neq 0\}$ and $U_2 = \{\theta = (\pi,\omega): \pi_2 \neq 0\}$. The standard coordinates in these regions are in U_1

$$\pi_2/\pi_1 = 1/\zeta, \quad \omega_1/\pi_1 = \nu, \quad \omega_2/\pi_1 = \mu/\zeta;$$

in U_2

$$\pi_1/\pi_2 = \zeta, \quad \omega_1/\pi_2 = \nu\zeta, \quad \omega_2/\pi_2 = \mu. \tag{2.7}$$

Let us choose a point $x_1(\theta)$ on each $\theta \in U_1$, and a point $x_2(\theta)$ on each $\theta \in U_2$. The restriction of a self-dual two form $F_{\mu\nu}$ to an anti-self-dual plane θ vanishes identically, so that the connection $T_\mu(x)$ is *flat* when restricted to θ. We may therefore define a vector bundle over $\mathbb{CP}^3 \setminus \mathbb{CP}^1$ whose fiber at θ consists of all covariant constants on θ. To any covariant constant f on the plane θ we may ascribe coordinates $f(x_1)$ if $\theta \in U_1$, or $f(x_2)$ if $\theta \in U_2$. If $\theta \in U_1 \cap U_2$ there are two sets of coordinates for f, related by a *transition matrix* $g(\theta)$:

$$f(x_1) = g(\theta) f(x_2), \tag{2.8}$$

where

$$g(\theta) = P \exp \int_{x_1}^{x_2} T_\mu(x) dx^\mu \tag{2.9}$$

and the integration is performed along any path from x_1 to x_2 in the plane θ. In principle, knowledge of $g(\theta)$ is sufficient to determine $T_\mu(x)$ up to a gauge transformation. For

$$g(\theta) = P \exp \int_{x_1}^{x} T_\mu \, dx^\mu \, P \exp \int_{x}^{x_2} T_\mu \, dx^\mu$$
$$= h(x,\zeta) k(x,\zeta)^{-1}, \tag{2.10}$$

where h is analytic for $\zeta \neq 0$ and k is analytic for $\zeta \neq \infty$. Liouville's theorem implies that h,k are uniquely determined by $g(\theta)$ up to transformations of the form $k(x,\zeta) \to k(x,\zeta)\,\gamma(x)$, $h(x,\zeta) \to h(x,\zeta)\,\gamma(x)$. But then for any vector V^μ tangent to the plane $\theta(x,\zeta)$,

$$\begin{aligned} V^\mu \partial_\mu h(x,\zeta) &= h(x,\zeta) V^\mu T_\mu(x), \\ V^\mu \partial_\mu k(x,\zeta) &= k(x,\zeta) V^\mu T_\mu(x). \end{aligned} \tag{2.11}$$

Now the tangent vectors of the plane $\theta(x,\zeta)$ are those which satisfy $V^\mu \partial_\mu \mu = V^\mu \partial_\mu \nu = 0$; recalling the expressions (2.6) for μ,ν we see that the tangent plane is spanned by the vectors $V^\mu \partial_\mu = \partial_y + \zeta \partial_{\bar z}$, $W^\mu \partial_\mu = \partial_z - \zeta \partial_{\bar y}$. Thus Eqs. (2.11) tell us that

$$T_y + \zeta T_{\bar z} = h^{-1}(\partial_y + \zeta \partial_{\bar z})h = k^{-1}(\partial_y + \zeta \partial_{\bar z})k, \tag{2.12}$$
$$T_z - \zeta T_{\bar y} = h^{-1}(\partial_z - \zeta \partial_{\bar y})h = k^{-1}(\partial_z - \zeta \partial_{\bar y})k.$$

The arbitrary transformation $k \to k\gamma$, $h \to h\gamma$, which we are allowed to perform without change in $g(\theta)$, gives rise to a gauge transformation $T_\mu \to \gamma^{-1} T_\mu + \gamma^{-1}\partial_\mu \gamma$.

Let us return to the situation described at the beginning of this section, in which T_μ depends on the single variable x^0, and $T_0 = 0$. It is possible to choose the two reference points x_1, x_2 so that they both lie on some given hyperplane $x^0 = \lambda$. Specifically, we may take

$$x_1 = \begin{pmatrix} \nu & 0 \\ (\mu+\nu-2\lambda)/\zeta & -\nu+2\lambda \end{pmatrix},$$
$$x_2 = \begin{pmatrix} -\mu+2\lambda & (\mu+\nu-2\lambda)\zeta \\ 0 & \mu \end{pmatrix}. \tag{2.13}$$

Indeed, it is easy to see that x_i is an analytic function of the coordinates (2.7) in the region U_i; in both cases $x^0 = (y+\bar y)/2 = \lambda$. For the path of integration in (2.9) we may choose a straight line lying within $x^0 = \lambda$; T_μ is constant along this path. We are now able to perform the integration

$$g(\lambda,\theta) = P\exp\int_{x_1}^{x_2} T_\mu\, dx^\mu = \exp T_\mu(\lambda)(x_2 - x_1)^\mu. \tag{2.14}$$

Let us make the following definition:

$$\tau(\lambda,\zeta) = T_z(\lambda) + (T_y(\lambda) - T_{\bar y}(\lambda))\zeta + T_{\bar z}(\lambda)\zeta^2$$
$$= \tfrac{1}{2}\{i(1-\zeta^2)T_1(\lambda) + (1+\zeta^2)T_2(\lambda) + 2i\zeta T_3(\lambda)\}. \tag{2.15}$$

Then $g(\lambda,\theta)$ is related to the "initial conditions" on T_i at $x^0 = \lambda$ by the following result:

$$g(\lambda,\theta) = \exp -(\tau(\lambda,\zeta)/\zeta)(\mu+\nu-2\lambda). \tag{2.16}$$

In the next section we shall consider some special cases of (2.16), and work through an example in which this result can be used to generate a solution to Nahm's equations.

III. EXAMPLES AND APPLICATIONS

The simplest nontrivial example of Nahm's equations arises when we consider a pair of SU(2) monopoles. In this case Eqs. (1.6) and (1.8) have $z_1 = -\frac{1}{2}$, $z_2 = \frac{1}{2}$, $k_1 = -2$, $k_2 = 2$; hence $k(z) = 2$ if $-\frac{1}{2} < z < \frac{1}{2}$, but $k(z) = 0$ otherwise. This means that the $T_i(z)$ are 2×2 skew-Hermitian matrices defined on the interval $(-\frac{1}{2}, \frac{1}{2})$. According to Nahm, T_i satisfies the boundary condition that it has *simple poles* at the end points. In order that the resultant $A_\mu(x)$ be an SU(2) potential, $T_i(z)$ must also satisfy a reality condition[8]

$$T_i(-z)^* = -cT_i(z)c^\dagger, \tag{3.1}$$

where the "charge conjugation" matrix c is chosen so that $cc^* = 1$. For our purposes it is convenient to choose a basis in which $c = \sigma_1$. A known solution of Nahm's equations with all the above properties is[11]

$$T_1(x^0) = -(i/2)\sigma_1(q/\mathrm{cn}\,px^0),$$
$$T_2(x^0) = -(i/2)\sigma_2(p\,\mathrm{dn}\,px^0/\mathrm{cn}\,px^0), \tag{3.2}$$
$$T_3(x^0) = -(i/2)\sigma_3(q\,\mathrm{sn}\,px^0/\mathrm{cn}\,px^0),$$

where $q = p\sqrt{1-k^2}$, k is the modulus of the elliptic functions, and in order to place the poles at $x^0 = \pm\frac{1}{2}$ we must set $p = 2K(k)$, where $K(k)$ is the smallest positive zero of cn θ. For this solution, the initial conditions at $x^0 = 0$ are

$$T_1(0) = -\tfrac{1}{2}i\sigma_1 q, \quad T_2(0) = -\tfrac{1}{2}i\sigma_2 p, \quad T_3(0) = 0. \tag{3.3}$$

We apply Eqs. (2.15) and (2.16) to calculate $g(\theta) \equiv g(0,\theta)$:

$$\tau(\zeta) = -(i/4)\{i(1-\zeta^2)q\sigma_1 + (1+\zeta^2)p\sigma_2\}. \tag{3.4}$$

Hence

$$g(\theta) = \exp(i/4\zeta)\{i(1-\zeta^2)q\sigma_1 + (1+\zeta^2)p\sigma_2\}(\mu+\nu). \tag{3.5}$$

To evaluate the exponential in this case we use the well-known identity $\exp i\alpha\mathbf{n}\cdot\boldsymbol{\sigma} = \cos\alpha + i\mathbf{n}\cdot\boldsymbol{\sigma}\sin\alpha$, where $\mathbf{n}$ is a unit vector. From (3.5) we read off the values of α and $\mathbf{n}$:

$$\alpha = (\sqrt{ab}/\zeta)(\mu+\nu), \tag{3.6}$$
$$\mathbf{n} = (1/2\sqrt{ab})(i(a-b), a+b, 0), \tag{3.7}$$

where the quadratic polynomials a and b are defined by

$$4a = p(1+\zeta^2) + q(1-\zeta^2), \quad 4b = p(1+\zeta^2) - q(1-\zeta^2). \tag{3.8}$$

Thus if $\sigma_\pm = \sigma_1 \pm i\sigma_2$, the explicit form for the transition matrix is

$$g(\theta) = \cos\frac{\sqrt{ab}}{\zeta}(\mu+\nu) + \frac{1}{2\sqrt{ab}}(b\sigma_+ - a\sigma_-)\sin\frac{\sqrt{ab}}{\zeta}(\mu+\nu)$$
$$= \begin{bmatrix} \cos\frac{\sqrt{ab}}{\zeta}(\mu+\nu) & \frac{b}{\sqrt{ab}}\sin\frac{\sqrt{ab}}{\zeta}(\mu+\nu) \\ -\frac{a}{\sqrt{ab}}\sin\frac{\sqrt{ab}}{\zeta}(\mu+\nu) & \cos\frac{\sqrt{ab}}{\zeta}(\mu+\nu) \end{bmatrix}. \tag{3.9}$$

It is possible to deduce the form (3.9) of $g(\theta)$ from a slightly different point of view. We are seeking a transition matrix $g(\theta)$ corresponding to a gauge field $T_\mu(x)$ dependent on x^0 alone. A gauge-equivalent field $\overline{T}_\mu(x)$ will not, in general, have the same property, but will satisfy the weaker condition: any translation of $\overline{T}_\mu(x)$ in x^1, x^2, or x^3 may be compensated by a gauge transformation. That is, if $a^0 = 0$ there exists a gauge transformation $\gamma_a(x)$ such that

$$\overline{T}_\mu(x+a) = \gamma_a(x)\overline{T}_\mu(x)\gamma_a(x)^{-1} - \partial_\mu\gamma_a(x)\gamma_a(x)^{-1}. \tag{3.10}$$

By contrast, the monopole gauge potentials $A_\mu(x)$ themselves are independent of x^0, and any equivalent potentials $\overline{A}_m(x)$ must satisfy an invariance condition analogous to (3.10) for translations $a = (a^0,0,0,0)$ in x^0. Guided by this precedent, we say that two transition matrices are equivalent if there exists a matrix $X(\theta)$ analytic in U_1 and a matrix $Y(\theta)$ analytic in U_2 [see definitions preceding (2.7)] such that

$$\bar g(\theta) = X(\theta)g(\theta)Y(\theta). \tag{3.11}$$

It is easy to prove that gauge-equivalence classes of vector potentials are in 1–1 correspondence with equivalence classes of transition matrices via the Atiyah–Ward construction. Therefore the property of translational invariance (3.10) is reflected in a similar property of the corresponding $g(\theta)$:

$$g(x+a,\zeta) = X_a(x,\zeta)g(x,\zeta)Y_a(x,\zeta). \tag{3.12}$$

The problem of finding the transition matrices for all solutions of Nahm's equations is equivalent to the problem of finding the most general matrix satisfying (3.12). Let us show that (3.9) indeed satisfies this constraint. Under an infinitesimal translation in x^1, x^2, and x^3 the angle α defined in (3.6) is changed by an amount

$$\delta\alpha = (\sqrt{ab}/\zeta)(\delta z\zeta - \delta\bar{z}/\zeta). \tag{3.13}$$

It follows that the increment in $g(\theta)$ is given by

$$\begin{aligned}\delta g(x,\zeta) &= \left(\delta z - \frac{\delta\bar{z}}{\zeta^2}\right)\begin{pmatrix} -\sqrt{ab}\sin\alpha & b\cos\alpha \\ -a\cos\alpha & -\sqrt{ab}\sin\alpha\end{pmatrix}\\ &= \delta\bar{z}\xi(x,\zeta)g(x,\zeta) + \delta z g(x,\zeta)z(x,\zeta),\end{aligned} \tag{3.14}$$

where the matrices ξ, η have been defined as

$$\begin{aligned}\xi(x,\zeta) &= -\frac{1}{\zeta^2}\begin{pmatrix} 0 & b \\ -a & 0\end{pmatrix},\\ \eta(x,\zeta) &= \begin{pmatrix} 0 & b \\ -a & 0\end{pmatrix}.\end{aligned} \tag{3.15}$$

It is easy to see that ξ, η are analytic for $\zeta \neq 0$, $\zeta \neq \infty$, respectively; (3.14) is just the infinitesimal form of the condition of translational invariance (3.12).

In Sec. II we calculated $g(\theta)$ from the initial values $T_i(\lambda)$; in some cases it is possible to go further, and derive the solution $T_i(x^0)$ itself. If $g(\theta)$ may be reduced to triangular form by an equivalence of type (3.11), then we may apply the usual *Ansätze*[2–4] to extract $T_\mu(x)$. One example for which this is possible is that of the axially symmetric charge 2 monopole.[12] In this case the quantities appearing in (3.2) and (3.8) are $k=0$, $q=p=\pi$, $a=\pi/2$, $b=\pi\zeta^2/2$, and $\sqrt{ab}=\pi\zeta/2$. Hence

$$g(\theta) = \begin{pmatrix}\cos\pi(\mu+\nu)/2 & \zeta\sin\pi(\mu+\nu)/2 \\ -\zeta^{-1}\sin\pi(\mu+\nu)/2 & \cos\pi(\mu+\nu)/2\end{pmatrix}. \tag{3.16}$$

In order to put g in upper-triangular form we perform the following sequence of equivalence transformations:

$$\begin{aligned}g(\theta) &= \begin{pmatrix}\cos\alpha & \zeta\sin\alpha \\ -\zeta^{-1}\sin\alpha & \cos\alpha\end{pmatrix}\\ &= \begin{pmatrix}1 & 0 \\ i\zeta^{-1} & 1\end{pmatrix}\begin{pmatrix}\cos\alpha & -i\zeta e^{i\alpha} \\ -i\zeta^{-1}e^{i\alpha} & 0\end{pmatrix}\begin{pmatrix}1 & i\zeta \\ 0 & 1\end{pmatrix}\\ &= \begin{pmatrix}1 & 0 \\ i\zeta^{-1} & 1\end{pmatrix}\begin{pmatrix}e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4}\end{pmatrix}\\ &\quad\times\begin{pmatrix}\zeta e^{i\alpha} & \cos\alpha \\ 0 & \zeta^{-1}e^{-i\alpha}\end{pmatrix}\begin{pmatrix}0 & -e^{i\pi/4} \\ e^{-i\pi/4} & 0\end{pmatrix}\begin{pmatrix}1 & i\zeta \\ 0 & 1\end{pmatrix}.\end{aligned} \tag{3.17}$$

Now let $\alpha = \alpha_+ + \alpha_-$, where α_+ is analytic for $\zeta \neq \infty$ and α_- is analytic for $\zeta \neq 0$; in this case $\alpha_+ = \pi\mu/2$, $\alpha_- = \pi\nu/2$. Then there is a further equivalence

$$\begin{aligned}&\begin{pmatrix}\zeta e^{i\alpha} & \cos\alpha \\ 0 & \zeta^{-1}e^{-i\alpha}\end{pmatrix}\\ &\quad= \begin{pmatrix}e^{i\alpha_-} & 0 \\ 0 & e^{-i\alpha_-}\end{pmatrix}\begin{pmatrix}\zeta & e^{i(\alpha_+ - \alpha_-)}\cos\alpha \\ 0 & \zeta^{-1}\end{pmatrix}\\ &\quad\times\begin{pmatrix}e^{i\alpha_+} & 0 \\ 0 & e^{-i\alpha_+}\end{pmatrix}.\end{aligned} \tag{3.18}$$

Hence $g(\theta)$ is equivalent to the upper-triangular matrix

$$\bar{g}(\theta) = \begin{pmatrix}\zeta & \frac{1}{2}(e^{i\pi\mu} + e^{-i\pi\nu}) \\ 0 & \zeta^{-1}\end{pmatrix}. \tag{3.19}$$

We can now recover the vector field $\overline{T}_\mu(x)$ corresponding to $g(\theta)$ as follows: Let $\rho(x,\zeta) = \frac{1}{2}(e^{i\pi\mu} + e^{-i\pi\nu})$, and let $\Delta_0(x)$ be the coefficient of ζ^0 in the Laurent expansion of ρ:

$$\Delta_0(x) = \frac{1}{2\pi i}\oint\frac{d\zeta}{\zeta}\rho(x,\zeta), \tag{3.20}$$

the integral being evaluated around some contour C encircling the origin in a positive direction. The matrix (3.19) is reminiscent of that encountered when applying the twistor method directly to the gauge field A_μ of a single monopole. Using the first *Ansatz*[13]

$$\overline{T}_\mu(x) = (i/2)\eta_{\mu\nu}\,\partial_\nu \ln\Delta_0, \tag{3.21}$$

where

$$\eta_{\mu\nu} = -\eta_{\nu\mu}, \quad \eta_{io} = \sigma_i, \quad \eta_{ij} = \epsilon_{ijk}\sigma_k, \tag{3.22}$$

and making the appropriate substitution

$$\Delta_0(x) = e^{-\pi x^3}\cos\pi x^0, \tag{3.23}$$

we find that the components of $\overline{T}_\mu(x)$ are as follows:

$$\begin{aligned}\overline{T}_0 &= \frac{i}{2}\pi\sigma_3, \quad \overline{T}_1 = \frac{i}{2}\pi\sigma_2 - \frac{i}{2}\pi\sigma_1\tan\pi x^0,\\ \overline{T}_2 &= -\frac{i}{2}\pi\sigma_1 - \frac{i}{2}\pi\sigma_2\tan\pi x^0,\\ \overline{T}_3 &= -\frac{i}{2}\pi\sigma_3\tan\pi x^0.\end{aligned} \tag{3.24}$$

This is certainly a self-dual gauge field; but to obtain a solution of Nahm's equations (2.1) we must make a gauge transformation $T_\mu = \gamma\overline{T}_\mu\gamma^{-1} - \partial_\mu\gamma\gamma^{-1}$ such that T_μ satisfies the additional constraints $T_0 = 0$, $\partial_i T_\mu = 0$. The group element $\gamma(x)$ must therefore have the property $\partial_0\gamma = \gamma\overline{T}_0$; in this case an appropriate choice is

$$\gamma(x) = \exp -(i/2)\sigma_3\pi(\tfrac{1}{2} - x^0). \tag{3.25}$$

Now the adjoint representative of $\gamma(x)$ is a rotation through an angle $\pi(\frac{1}{2} - x^0)$ in the $\sigma_1\sigma_2$ plane. It follows that the resulting solution of (2.1) is

$$\begin{aligned}T_1 &= -\frac{i}{2}\pi\frac{\sigma_1}{\cos\pi x^0}, \quad T_2 = -\frac{i}{2}\pi\frac{\sigma_2}{\cos\pi x^0},\\ T_3 &= -(i/2)\pi\sigma_3\tan\pi x^0,\end{aligned} \tag{3.26}$$

which is just a special case of the known solution (3.2) in which $k=0$.

As a second example, we shall solve (2.1) subject to the boundary conditions $T_i(1) = \frac{1}{2}i\sigma_i$. In this case (2.15) and (2.16) are used with $\lambda = 1$, and $\tau(1,\zeta)^2 = 0$. Therefore

$$\begin{aligned}g(1,\theta) &= 1 - [\tau(1,\zeta)/\zeta](\mu+\nu-2)\\ &= \begin{pmatrix}(\mu+\nu)/2 & -\zeta(\mu+\nu-2)/2 \\ \zeta^{-1}(\mu+\nu-2)/2 & 2-(\mu+\nu)/2\end{pmatrix}.\end{aligned} \tag{3.27}$$

The triangular matrix equivalent to (3.27) is

$$\bar{g}(1,\theta)=\begin{pmatrix}\zeta & \frac{1}{2}(\mu+\nu)\\ 0 & \zeta^{-1}\end{pmatrix}. \tag{3.28}$$

Using the first *Ansatz* as above, with the function $\Delta_0(x)=\frac{1}{2}(y+\bar{y})=x^0$, we find the gauge field

$$T_i(x)=(i/2)(\sigma_i/x^0),\quad T_0(x)=0. \tag{3.29}$$

This field already satisfies the conditions $T_0=0$, $\partial_i T_\mu=0$, so no gauge transformation is necessary. It represents a solution of Nahm's equations (2.1) having a *single* pole (at $x^0=0$). Such solutions become relevant when one considers gauge groups larger than SU(2).

It is easy to see that if T_μ and $\gamma T_\mu \gamma^{-1}-\partial_\mu \gamma\gamma^{-1}$ are two gauge-equivalent solutions of Nahm's equations [that is, both satisfy (2.1) and $T_0=\partial_i T_\mu=0$] then the gauge transformation γ must be constant. Therefore each equivalence class of fields with the property (3.10) contains a solution of Nahm's equations which is unique up to a global transformation $T_\mu \to \gamma T_\mu \gamma^{-1}$.

IV. CONCLUSION

By remarking that Nahm's equations are the self-duality conditions for a one-dimensional gauge field and applying the twistor construction we have shown the equivalence of these equations to the following *Riemann–Hilbert problem* (RHP). Let C be a contour encircling the origin in the ζ-plane, and let D_+,D_- be the interior and exterior domains separated by C; the matrix function $g(x,\zeta)$ of Eq. (2.16) is defined and analytic in a neighborhood of C. The RHP to be solved is to find two matrices $h(x,\zeta)$, $k(x,\zeta)$ analytic in D_-,D_+ respectively, such that $g=hk^{-1}$ on C. For simplicity we impose the boundary condition $h(\infty)=I$.

If we represent the functions h,k as

$$h(\lambda)=I+\frac{1}{2\pi i}\oint\frac{\sigma(\zeta)}{\zeta-\lambda}\,d\zeta,\quad \lambda\in D_-,$$

$$k(\lambda)=I+\frac{1}{2\pi i}\oint\frac{\sigma(\zeta)}{\zeta-\lambda}\,d\zeta,\quad \lambda\in D_+, \tag{4.1}$$

then the function $\sigma(\zeta)$ satisfies a linear singular integral equation

$$\sigma(\lambda)+2[g(\lambda)+I]^{-1}[g(\lambda)-I]\times\left(I+\frac{1}{2\pi i}P\oint\frac{\sigma(\zeta)}{\zeta-\lambda}\,d\zeta\right)=0, \tag{4.2}$$

where P denotes the principal value of the contour integral.

We have seen that if $g(\theta)$ can be put into triangular form by an equivalence transformation then the Riemann–Hilbert problem may be solved by a standard *Ansatz*.[2–4] However, we have not been able to find a triangular form for the general transition matrix (2.16), nor indeed for the general two-monopole situation of Sec. III in which $k\neq 0$. It is possible to argue that no such standard form exists in this case; for suppose that $g(\theta)$ is equivalent to

$$\bar{g}(\theta)=\begin{pmatrix}\zeta^l & \rho(x,\zeta)\\ 0 & \zeta^{-l}\end{pmatrix} \tag{4.3}$$

for some positive integer l and some $\rho(x,\zeta)$ analytic for $\zeta\notin\{0,\infty\}$. Suppose also that $\bar{g}(\theta)$ obeys the condition of translational invariance with respect to x^i in its infinitesimal form

$$\frac{\partial}{\partial x^i}\begin{pmatrix}\zeta^l & \rho\\ 0 & \zeta^{-l}\end{pmatrix}=\begin{pmatrix}\alpha & \beta\\ 0 & \gamma\end{pmatrix}\begin{pmatrix}\zeta^l & \rho\\ 0 & \zeta^{-l}\end{pmatrix}+\begin{pmatrix}\zeta^l & \rho\\ 0 & \zeta^{-l}\end{pmatrix}\begin{pmatrix}a & b\\ 0 & c\end{pmatrix} \tag{4.4}$$

for some functions of x and ζ (depending on i) of which α,β,γ are analytic for $\zeta\neq 0$ and a,b,c are analytic for $\zeta\neq\infty$. It is straightforward to deduce that α,γ,a,c are independent of ζ and that

$$\partial_i\rho=(\alpha+c)\rho+\beta\zeta^{-l}+\zeta^l b. \tag{4.5}$$

If we now consider the Laurent coefficients

$$\Delta_r(x)\equiv\frac{1}{2\pi i}\oint\frac{d\zeta}{\zeta}\,\zeta^r\rho(x,\zeta),$$

we see that for $-l+1\leqslant r\leqslant l-1$

$$\partial_i\Delta_r=(\alpha+c)_i\Delta_r=k_i\Delta_r, \tag{4.6}$$

so that each Δ_r depends exponentially upon $\mathbf{x}$:

$$\Delta_r(x)=\Delta_r{}^0(x^0)\exp\mathbf{k}\cdot\mathbf{x}. \tag{4.7}$$

Moreover, the functions Δ_r satisfy the Laplace equation[3] $\partial^2\Delta_r=0$, which by virtue of (4.7) becomes

$$\Delta_r{}^{0\prime\prime}=-k^2\Delta_r{}^0. \tag{4.8}$$

Finally we obtain the result that the coefficients $\Delta_r(x)$ for $-l+1\leqslant r\leqslant l-1$ must take the form

$$\Delta_r(x)=(A_r\exp i|\mathbf{k}|x^0+B_r\exp -i|\mathbf{k}|x^0)\exp\mathbf{k}\cdot\mathbf{x}. \tag{4.9}$$

If these functions are substituted into the *Ansätze* of Refs. 2–4 they will yield potentials T_μ involving exponential and trigonometric functions; never elliptic functions. Hence the solution (3.2) cannot arise from a transition matrix such as (4.3).

Many questions remain to be answered in this approach. For example, it is clear that the solution $T_\mu(x)$ will be singular at points x where the RHP $g(x,\zeta)=h(x,\zeta)k(x,\zeta)^{-1}$ is insoluble, but the location of these points is not at all obvious from (2.16). If we could deduce from $g(x,\zeta)$ the location and nature of the singularities in $T_\mu(x)$, then we would be able to find the constraints on $T_\mu(0)$ which ensure Nahm's boundary condition (that T_μ has simple poles at the endpoints). Such constraints must be sufficient to reduce the number of parameters of an SU(2) multimonopole to the value $4n-1$ (see Ref. 14).

It is interesting to note that the characteristic equation of $\tau(\lambda,\zeta)$, $\det(\eta-\tau(\lambda,\zeta))=0$ (which is independent of λ) defines what Hitchin calls the *spectral curve*.[15] The reason for the appearance of the spectral curve in this approach is not clear to the author.

One of the most interesting features of Nahm's construction is the "reciprocity" of the relationship between the gauge potential A_μ and the matrices T_μ (see Ref. 9). This leads us to expect a similarly reciprocal relationship between the transition matrices for the two connections A_μ and T_μ.

ACKNOWLEDGMENTS

The author is grateful to E. Corrigan and S. Rouhani for helpful discussions. He would like to thank the Department of Applied Mathematics and Theoretical Physics and

Trinity College, Cambridge for their hospitality during the writing of this paper, and the U.K. Science and Engineering Research Council for a studentship.

[1]R. Ward, Phys. Lett. A **61**, 81 (1977).
[2]M. Atiyah and R. Ward, Commun. Math. Phys. **55**, 117 (1977).
[3]E. Corrigan, D. Fairlie, P. Goddard, and R. Yates, Commun. Math. Phys. **58**, 223 (1978).
[4]E. Corrigan and P. Goddard, Commun. Math. Phys. **80**, 575 (1981).
[5]M. Atiyah, C. Drinfield, N. Hitchin, and Yu. Manin, Phys. Lett. A **65**, 185 (1978).
[6]E. Corrigan, D. Fairlie, P. Goddard, and S. Templeton, Nucl. Phys. B **140**, 31 (1978); N. Christ, N. Stanton, and E. Weinberg, Phys. Rev. D **18**, 2013 (1978).
[7]W. Nahm, Phys. Lett. B **90**, 413 (1980).
[8]W. Nahm, "All self-dual multimonopoles for arbitrary gauge groups," in *Structural Elements in Particle Physics and Statistical Mechanics*, edited by J. Honerkamp, K. Pohlmeyer, and H. Römer (NATO, New York, 1981), p. 301; W. Nahm, "Multimonopoles in the ADHM construction," in *Proceedings of the Symposium on Particle Physics*, edited by Z. Horvath *et al.* (Visegrad, 1981), p. 397.; W. Nahm, "The construction of all self-dual multimonopoles by the ADHM method," in *Monopoles in Quantum Field Theory*, edited by N. Craigie, P. Goddard, and W. Nahm (World Scientific, Singapore, 1982), p. 87.
[9]E. Corrigan and P. Goddard, "Construction of instanton and monopole solutions and reciprocity," Ann. Phys. **154**(1), 253 (1984); W. Nahm, "Self-dual monopoles and calorons," Bonn preprint HE-83-16.
[10]N. Manton, Nucl. Phys. B **135**, 319 (1978).
[11]S. Brown, H. Panagopoulos, and M. Prasad, Phys. Rev. D **26**, 854 (1982).
[12]R. Ward, Commun. Math. Phys. **79**, 317 (1981); P. Forgacs, Z. Horvath, and L. Palla, Phys. Lett B **99**, 232 (1981).
[13]R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977); E. Corrigan and D. Fairlie, Phys. Lett. B **67**, 69 (1977).
[14]E. Weinberg, Phys. Rev. D **20**, 936 (1979).
[15]N. Hitchin, Commun. Math. Phys. **83**, 579 (1982); N. Hitchin, Commun. Math. Phys. **89**, 145 (1983).

Electrodynamics in eight-dimensional space

H. C. Chandola and B. S. Rajput
Department of Physics, Kumaun University, Naini Tal-263002, India

(Received 19 December 1983; accepted for publication 1 June 1984)

Generalized electromagnetic fields and the corresponding generalized electromagnetic field tensor in a real eight-dimensional space-time manifold have been constructed, in terms of which the generalized Maxwell's field equations have been derived. The Lorentz force acting on a charged particle moving in generalized electromagnetic fields has been derived by constructing an eight-velocity vector in this space.

I. INTRODUCTION

During the past few years there has been considerable interest in higher-dimensional kinematical models[1–8] for a proper and unified representation of every type of relativistic object, bradyonic as well as tachyonic (including those with internal structure). The Yang–Mills field, like the electromagnetic field, appears as part of the metric tensor for the higher-dimensional space as shown by Thirring[9] for the five-dimensional case (the gauge group being abelian) in which a gauge invariant effective Lagrangian for scalars and spinors, coupled to the electromagnetic field, can be derived from a free field Lagrangian. Furthermore, the Einstein–Yang–Mills Lagrangian can also be derived from a higher-dimensional gravity Lagrangian, as Mecklenburg[10] has used a seven-dimensional space with its internal dimensions spacelike in order to reproduce the right coupling of gravity to the Yang–Mills field. Quite generally, there is the hope that spontaneous symmetry breakdown may also be achieved by the use of higher dimensionality. In a similar way, it has been speculated[11–17] that the problem of representation and localization of tachyonic objects may be solved only via the use of a higher-dimensional space since they cannot be fully localized in ordinary four-dimensional Minkowski space. In an attempt to achieve a unified and consistent representation of both bradyonic and tachyonic objects (extended one, i.e., with internal structure also), we have constructed[18] an eight-dimensional space-time structure (D^8) as the union of two different subspaces $R^4(\vec{r},t)$ and $T^4(\vec{t},r)$ which act as the most natural spaces for the individual representations of bradyonic and tachyonic objects, respectively. This new eight-dimensional space-time geometry can be represented as the topological summation of R^4 and T^4 subspaces, such that

$$D^8 \equiv R^4 \cup T^4.$$

While passing from R^4 space to T^4 space, the timelike event is observed as a spacelike one and the usual bradyons appear to behave as tachyons and vice versa. Thus, a subluminal speed in R^4 subspace leads to a superluminal speed in T^4 subspace and consequently, the bradyons move with subluminal velocity in subspace R^4 and superluminal velocity in T^4 subspace in such a way that the internal and external spaces for bradyons are designated as T^4 and R^4 subspaces, respectively, while the situation is just reversed for tachyonic objects.

Furthermore, it has already been demonstrated in our earlier paper[19] that in eight-dimensional space-time one gets two types of Doppler effect of sub- and superluminal sources, the usual one associated with their appropriate external subspaces and the other one assoicated with their internal subspaces which is pronounced as a hidden Doppler effect since it is associated with the internal degrees of freedom and internal symmetry properties of different objects. It has also been shown that if a macroscopic phenomena is known to produce a radio emission in bradyonic (or tachyonic) external subspace obeying a certain chronological law and one happens to detect the reversed radio emission, it should correspond to a hidden Doppler effect associated with the corresponding internal subspaces. Moreover, the built-in duality associated with the combination of symmetries of R^4 and T^4 subspaces may be useful to understand the problem of quark confinement in quantum chromodynamics where the role of tricolors would be played by three times for bradyons and three space coordinates for tachyons. In the present paper, we have formulated the unified electrodynamics in this extended Minkowski space (real eight-space geometry) and the field equations have been derived in terms of an eight-dimensional electromagnetic field tensor and eight-current densities. The Lorentz force on a charged particle moving in these generalized electromagnetic fields has also been constructed in terms of generalized field tensor and eight-velocity of the particle.

II. GENERALIZED ELECTROMAGNETISM IN EIGHT-SPACE

As mentioned earlier, the eight-dimensional space-time manifold is the union of two subspaces $R^4(\vec{r},t)$ and $T^4(\vec{t},r)$, in which a space-time eight-vector is defined as the following column vector (in natural units $\hbar = c = 1$):

$$\{x^\rho\} \equiv [\{r^\mu\};\quad \{t^\mu\}]^T \equiv [(\vec{r},t);(\vec{t},r)]^T, \tag{1}$$

where $\mu = 1,...,4, \rho = 1,...,8$, and T denotes the transpose. In a similar way, the eight-potential vector may be defined by the following column vector:

$$\{A^\rho\} \equiv [\{A^\mu\};\quad \{\phi^\mu\}]^T \equiv [(\vec{A},\phi);(\vec{\phi},A)]^T, \tag{2}$$

while the eight-dimensional differential operator is given by

$$\{\partial_\rho\} \equiv [\{(\partial_r)_\mu\};\quad \{(\partial_t)_\mu\}]^T \equiv [(\vec{\partial}_r,\partial_t);(\vec{\partial}_t,\partial_r)]^T. \tag{3}$$

The four-vector $\{A^\mu\}$ generates the electromagnetic fields in R^4, and the same fields when viewed upon from T^4 appear as if generated by the second four-vector $\{\phi^\mu\}$ in T^4 subspace. As such, the electromagnetic fields in R^4 and T^4 subspaces may be defined by the following relations:

$$\vec{E}_r = -\vec{\partial}_r\phi - \partial_t\vec{A}, \quad \vec{H}_r = \vec{\partial}_r\times\vec{A} \quad (\text{in } R^4), \tag{4a}$$

and

$$\vec{E}_t = -\vec{\partial}_t A - \partial_r\vec{\phi}, \quad \vec{H}_t = \vec{\partial}_t\times\vec{\phi} \quad (\text{in } T^4), \tag{4b}$$

which show that the space rotation of the spatial part of $\{\phi^{\mu}\}$ in T^4 produces the magnetic field $\vec{H}_t$ in this subspace in a similar manner as the space rotation of the spatial part of $\{A^{\mu}\}$ in R^4 produces the magnetic field $\vec{H}_r$ in this subspace. In order to construct the field equations taking into account both these types of fields as observed in R^4 and T^4 subspaces separately, we introduce the eight-current density vector as follows:

$$\{J^{\rho}\} \equiv [\{j^{\mu}\}; \ \{\zeta^{\mu}\}]^T \equiv [(\vec{J},\zeta);(\vec{\zeta},J)]^T, \tag{5}$$

where the four-vector $\{J^{\mu}\}$ provides the prescribed four-current density for fields $\vec{E}_r$ and $\vec{H}_r$ in subspace R^4, and $\{\zeta^{\mu}\}$ does so for fields $\vec{E}_t$ and $\vec{H}_t$ in T^4 subspace. Then, the field equations in subspace R^4 may be written as follows:

$$\begin{aligned} &\vec{\partial}_r\cdot\vec{E}_r = \zeta, \quad \vec{\partial}_r\times\vec{H}_r - \partial_t\vec{E}_r = \vec{J}, \\ &\vec{\partial}_r\cdot\vec{H}_r = 0, \quad \vec{\partial}_r\times\vec{E}_r + \partial_t\vec{H}_r = 0, \end{aligned} \tag{6}$$

and those in subspace T^4 are given by

$$\begin{aligned} &\vec{\partial}_t\cdot\vec{E}_t = -J, \quad \vec{\partial}_t\times\vec{H}_t - \partial_r\vec{E}_t = -\vec{\zeta}, \\ &\vec{\partial}_t\cdot\vec{H}_t = 0, \quad \vec{\partial}_t\times\vec{E}_t + \partial_r\vec{H}_t = 0. \end{aligned} \tag{7}$$

In physical terms, four-current vector $\{\zeta^{\mu}\}$ gives the charge and current source densities in subspace T^4 which acts as internal (or external) space for bradyons and subluminal fields (or tachyons and superluminal fields). Four-current density $\{J^{\mu}\}$ producing electromagnetic fields $\vec{E}_r$ and $\vec{H}_r$ in subspace R^4 when viewed upon from subspace T^4, appears as four-current $\{\zeta^{\mu}\}$ producing the electromagnetic fields $\vec{E}_t$ and $\vec{H}_t$ in subspace T^4. The electromagnetic field tensors in these two subspaces, may easily be constructed in the following form, by using the above definitions of fields given by Eqs. (4a) and (4b):

$$\xi_{\mu\nu} \equiv (\partial_r)_\nu A_\mu - (\partial_r)_\mu A_\nu = \begin{pmatrix} 0 & H_z & -H_y & E_x \\ -H_z & 0 & H_x & E_y \\ H_y & -H_x & 0 & E_z \\ -E_x & -E_y & -E_z & 0 \end{pmatrix} \quad (\text{in } R^4), \tag{8}$$

and

$$\mathscr{H}_{\mu\nu} \equiv (\partial_t)_\nu \phi_\mu - (\partial_t)_\mu \phi_\nu = \begin{pmatrix} 0 & H_{t_z} & -H_{t_y} & E_{t_x} \\ -H_{t_z} & 0 & H_{t_x} & E_{t_y} \\ H_{t_y} & -H_{t_x} & 0 & E_{t_z} \\ -E_{t_x} & -E_{t_y} & -E_{t_z} & 0 \end{pmatrix} \quad (\text{in } T^4). \tag{9}$$

In terms of these field tensors, the field equations (6) and (7) may be expressed as follows:

$$\partial_r^{\nu}\xi_{\mu\nu} = J_\mu, \quad \partial_r^{\nu}(\xi^d_{\mu\nu}) = 0 \quad (\text{in } R^4), \tag{10a}$$

$$\partial_t^{\nu}\mathscr{H}_{\mu\nu} = \zeta_\mu, \quad \partial_t^{\nu}(\mathscr{H}^d_{\mu\nu}) = 0 \quad (\text{in } T^4). \tag{10b}$$

The negative sign occurring with source densities in the second set of these equations may be explained in the light of results of Mignani and Recami[14] and Dattoli and Mignani,[15] where it has been shown that imaginary superluminal Lorentz transformations when applied to usual (subluminal) Maxwell's equations introduce such negative sign in four-current density. As such, by comparing the general field equations (10a) and (10b), we may conclude that electromagnetic fields of subspace T^4 behave as superluminal fields when observed from subspace R^4.

Using the field tensors, defined by Eqs. (8) and (9), the generalized electromagnetic field tensor in eight-dimensional space may be constructed as the following 8×8 matrix:

$$\hat{F}_{\rho\lambda} \equiv \begin{pmatrix} \hat{\xi}_{\mu\nu} & \hat{0} \\ \hat{0} & -\hat{\mathscr{H}}_{kl} \end{pmatrix}$$

$$(\mu,\nu = 1,\ldots,4; \ k = \mu + 4, \ l = \nu + 4), \tag{11}$$

where $\hat{\xi}_{\mu\nu}$, $\hat{\mathscr{H}}_{kl}$ are 4×4 matrices defined by Eqs. (8) and (9) are $\hat{0}$ represents a null matrix (4×4). It yields the following generalized field equations with the eight-current density defined by Eq. (5):

$$\partial^{\lambda}F_{\rho\lambda} = J_\rho \quad \text{and} \quad \partial^{\lambda}F^d_{\rho\lambda} = 0, \tag{12}$$

which are similar to usual Maxwell's equations and reproduce the sets of field equations (6), (7), and (10). In terms of the generalized eight-dimensional electromagnetic field tensor $F_{\rho\lambda}$, the Lorentz force acting on a charged particle $(-e)$ moving in such an electromagnetic field is given by the following relation:

$$f_\rho = -e F_{\rho\lambda} U^{\lambda}, \tag{13}$$

where U^{λ} is the eight-velocity of the particle defined in the following form as the combination of forward and inverse velocity four-vectors associated with subspace R^4 and T^4, respectively,

$$\{U^{\lambda}\} \equiv [\{v^{\mu}\};\{u^{\mu}\}]^T \equiv [(\vec{v},1);(\vec{u},1)]^T, \tag{14}$$

which contains both the forward $(\vec{v} = d\vec{r}/dt)$ and inverse $(\vec{u} = d\vec{t}/dr)$ velocities of a moving charge in R^4 and T^4, respectively. Substitution of the expressions of generalized electromagnetic field tensor and eight-velocity in Eq. (13) directly leads to the following Lorentz force equations in subspaces R^4 and T^4 separately:

$$\vec{F}_r = e[\vec{E}_r + (\vec{v}\times\vec{H}_r)], \quad f_t = e(\vec{E}_r\cdot\vec{v}), \tag{15a}$$

$$\vec{F}_t = -e[\vec{E}_t + (\vec{u}\times\vec{H}_t)], \quad f_r = -e(\vec{E}\cdot\vec{u}), \tag{15b}$$

where the eight-force vector has been taken as

$$\{f^{\rho}\} \equiv [(\vec{F}_r,f_t);(\vec{F}_t,f_r)]^T.$$

Equations (15a) represent the usual Lorentz force acting on a moving charged particle $(-e)$ and the rate of change of its energy in R^4 (with velocity as $\vec{v}$), while Eqs. (15b) give the quantities in T^4 (with velocity parameter $\vec{u}$) with negative strength.

As an alternative, if we interchange the role of electric and magnetic fields while passing to T^4 subspace from subspace R^4, then we have the following equations in place of those given by (4b):

$$\vec{E}_t = -\vec{\partial}_t\times\vec{\phi} \quad \text{and} \quad \vec{H}_t = -\vec{\partial}_t A - \partial_r\vec{\phi}, \tag{4b'}$$

and consequently, the field equations in subspace T^4 reduce to the following forms:

$$\vec{\partial}_t \cdot \vec{H}_t = -J, \quad \vec{\partial}_t \times \vec{E}_t + \partial_r \vec{H}_t = \vec{\zeta},$$

$$\vec{\partial}_t \cdot \vec{E}_t = 0, \quad \vec{\partial}_t \times \vec{H}_t - \partial_r \vec{E}_t = 0. \tag{7'}$$

These field equations when compared with those given by Eqs. (7) show that the fields defined by Eqs. (4b′) incorporate the interchange $\vec{E} \to \vec{H}$ and $\vec{H} \to -\vec{E}$ with charge and current source densities suitably interchanged while passing from subspace R^4 to subspace T^4. It shows the dual invariance of Maxwell's equations and reveals the behavior of electric charge in R^4 subspace as that of magnetic charge in T^4 subspace. It is also clear from these equations that the current density $\vec{\zeta}$ in subspace T^4 produces the electric field $\vec{E}_t$ in the similar manner as the current density $\vec{J}$ in subspace R^4 produces the magnetic field $\vec{H}_r$. Furthermore, the charge density J in subspace T^4 creates the longitudinal part of the magnetic field $\vec{H}_t$ in the similar manner as the charge density ζ in subspace R^4 produces the longitudinal electric field. This interchange in the roles of current and charge source densities while passing from subspace R^4 to subspace T^4 is consistent with the dual invariance of Maxwell's equations under superluminal Lorentz transformations. The electromagnetic field tensor corresponding to definition (4b′) is given by

$$\mathscr{H}_{\mu\nu} \equiv (\partial_t)_\nu \phi_\mu - (\partial_t)_\mu \phi_\nu \equiv \begin{pmatrix} 0 & -E_{t_z} & E_{t_y} & H_{t_x} \\ E_{t_z} & 0 & -E_{t_x} & H_{t_y} \\ -E_{t_y} & E_{t_x} & 0 & H_{t_z} \\ -H_{t_x} & -H_{t_y} & -H_{t_z} & 0 \end{pmatrix}. \tag{9'}$$

Using this field tensor (in subspace T^4) with that given by Eq. (8), the second type of generalized field tensor may be constructed in a similar way as given by relation (11). The resulting expression for Lorentz force remains the same as given by Eq. (13) which has its first four components ($\vec{F}_r$ and f_t) similar to those given by Eq. (15a), while the remaining four components are given by the following relations [in place of those given by Eq. (15b)]:

$$\vec{F}_t = -e[\vec{H}_t - (\vec{u} \times \vec{E}_t)] \quad \text{and} \quad f_r = -e(\vec{H}_t \cdot \vec{u}), \tag{15b'}$$

which are the force equations similar to those for a magnetic charge of strength $(-e)$ moving through the electromagnetic fields $\vec{E}_t$ and $\vec{H}_t$ with the subluminal inverse velocity $\vec{u}$ in subspace T^4. In other words, the electric charge in subspace R^4 when observed from T^4 subspace (i.e., seen as moving with superluminal velocity) appears as a magnetic charge moving with the subluminal velocity in T^4 subspace. It is in agreement with the result of Dattoli and Mignani.[15]

III. DISCUSSION

A unified electromagnetic theory in the proposed real eight-dimensional space-time geometry has been developed for extended objects (bradyonic as well as tachyonic), where for the bradyonic bodies the subspace R^4 is the representation space of the observables and the subspace T^4 is the internal subspace in which internal degrees of freedom of these bodies be represented. On the other hand, for tachyonic bodies T^4 is the representation space of the observables and the internal degrees of freedom are associated with the internal space R^4. A bradyonic body moving with the velocity $|\vec{v}| < 1$ in subspace R^4 (forward velocty $\vec{v} = d\vec{r}/dt$ in eight-dimensional space D^8), when viewed upon from the subspace T^4 appears as moving with the inverse velocity $\vec{u} = d\vec{t}/dr$ with $|\vec{u}| > 1$. Similarly, a tachyonic body moving with the velocity $|\vec{v}| > 1$ in subspace R^4 appears in subspace T^4 as moving with the inverse velocity $|\vec{u}| < 1$. As such, bradyons are localized in subspace R^4 while tachyons are localized in subspace T^4. Using the field definitions given by Eqs. (4), the electromagnetic field tensors in subspaces R^4 and T^4 and the corresponding field equations have been derived as Eqs. (8), (9), and (10), respectively, which show that while passing from R^4 to T^4, the bradyonic fields appear to behave as tachyonic and vice versa. In terms of these field tensors, the generalized field tensor in eight-dimensional space has been constructed as an 8×8 matrix given by Eq. (11) which has been shown to satisfy the general field equations (12) for the generalized current defined by Eq. (5), thereby showing the dual invariance of the field tensor. The conservation of the generalized eight-current also leads to the charge conservation in subspaces R^4 and T^4 separately. The eight-force experienced by a moving charge in these generalized fields with eight-velocity defined by Eq. (14), has been derived as Eq. (13) which is expressible in terms of Lorentz forces and energy rate changes in both the subspaces. Besides the usual fields (in subspace R^4), in order to incorporate the magnetic monopole type behavior, the generalized electromagnetic field tensor has also been constructed after interchanging the roles of electric and magnetic fields in subspace T^4 [Eq. (4b′)], which is shown to satisfy the same generalized field equations (12); but the second set of eight-force equations (15) changes to Eq. (15b′), which reveals that the electric charge (e) in R^4 when observed by an observer in T^4 behaves as a magnetic charge of strength $(-e)$.

ACKNOWLEDGMENT

One of the authors (H. C. C.) is thankful to the Council of Scientific and Industrial Research, New Delhi, for financial assistance.

[1] E. Recami and R. Mignani, Riv. Nuovo Cimento **4**, 209 (1974) and references therein.
[2] A. F. Antippa, Phys. Rev. D **11**, 724 (1975).
[3] P. Demers, Canad. J. Phys. **53**, 1687 (1975).
[4] E. A. B. Cole, Phys. Lett. A **76**, 371 (1980).
[5] P. T. Pappas, Lett. Nuovo Cimento **25**, 429 (1979).
[6] M. Pavšič, J. Phys. A **14**, 3217 (1981); Lett. Nuovo Cimento **30**, 111 (1981).
[7] C. Ramon, Found. Phys. **10**, 718 (1980).
[8] O. P. S. Negi and B. S. Rajput, Lett. Nuovo Cimento **32**, 117 (1981); J. Math. Phys. **23**, 1964 (1982).
[9] W. Thirring, Acta Phys. Austriaca Suppl. **9**, 256 (1972).
[10] W. Mecklenburg, Phy. Rev. D **21**, 2149 (1980).
[11] O. P. S. Negi, H. C. Chandola, K. D. Purohit, and B. S. Rajput, Phys. Lett. B **105**, 281 (1981).
[12] O. P. S. Negi and B. S. Rajput, Phys. Lett. B **113**, 183 (1982); Lett. Nuovo Cimento **34**, 180 (1982); Indian J. Pure Appl. Phys. **20**, 472 (1982).
[13] H. C. Chandola and B. S. Rajput, Nuovo Cimento B **72**, 21 (1982); Indian J. Pure Appl. Phys. **21**, 651 (1983).
[14] R. Mignani and E. Recami, Lett. Nuovo Cimento **16**, 449 (1976).
[15] G. Dattoli and R. Mignani, Lett. Nuovo Cimento **22**, 65 (1978).
[16] V. Vyšin, Lett. Nuovo Cimento **22**, 76 (1978).
[17] G. D. Maccarrone and E. Recami, Lett. Nuovo Cimento **34**, 251 (1982); Franscati Report INFN/AE-82/12, 1982.
[18] H. C. Chandola and B. S. Rajput, Indian J. Pure Appl. Phys. **22**, 1 (1984).
[19] H. C. Chandola and B. S. Rajput, Lett. Nuovo Cimento (in press).

Resolving the discrete ambiguities in amplitude determinations

Michael J. Moravcsik
Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403

(Received 13 January 1984; accepted for publication 25 May 1984)

Necessary and sufficient criteria are presented for eliminating discrete ambiguities in the determination of reaction amplitudes from the bilinear products of amplitudes that are measured in experiments. In terms of a geometrical analog, the criteria require closed loop diagrams with an odd number of broken lines and an odd number of solid lines. The results are applicable to the general mathematical problems of solving a set of simultaneous bilinear algebraic equations for several unknowns.

I. INTRODUCTION

The determination of reaction amplitudes from experimental measurements on reactions in molecular, atomic, nuclear, and particle physics is an interesting and subtle problem because it is nonlinear: The experimental observables are linear combinations of bilinear products of reaction amplitudes with their complex conjugates. For this reason, when we want to determine n complex reaction amplitudes (apart from one overall phase which is undeterminable from experiments on a purely phenomenological level), even if we choose an appropriate set of $2n-1$ experiments which fix these amplitudes so that no *continuum* of solutions can exist, there can still be several discrete solutions which then can presumably be reduced to only one acceptable solution by further experiments.

While the question of how to construct a set of experiments (which we will call a *complete* set) which eliminates *continua* of ambiguities for an arbitrary reaction has been discussed in the literature from many points of view,[1,2] the requirements for also eliminating *discrete* ambiguities (i.e., constructing what we will call a *fully complete set*) have been discussed only in connection with a few simple specific reactions. Yet the question is of importance in designing experimental programs for polarization experiments and for the analysis of the already large body of polarization data that is available. In particular, it was found recently[3] that the existing data for p.–p. elastic scattering at 6 GeV/c admit four amplitude solutions of almost equal statistical credentials, and this ambiguity makes dynamical deductions from those data less convincing.

The present paper therefore develops criteria for ways of reducing these discrete ambiguities. As we will see, the answer is simple and at the same time interesting, because the additional number of experiments needed is by no means a fixed number but depends on the type of measurements already performed. This suggests, therefore, that in order to be most economical and yet most impactful, one should plan the whole set of polarization experiments together, including the "additional" experiments that resolve discrete ambiguities.

The utility of the results of this paper may, however, be broader than merely in amplitude determination through polarization experiments. The theory of a set of bilinear algebraic equations for several unknowns is, as far as I could ascertain to my great surprise, nonexistent. Since such sets of equations appear very frequently in mathematical and physical situations, a contribution to their understanding may have a very broad impact.

This note will therefore discuss the nonlinear part of the problem: We are given the absolute value squares of n complex amplitudes a_i ($i=1,2,...,n$) as well as the real and imaginary parts of the $a_i\, a_j^*$'s. What subset of these n^2 quantities will allow us to determine the value of the a_i's without discrete ambiguities?

II. THE PROCEDURE

In the analysis of the situation I will consider the procedure in which the n magnitudes of the amplitudes are first determined by n measurements involving *only* the absolute value squares of the amplitudes. It has been shown that in any optimal formalism[4] there is such a set of n measurements. This determination of the magnitudes involves *no* discrete ambiguities, and the determination can usually be carried out to a considerable degree of accuracy even with measurements whose precision is short of spectacular.[5] The reason for this ease and accuracy is that once the absolute value *squares* are determined from the measurements from a set of *linear* equations, the magnitudes themselves, which are of course always positive, can be determined unambiguously and accurately. Thus it is very advantageous for any experimental program to start with this well-defined subset of experiments giving the magnitudes of the amplitudes.

From a purely mathematical point of view, of course, there are sets of bilinear combinations of amplitudes which do not admit even discrete ambiguities but which do not start with the determination of all magnitudes of the amplitudes. For example, the set $|a_1|^2$, Re $a_1a_i^*$, Im $a_1a_i^*$($i=2,3,...,n$) is an example for this. We know, however, from the structure of optimal formalisms that the actual polarization experiments do not lend themselves readily to the determination of just one magnitude by itself. Hence in practice, sets of this type would be more difficult to acquire than sets in which first all magnitudes are determined. It is for this practical reason that the present paper considers this latter case.

Once these magnitudes are known from the set of n measurements, an additional set of at least $n-1$ measurements are needed to determine the relative phases of the amplitudes in the complex plane, and it is at this stage that

 0022-2488/85/010211-03$02.50

discrete ambiguities can arise, since, for example, the determination of Re ab^* gives the angle between a and b, even if we know $|a|$ and $|b|$, only to within a double discrete ambiguity, namely, if a relative phase angle fits this Re ab^*, the negative of it also does.

Since the overall phase is arbitrary, as said before, one of the n amplitudes, let us say a_1, can be made pure real without sacrificing generality. Having done so, we have broken the symmetry between "real" and "imaginary," something that will reflect in the demonstration of the results. Had we, equally justifiably, set this first amplitude to be pure imaginary, the demonstration of the results would have held without modification except for the interchange "real"↔"imaginary." Had we set the phase of the first amplitude to be something other than a multiple of $\frac{1}{2}\pi$, the results, although substantially still the same, would have been formulated in a more elaborate format without gaining anything in the process.

III. THE THEOREM AND ITS PROOF

I will now state the main result of the discussion, namely a necessary and sufficient criterion for the elimination of even discrete ambiguities. This will be followed by a proof of the validity of that criterion.

The criterion and its proof will be described in the language of a geometrical analog, similar to the one used in a previous paper[2] discussing the determination of amplitudes. Let us denote each amplitude by a point, and each bilinear amplitude product by a line connecting the two points that correspond to the amplitudes that appear in the product. The line is solid if we have Re $a_i a_j^*$, and is broken (dashed) for Im $a_i a_j^*$.

For even just a complete (and not fully complete) determination of the amplitudes, the set of lines in our diagram corresponding to a complete set of bilinear combinations of amplitudes ("bicoms") must touch each amplitude point and must form a connected network. To be *fully* complete, the network must also satisfy the following two criteria.

(A) Each amplitude point must be included in a *closed loop*.

(B) At least one closed loop belonging to each amplitude point must have an odd number of broken lines and an odd number of solid lines in it.

To reiterate, these are criteria for a fully complete determination of reaction amplitudes for an arbitrary particle reaction (that is, a reaction containing particles with arbitrary spins) when the procedure for the determination of the amplitudes starts with a subset of experiments determining the magnitudes of the amplitudes.

In order to understand why these criteria are valid, let us first consider a set of lines in our diagram which starts at the "anchor" amplitude a_1 (the one we arbitrarily set to be pure real) and connects end-to-end to form an open and unbranched chain. In other words, we consider the bicoms $a_1 a_i^*, a_i a_j^*, \ldots, a_r a_s^*$ such that all the indices $1, i, j, \ldots, r$, and s are all different.

Let us first assume that all lines in this pattern are solid. In this case, let us track down the discrete ambiguities that exist in the knowledge of the amplitudes $a_1, a_i, a_j, \ldots, a_s$ when we already know the magnitudes $|a_1|, |a_i|, \ldots, |a_s|$ and we have exact measurements of Re $a_1 a_i^*$, Re $a_i a_j^*$, ..., Re $a_r a_s^*$. Let us denote by α_{ij} the relative phase angle between a_i and a_j, that is, if $a_1 = |a_i| e^{i\phi_i}$, then $\alpha_{ij} = \phi_i - \phi_j$. If we then start with a_1 (for which $\phi_1 = 0$ by convention), measuring Re $a_1 a_i^*$ yields ϕ_i or $-\phi_i$, thus leaving a two-fold ambiguity. Taking the next step by adding Re $a_i a_j^*$ will produce a similar twofold ambiguity in α_{ij}, with respect to ϕ_i, and hence, putting this together with the ambiguity in α_{1i} discussed above, we have now a fourfold ambiguity of the form $+\phi_i + \phi_j$, $+\phi_i - \phi_j$, $-\phi_i + \phi_j$, and $-\phi_i - \phi_j$. Continuing this process, after l steps we end up with a 2^l-fold ambiguity which can be described as

$$\pm \phi_i \pm \phi_j \pm \cdots \pm \phi_r \pm \phi_s. \tag{1}$$

Let us now add to the above set the solid line connecting point 1 with point s, that is, we add Re $a_1 a_s^*$. In other words, we make a closed loop out of the previous patterns of an open chain. In this case we now have the condition

$$\alpha_{1i} + \alpha_{ij} + \cdots + \alpha_{rs} + \alpha_{s1} = 0. \tag{2}$$

This condition will, in general, not be satisfied for all possible sets of ϕ_i's which were allowed by the ambiguities in Eq. (1). In fact, for *arbitrary* values of the ϕ_i's, only two of the 2^l previous solutions will satisfy Eq. (2), the two differing only in an *overall* sign. Thus we find that for a closed loop containing the "anchor" amplitude and containing only solid lines, we end up with one twofold discrete ambiguity regardless of the size of the loop.

For *special* sets of values for the ϕ_i's, the number of ambiguities may be large. To illustrate this, consider the highly unlikely and artificial situation in which $\alpha_{1i} = \alpha_{ij} = \cdots = \alpha_{rs} \equiv \beta$, and, say, $\alpha_{1s} = k\beta$, where k is an integer and $k < m$, where m is the number of indices in the set $\{i,j,\ldots,r,s\}$. In that case, even if we measure α_{1s}, that value of $\alpha_{1s} = k\beta$ can be reconciled with 2^{m-k} choices of signs in Eq. (1), even apart from the twofold ambiguity due to the overall sign change mentioned earlier.

While such special sets of values of the ϕ's are unlikely to be relevant when we deal with measurements of *infinite* accuracy, their case is pertinent when we deal with experiments of *finite* precision, especially if this precision is quite limited. Although the method of analysis of such "real-life" situations is given by the above observations, the conclusions of such an analysis must be determined for each situation separately, since they depend on the individual values of the ϕ's as well as on the experimental uncertainties in the set of measurements under consideration.

Let us now return to the consideration of the case of precise experiments and the chain of measurements discussed above. So far we considered only chains containing only solid lines. Let me now take a similar open chain, starting from the "anchor" amplitude, in which the first few links are solid lines and the remaining links are broken lines. In a way which is directly analogous to the procedure we used for the all-solid chain, we get the ambiguity pattern [corresponding to Eq. (1)], which is

$$\pm \phi_i \pm \phi_j \pm \cdots \pm \phi_p + \begin{Bmatrix} \phi_q \\ \pi - \phi_q \end{Bmatrix} + \cdots + \begin{Bmatrix} \phi_s \\ \pi - \phi_s \end{Bmatrix} \tag{3}$$

and if we close *this* chain by, say, Im $a_1 a_s^*$, we get the condition analogous to Eq. (2),

$$\pm\alpha_{1i}\pm\alpha_{ij}\pm\cdots\pm\alpha_{p-1,p}+\begin{Bmatrix}\alpha_{pq}\\ \pi-\alpha_{pq}\end{Bmatrix}+\cdots+\begin{Bmatrix}\alpha_{s1}\\ \pi-\alpha_{s1}\end{Bmatrix}=0. \tag{4}$$

Equation (4) can be rewritten, remembering that a phase has meaning only *modulo* 2π. The ambiguity then is

$$\pm\alpha_{1i}\pm\cdots\pm\alpha_{p-1,p}+\begin{Bmatrix}\sum\limits_{i+1=q}^{1}{}''\alpha_{i,i+1}\\ \pi+\sum\limits_{i+1=q}^{1}{}'\alpha_{i,i+1}\end{Bmatrix}=0, \tag{5}$$

where Σ'' is any of the many summations of the α_{ij}'s in which an even number of coefficients of the α's are -1 and the rest $+1$, and where in the similar Σ' there is an odd number of -1's among the coefficients.

We can now ask if the ambiguity we found in the case of an all-solid line loop exists here or not. To find the answer, let us consider the situation in which

$$\alpha_{1i}+\cdots+\alpha_{p-1,p}+\alpha_{pq}+\cdots+\alpha_{s1}=0. \tag{6}$$

Is there now another combination of α's of the form given by Eq. (5) which is 0? If the number of broken lines is even, the answer is "yes," namely

$$-\alpha_{1i}-\cdots-\alpha_{p-1,p}-\alpha_{pq}-\cdots-\alpha_{s1}=0 \tag{7}$$

also holds, and this combination also appears among those in the *upper* line of Eq. (5). If, however, the number of broken lines is odd, the combination on the left-hand side of Eq. (7) appears in the *lower* line of Eq. (5) and hence with an extra π in the combination, which should vanish with this extra π and hence will not without it. Thus the twofold ambiguity experienced for loops with all solid lines does not exist for loops with an odd number of broken lines. The proof is similar in the case when in Eq. (6) some of the signs are negative rather than positive.

In the above proof we considered the case when the first few lines are solid and the remaining ones broken, but that restriction is clearly inessential: The structure of Eqs. (3)–(7) is evidently independent of this restriction and so are the conclusions derived from it.

So far we considered only bicom sets which form a loop in the geometrical analog, with no additional lines in the diagram. If the loop also has "dangling" open chains sticking out of it, for each such chain consisting of z segments we have a 2^z-fold ambiguity. To eliminate such a set of ambiguities, we must close the chain into a loop the same way as we did for the original loop, making sure that the newly closed loop now has an odd number of broken lines in it. If by closing the chain into a loop we make the chain part of several loops, only one of them needs to contain an odd number of broken lines.

As we said earlier, any complete (but not yet fully complete) pattern will have at least $n-1$ lines in it (after we have already determined the n magnitudes of the amplitudes). Since the closing of *each* loop in a pattern requires one appropriately chosen line, it is clear that, from the point of view of eliminating discrete ambiguities, the most economical complete (but not yet fully complete) set is one large open chain including all amplitudes, since in that case just *one additional* measurement can eliminate *all* discrete ambiguities. The least economical set is the one with a "porcupine" diagram in which all $n-1$ lines emanate from the "anchor" amplitude and connect it with each of the other $n-1$ amplitude points. In that case, we need $\frac{1}{2}(n-1)$ additional bicoms to close all chains into loops if n is odd, and $\frac{1}{2}n$ if n is even.

In order to complete the proof, all we have to remark now is that the whole argument above could have been made with "real" and "imaginary" (or "solid" and "broken") interchanged. For example, in connection with an all-broken (rather than all-solid) chain, we would have, instead of Eq. (1),

$$\begin{Bmatrix}\phi_i\\ \pi-\phi_i\end{Bmatrix}+\begin{Bmatrix}\phi_j\\ \pi-\phi_j\end{Bmatrix}+\cdots+\begin{Bmatrix}\phi_r\\ \pi-\phi_r\end{Bmatrix}+\begin{Bmatrix}\phi_s\\ \pi-\phi_s\end{Bmatrix}, \tag{8}$$

etc. It is for this reason that criterion B calls for an odd number of solid lines as well as an odd number of broken lines.

This completes the demonstration of the validity of the criteria for a fully complete set of bicoms. The criteria give a simple and general prescription for how to augment an already complete set into a fully complete one when we deal with measurements of infinite accuracy. As mentioned earlier, the prescription also gives the *minimal* augmentation needed when we take into account experimental errors. The *maximal* augmentation in such a case has not much meaning, because clearly measurements with very large errors may serve no purpose at all. Even in somewhat better situations the judgment of whether a discrete ambiguity is eliminated or not will depend on judgments on what differences in chi-squares are found convincing. Nevertheless, the procedure outlined in this discussion pertaining to exact measurements also serves as a guide for the analysis of measurements with finite uncertainties.

[1]See, for example, C. Schumacher and H. A. Bethe, Phys. Rev. **121**, 1534 (1961); M. J. Moravcsik, Phys. Rev. **125**, 1088 (1962); B. Anderson, Nucl. Phys. **60**, 144 (1964); M. J. Moravcsik, in *Proceedings of the Second International Symposium on Polarization Phenomena of Nucleons* (Basel, 1966), p. 159; M. Simonius, Phys. Rev. Lett. **19**, 279 (1967); P. Csonka, M. Moravcsik, and M. Scaddron, Rev. Mod. Phys. **39**, 178 (1967); M. J. Moravcsik, Phys. Rev. **170**, 1440 (1968); M. Moravcsik and W. Yu, J. Math. Phys. **10**, 925 (1969); C. Bourrely and J. Soffer, Phys. Rev. D **12**, 2932 (1975); P. W. Johnson, R. C. Miller, and G. H. Thomas, Phys. Rev. D **15**, 1895 (1977); P. La France and P. Winternitz, Phys. Rev. D **27**, 112 (1983); and many others.

[2]G. Goldstein, M. Moravcsik, and D. Bregman, Lett. Nuovo Cimento **II**, 137 (1974).

[3]N. Ghahramany, G. R. Goldstein, and M. J. Moravcsik, Phys. Rev. D **28**, 1006 (1983).

[4]M. J. Moravcsik, Phys. Rev. D **22**, 135 (1980), Table I; G. R. Goldstein and M. J. Moravcsik, Ann. Phys. (N.Y.) **142**, 219 (1982).

[5]G. R. Goldstein and M. J. Moravcsik, Nuovo Cimento A **73**, 196 (1983); G. R. Goldstein and M. J. Moravcsik, Phys. Lett. B **102**, 189 (1981); and Ref. 3.

ERRATA

Erratum: On the hydrodynamic self-similar cosmological models [J. Math. Phys. 24, 2532 (1983)]

L. K. Chi
Computer Science Department, United States Naval Academy, Annapolis, Maryland 21402

(Received 1 October 1984; accepted for publication 12 October 1984)

The first equation of Eqs. (1) should read

$$\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial r} = -\frac{\rho}{r^2}\frac{\partial}{\partial r}(r^2 v).$$

Also, the correct value of the constant η_0 is $1/6\pi$, instead of the value $\frac{1}{6}\pi$ quotated in the paper.

All other equations and results remain unaltered.

Erratum: Higher-dimensional Riemannian geometry and quaternion and octonion spaces [J. Math Phys. 25, 347 (1984)]

J. W. Moffat
Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7

(Received 3 October 1984; accepted for publication 12 October 1984)

Equation (2.7) should read:

$$\Omega_\Sigma = \frac{1}{2}\sigma^{[AB]}\, e^\Lambda_A(x)\left[\frac{\partial}{\partial x^\Sigma} e_{B\Lambda}(x) - e_{B\Omega}\,\Gamma^\Omega_{\Sigma\Lambda}\right]. \quad (2.7)$$

 0022-2488/85/010214-01$02.50